UNIVERSITÀ DEGLI STUDI DI CATANIA
SCUOLA SUPERIORE DI CATANIA

International PhD in
Nuclear and Particle Astrophysics
XXIV cycle

PROPAGATION OF ULTRA-HIGH ENERGY COSMIC RAYS AND ANISOTROPY STUDIES WITH THE PIERRE AUGER OBSERVATORY: THE MULTISCALE APPROACH

Manlio De Domenico

A dissertation submitted in partial fulfillment of the requirements for the degree Doctor of Philosophy in Nuclear and Astroparticle Physics

Coordinator of PhD:
Prof. U. Lombardo

Advisors:
Prof. A. Insolia
Dr. P. L. Ghia

2008-2011

Abstract

The main goal of this thesis is the investigation of the arrival directions of ultra-high energy cosmic rays (UHECRs). The clustering of such events, collected by the Pierre Auger Observatory, is used to infer the nature of sources and their density, as well as other physical unknowns. Results are based on the comparison between real and simulated data. Hence, an *ad hoc* Monte Carlo code (HERMES) for the realistic simulation of UHECR propagation in the Universe and a novel method (MAF) to quantify the amount of clustering in a data set of few UHECR events have been developed.

In the first chapter, a general overview of UHECR physics will be given, with particular attention to the Pierre Auger Observatory and the most recent results regarding its measurements.

In the second and third chapters, we will present the general structure of our propagation code and we will discuss all the models, the parameterizations and the procedures adopted to simulate the propagation of UHECRs in a magnetized Universe. In the second chapter, magnetic fields are treated, and their impact on the propagation of UHECRs is discussed. In particular, we will simulate the diffusion of charged particles in both turbulent and structured magnetic fields for energy values ranging from 10^{17} eV to 10^{21} eV. In the third chapter, the propagation of protons, heavier nuclei, photons and neutrinos, will be treated in the absence of magnetic fields. We will define the cosmological framework of HERMES and we will treat the parameterizations adopted to simulate the extragalactic background radiation. The parameterizations for the cross section of $p\gamma_{\rm EBR}$, $A\gamma_{\rm EBR}$ and $\gamma\gamma_{\rm EBR}$ interactions are discussed, as well as the relevant energy-loss processes. In particular, pair and pion production, as well as photo-disintegration in the case of heavy nuclei, are treated with great detail. Mean free path and energy-loss length are numerically estimated with HERMES: their dependence on nuclear mass and their evolution with redshift are also discussed. The impact of propagation effects on the GZK horizon of UHE protons is investigated, and some comparisons between our results and those obtained with other simulators available to the UHECR community are presented.

In the fourth chapter, a novel method is introduced to estimate the statistical significance of clustering in the arrival direction distribution of few events, a necessary requirement because of the current small number of events observed above 5×10^{19} eV. The method involves a multiscale procedure, based on information theory and extreme value statistics, providing high discrimination power, even in presence of strong background isotropic contamination. It is naturally extended to allow multiscale cross-correlation analysis with candidate sources of UHECRs. Here, the term "multiscale" explicitly indicates the dependence on the angular scale adopted to investigate the arrival directions of UHECRs. It is shown that multiscale methods have some valuable features: i) they are semi-analytical, drastically reducing computation and allowing a larger parameter space to be explored in reasonable amount of time, ii) they are very sensitive to small, medium and large scale clustering, iii) they are not biased against the null hypothesis.

Finally, in the fifth, sixth and seventh chapters, Monte Carlo simulations, required because of the stochastic nature of some interactions between UHECRs and background photons, are extensively adopted to investigate real data. In the fifth chapter we use multiscale methods to explore the effects of experimental uncertainties on clustering and correlation of UHECRs with catalogs. In the

sixth chapter, energy losses due to secondary particle production (electron/positron pairs and pions) or photo-disintegration, as well as deflections due to galactic and extragalactic magnetic fields, are taken into account. All of such interactions, together with the distribution of (still unknown) sources, produce different distributions of arrival directions of events observed at Earth. Hence, multiscale clustering in events detected with the Pierre Auger Observatory and in simulated sky maps of UHECRs, mimicking realistic astrophysical scenarios, is used to put bounds on some relevant unknowns, as the fraction of protons in the data, the density of sources and the strength of the turbulent component of the extragalactic magnetic field. Moreover, the possibility that nearby active galactic nuclei and black holes could be responsible for the observed flux of UHECRs is explored in detail. In the seventh chapter, we perform a more extended study which takes into account additional observables, as the elongation rate and the energy spectrum. By varying the underlying assumptions, as for instance those ones on the mass composition and the intensity of magnetic fields, we have outlined an astrophysical scenario able to explain Auger data from a phenomenological point of view.

Acknowledgements

My PhD has been a very intense period of activities, characterized by so many experiences and involving so many people, that it is difficult to "backtrack" step-by-step up to the first day. The best way is to walk like a drunk guy in the turbulent space of memory, wishing to visit any cell of brain without undergoing photo-disintegration (that is a complicated way to express a simple concept: I hope to remember here as much as possible in a finite amount of time and space).

My first acknowledgment goes to two people who, unfortunately, are no more among us, at least not in their traditional shape. The first one, Prof. Giovanni Raciti, has transmitted to me his passion for physics and his personal approach to problem solving, in a so funny (and sometimes severe) way that it is not possible to forget him, for me. The second one, Gigi, the brother I have never had, has transmitted to me his passion for life and for human beings in a so original way that he has a non-negligible impact on my life, still now: he has given me his friendship every day during the last seven years, with a constant interest for my activities and with constant complicated (still unanswered) questions about nature and about life.

I cannot miss here to thank my family, for the patience and for the support, and in particular to thank my sister for having care of me (in her personal approach) in the last two years of my life and for being so funny (at least 90% of time): awakenings when she was dancing on my stomach will be impressed on my mind for ever.

It is not possible to miss to thank Cosimo, Danilo, Fabio, Serena and "the Giuntas" for their friendship and another uncountable number of reasons. Mr. Giacomo should be acknowledged because of his invaluable help and the time spent to repair my car. I would like to include a special acknowledgement to Claudio, for being like a brother almost every day of his life in the last seven years and for inciting me to give always my best.

I would like to thank the Auger group in Catania: Antonio (a.k.a. Prof. Insolia) for keeping me free to follow my interests and for supporting me and my work in the best possible way, Rossella for travelling with me almost 40,000 miles by flight, for her support (especially in the last year) and for funny moments together around the world, and Enrica, Mario and Simone for the nice moments in our small lab.

Of course I can not miss to thank Vito (a.k.a. Prof. Latora), for inviting me to join the Laboratory for Complex Systems and, overall, for transmitting to me his passion for anything is complex and for the funny late evenings in working to papers. I also thank "Vito's guys", for accepting my presence in the Lab and providing a nice working place in a nice group of people: Alessio, Enzo, Massimo, Roberta and Salvo.

I would like to acknowledge the coordinator of my PhD, Prof. Lombardo, as well as the Scuola Superiore di Catania, for supporting my studies in the best possible way and for following with interest their development.

I can not miss to thank Maria Ippolito and Bice Immè for the help they have given to me for what concerns one of the last unsolved mysteries of XXI century: bureaucracy. Without their help, I would have wasted a non-negligible amount of my time in solving issues too much long and complicated for

my simple mind.

I guess that I can not add acknowledgments to the other chapters of my thesis, thus I am going to conclude with people from Auger. I would like to thank any collaborator for his/her patience to listen to my talks and a large number of people for many interesting and useful discussions about the physics of cosmic rays. In particular, I would like to thank Laurentiu Caramete and Peter Biermann, for their kindness and their spirit of collaboration, as well as for very interesting discussions.

I would like to thank Piera (a.k.a. Dr. Ghia) for supporting me and my work since the first day, and for believing in my studies more than me. The period of time dedicated to the correction of my thesis has been one of the most difficult in the last three years for several reasons, and I have to acknowledge her passion for CR physics (and, in particular, for Auger) and her patience with me, if the final result is what it is. We have worked hard, but we did it!

I would like to thank the Auger group in Orsay, Haris, Isabelle, Olivier and Tiina, for making me feel like at home during my short permanence in Orsay (as well as Mario Chavez and his wife Laura, even if they are not in the Auger collaboration), and, in particular I would like to thank Olivier for his great patience and his invaluable help in the key parts of my studies. Of course I can not miss to thank people of the Large Scale Anisotropy task for the huge amount of useful discussions, informations and feedbacks provided during my studies.

I would like to thank the young "italian guys" (Fausto included, of course) of the Auger collaboration for nice moments around the world, with more regards for Denise and Mariangela, because of interesting discussions and the interesting studies that we have done, we are doing and, probably, we will do. In particular, I would like to thank Mariangela for her precious friendship and the very funny moments in Beijing, in Catania and in Malargüe.

A special thank is for Haris, for sharing with me his passion for physics, his philosophy, a non-negligible fraction of his time during the last 2 years and, overall, for his precious friendship, as well as for a lot of funny moments (and many beers) around the world. I must acknowledge his mother Nicole, for kindly having me at her home during my permanence in Paris.

Last, but not least for importance, is the more special thank to Serafina (and her enlarged family), simply for her existence and any of her expressions. I thank her for the patience with me, for having care of me, for supporting me, for her kindness, for her help and for her time, but, overall, for her crazy idea to share her whole life with me.

In the last three years I had the great opportunity to travel a lot to deepen my studies. However, I have also seen Galician bagpipers playing for Galician dancers in the square, Chinese old musicians playing the erhus in Beijing's hutongs, guanacos running free in the Pampa Amarilla, street performers from all regions of the world in London, Paris and Seattle, as well as the sunlight and the sound of rain in all of such wonderful places. I am glad for this.

Within his undergraduate thesis in 2004-2005, Mario Scuderi adopted a physical language for acknowledgments. In 2007-2008 I have extended such an approach in the framework of chaos theory for my Msc thesis and he has acknowledged such an extension in his PhD thesis, in the same year. Here, I would like to keep trace of such thoughts, by concluding with

"I thank all the initial conditions and all the external factors, appeared or disappeared over time, which allowed me to reach this final status."

Contents

List of Tables

List of Abbreviations

ΛCDM	Λ−Cold Dark Matter
ACF	AutoCorrelation Function
ASS	AxiSymmetric Spiral
BSS	BiSymmetric Spiral
CEL	Continuous Energy Loss
CR	Cosmic Rays
CGB	Cosmic $\gamma-$ray Background
CIB	Cosmic Infrared Background
CIOB	Cosmic Infrared/Optical Background
CMB	Cosmic Microwave Background
COB	Cosmic Optical Background
CRB	Cosmic Radio Background
CUVOB	Cosmic UltraViolet/Optical Background
CXB	Cosmic X-ray Background
EAS	Extensive Air Shower
EBL	Extragalactic Background Light
EBR	Extragalactic Background Radiation
EMF	Extragalactic Magnetic Field
GMF	Galactic Magnetic Field
GZK	Greisen-Zatsepin-Kuzmin
HECR	High Energy Cosmic Rays
HMR	Harari-Mollerach-Roulet
IRB	InfraRed Background
ISM	InterStellar Medium
LDF	Lateral Distribution Function
MHD	MagnetoHydroDynamics
SFR	Star Formation Rate
SNR	SuperNova Remnant
UHECR	Ultra HECR

By striving to do the impossible, man has always achieved what is possible. Those who have cautiously done no more than they believed possible have never taken a single step forward

M. A. Bakunin

Chapter 1

Ultra-high energy cosmic rays and the Pierre Auger Observatory

Cosmic rays (CRs) are generally considered to be charged nuclei originating outside the solar system. Such nuclei have been detected and their energy spectrum spans more than 14 orders of magnitude, from few MeV to hundreds of EeV (1 EeV = 10^{18} eV). As many other topics, cosmic ray physics was born as an observational science. In this chapter, we will briefly describe the history of measurements that carried such an exciting field to the ultra-high energy (UHE) era and to the possibility of particle astronomy. Moreover, we will present and discuss the main experimental results regarding the measurement of the energy spectrum and the mass composition of CRs, with particular attention to those events of interests for the present thesis, above 1 EeV.

We will overview the possible production mechanisms responsible for the acceleration of charged nuclei up to highest energies and we will identify the candidate sources of such events. We will sketch the theoretical description of the extensive air shower, together with the corresponding detection techniques.

Successively, we will focus our attention to the largest observatory of ultra-high energy CRs (UHE-CRs), the Pierre Auger Observatory. We will describe its innovative detection technique, with particular attention to event selection and reconstruction. Finally, we will show the up-to-date results obtained with Auger measurements. In particular, we will show energy spectrum and mass composition measurements, as well as the intrinsic anisotropy of their arrival direction at Earth and their correlation with the position of active galactic nuclei (AGN), one of the most promising acceleration sites for such extraordinary events.

1.1 Historical highlights

The first experimental evidences of the existence of cosmic rays are due to the pioneeristic measurements led by D. Pacini between 1907 and 1911 and by V. Hess between 1910 and 1912. By observing the variations of the rate of the ionization on mountains, over a lake, over the sea, and underwater, Pacini concluded that ionization underwater was significantly lower than on the sea surface, showing that such a behavior could not be attributed to the radioactivity of the Earth [1]. By measuring the ionization rate of air as a function of altitude, by mean of hydrogen-filled balloons, Hess found an increase of ionizing radiation with increasing height and he concluded that radiation penetrates from outer space into the atmosphere [2]. In 1913 and 1914, W. Kolhörster repeated the experiment led by Hess up to an altitude of 9 km. Such experimental results confirmed that ionization rate from this radiation was greater at that altitude than at sea level and that the source for these ionizing rays was of extraterrestrial origin [3, 4]. For his discovery, Hess received the Nobel Prize in Physics in 1936. The term “cosmic rays” was

coined by R. Millikan [5] who confirmed in 1926 they were extraterrestrial in origin, and not produced by atmospheric electricity [6–8].

In 1929, Bothe and Kolhörster used for the first time two Geiger-Müller counters to measure coincident signals, and by placing absorber material in between the two counters they concluded that the cosmic radiation consists of charged particles [9]. Just one year after, J. Clay showed that the intensity of cosmic rays depends on the (magnetic) latitude of the observer, further corroborating the hypothesis that ionization radiation from the outer space consists of charged particles [10].

In 1930, B. Rossi used several Geiger counters, powered with a bank of batteries, and invented the first coincidence circuit for particle detection [11]. In the same year, from studies of Störmer's theory about the motion of a charged particle in a dipole magnetic field like that of the Earth, he predicted that if the primary cosmic rays were charged particles with one charge sign, then a difference between the cosmic ray intensities from the east and the west should be measured [12]. The experimental evidences for such a prediction has been found later, in 1934 [13–15]. Between 1931 and 1932, Rossi carried out three experiments, based on triple coincidence among three Geiger counters, of fundamental importance in the development of cosmic-ray and particle physics. The choice of the triple coincidence requirement dramatically reduced the rate of accidental coincidences and he proved the presence in cosmic rays of penetrating particles with energies more than 2 GeV. In another experiment, Rossi placed three counters in a triangular configuration surrounded by lead shielding. Since a single penetrating particle could not discharge all three counters, triple coincidences showed that interactions of cosmic rays in the shield above the counters produced showers of particles. The result was so astonishing that one journal refused to publish it, and another accepted it only after Werner Heisenberg vouched for Rossi's reliability. In 1933, Rossi measured the occurrence rate of showers as a function of the thickness of a screen of lead or iron placed above the triangular arrangement of counters [16]. The result demonstrated the presence in cosmic rays of two distinct components: a soft component, soon identified as electrons and photons, which readily produces showers and is rapidly absorbed in lead, and a hard component, consisting of penetrating particles (muons) that produce showers much less readily, and is only gradually attenuated in lead. The measurements led by Rossi and their dependence on the atomic number of the screen provided a critical test of the Bhabha-Heitler theory of electron-photon cascade showers. It is worth noticing that muons were discovered after just three years by C. Anderson and S. Neddermeyer, while studying cosmic radiation. In 1934, Rossi discovered extensive cosmic-ray air showers. In order to check the rate of accidental coincidences between pulses from the two Geiger counters he was using in his experiment, he placed the two counters some distance apart in a horizontal plane. Rossi found the coincidence rate was larger than the expected accidental rate derived from the counting rates of the individual Geiger counters and the resolving time of the coincidence circuit[1] [18]. In practice, cosmic rays penetrating the atmosphere with very high energy produce a shower of secondary particles, whose detection in coincidence, by two or more detectors, is still the basic scheme of a large class of modern detectors.

In 1933, A. Compton published the measurements of the intensity of cosmic rays by 8 different expeditions at 69 stations distributed at representative points over the Earth's surface, concluding that the data can be quantitatively explained on the basis of Lemaitre and Vallarta's theory of electrons approaching the earth from remote space [19]. Such an interpretation contradicted that of Millikan, who believed that cosmic rays were high-energy photons with some secondary electrons produced by Compton scattering of gamma rays [20].

In 1934, Compton confirmed that photons could not constitute more than a negligible part of primary cosmic rays: in fact, comparison of high altitude measurements at the equator with those at high latitudes showed that the equatorial cosmic rays were closely similar to the magnetically deviable rays

[1] The brief summary of the extraordinary activity of B. Rossi is partially extracted from the recent review of G. Clark [17].

both in their absorption in air and in their transition effects. This effect indicated that most of the cosmic rays which reached the earth were similar in character to the deviable rays, most probably protons [21]. Moreover, he reported that the coincidence experiments of Bothe-Kolhörster, Rossi and others, considered in the light of various alternative interpretations, showed the existence of penetrating, electrically charged particles which were either primary cosmic rays or were secondaries of primaries absorbed high in the atmosphere. The shower-producing radiation, which seemed to consist of photons, had to be produced by some electrically charged primary rays, which he provisionally identified as electrons [22].

In 1938, Kolhörster reported the discovery of coincident signals between two tubes set as far as 75 m apart, by using Geiger-Müller tubes operated in coincidence: he concluded that the tubes were hit by secondary particles or showers generated by cosmic rays in the atmosphere [23]. With a similar technique, Pierre Auger investigated air showers and, apparently unaware of Rossi's report, claimed their discovery, concluding that the registered particles are secondaries generated in the atmosphere, originating from a single primary cosmic ray [24].

The theoretical background for such discoveries has been developed in the successive years, together with more sophisticated experiments devoted to measure the mass composition of primary cosmic rays and their energy spectrum, discussed in the next sections.

1.2 The energy spectrum and the mass of CRs

The energy of cosmic rays extend from the MeV range to at least 10^{20} eV. The differential energy spectrum of all cosmic-ray particles (i.e. integrated over mass) is shown in Fig. 1.1, multiplied by a factor E^2 to facilitate visualization. It falls steeply as function of energy, decreasing by about a factor 500 per decade in energy. At GeV energies, more than 1000 CRs per second and square meter are observed: such a flux decreases to about one particle per square meter and per year at a PeV, and further to less than one particle per km^2 and per century above 100 EeV. Such a dramatic decrease in flux represent one of the main obstacles to the detection of high energy CRs. At sub-GeV energies individual isotopes are measured with small detectors in outer space and individual elements can be resolved with balloon-borne detectors in the TeV regime. Above energies exceeding 100 TeV large detection areas are required to collect a suitable number of particles in a reasonable time. At present, such detectors are realized at ground level only and secondary particles generated in the atmosphere, namely the extensive air showers, that will be described in Sec. 1.5, are registered. At PeV energies, mass identification becomes difficult and model-dependent, although groups of elements may still resolved. At the highest energies even a classification into "light" and "heavy" particles becomes an even harder experimental challenge.

Fig. 1.2 shows the relative abundances of CRs, as a function of nuclear charge, measured at the lowest energies (around 1 GeV/n), normalized to Si = 100. Qualitatively, it is similar to that in the solar system (solid line), with differences around $Z = 2 - 7$ (hydrogen isotopes, helium, lithium), $Z = 9$ (Berillium) and $Z = 20 - 26$ (Iron group). Such elements are more abundant in CRs than in solar system, because they are assumed to be produced in spallation processes of the more abundant particles of the CNO, iron, and lead groups during the propagation of cosmic rays through the Galaxy. Hence, they are frequently referred to as secondary cosmic rays. As the spallation cross section of the relevant nuclei is known at GeV energies, the ratio of secondary to primary cosmic rays is used to infer the propagation path length of cosmic rays in the Galaxy (see [26] and Ref. therein).

At the highest energies, an indicator of the mass of UHECRs is the average depth of the shower maximum of $X_{\max}$. We will define it further in the text, where we will describe extensive air showers: within this context, we are interested only in showing how the mass composition changes with energy, from 10^{14} eV up to $\approx 10^{20}$ eV. In Fig. 1.3 such an indicator is shown, compared to predictions for proton

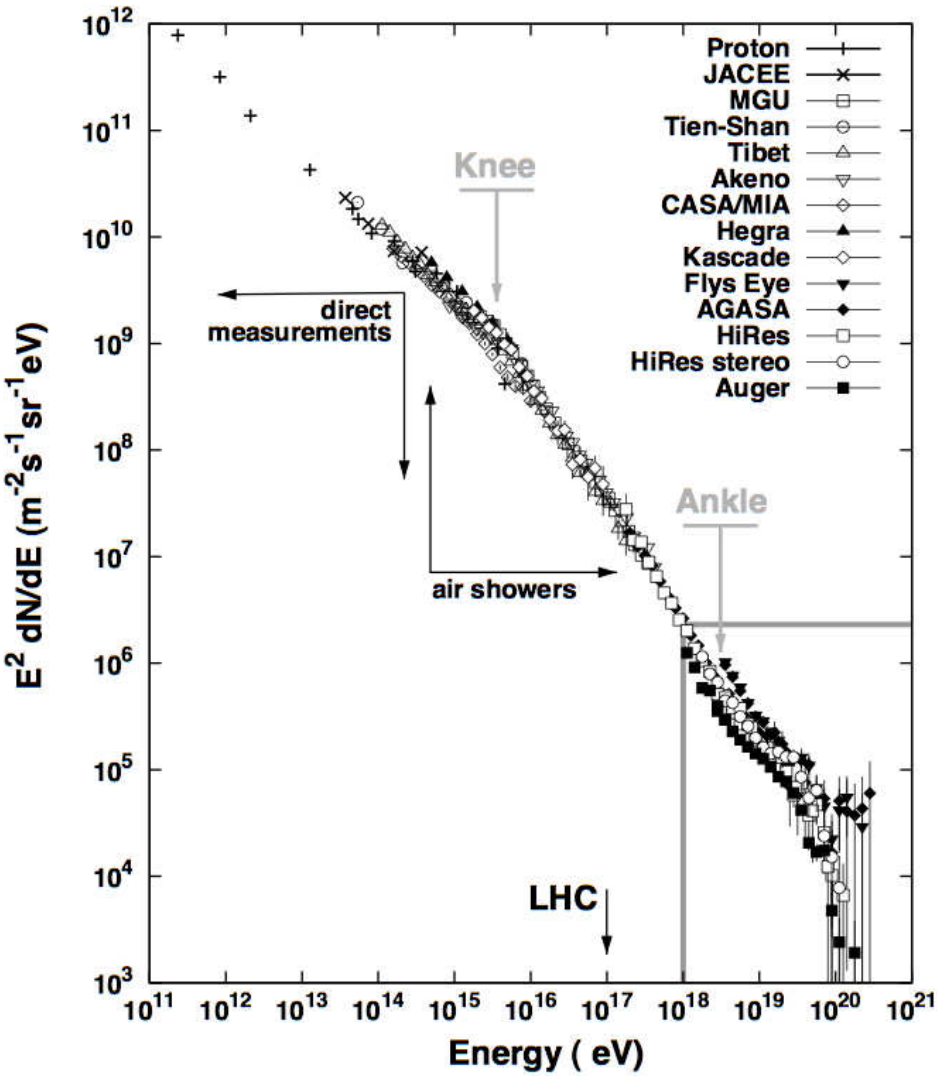


Figure 1.1: All-particle energy spectrum of UHECRs measured by several experiments [25]

and iron induced showers obtained with different hadronic models.

The energy spectrum of single species, as well as the all-particle one, behaves like $\sim E^{-\gamma}$ over a wide energy range, indicating non-thermal acceleration processes. At a first sight, it appears rather featureless, although it is not the case after a more careful inspection. To put in evidence its structures, the flux is generally multiplied by factors of type E^{κ}, as shown, for instance, in the right panel of Fig. 1.4 (where $\kappa = 2.7$). The spectral index is rather constant ($\gamma \simeq 2.7$) up to $10^{15.5}$ eV (see the left panel of Fig. 1.4, showing the measured energy spectrum from GeV to PeV energies for different species), where a steepening is observed ($\gamma \simeq 3.0$), the so-called knee. At about $10^{17.5}$ eV a further steepening ($\gamma \simeq 3.3$) is observed, the second knee. At about $10^{18.6}$ eV a new change appears, i.e. the ankle. Above the ankle, the spectrum flattens again ($\gamma \simeq 2.6$). Above $10^{19.5}$ eV the flux dramatically drops down ($\gamma \simeq 4.3$) [28]. Intriguingly, in this last region a strong suppression of the flux is expected because of the so-called Greisen-Zatsepin-Kuzmin (GZK) effect [29, 30], that will be described in much more detail successively in this thesis, where we will discuss the propagation of UHECRs. On of the main results of the Pierre Auger deals with the experimental evidence of such a suppression, and will be discussed at the end of

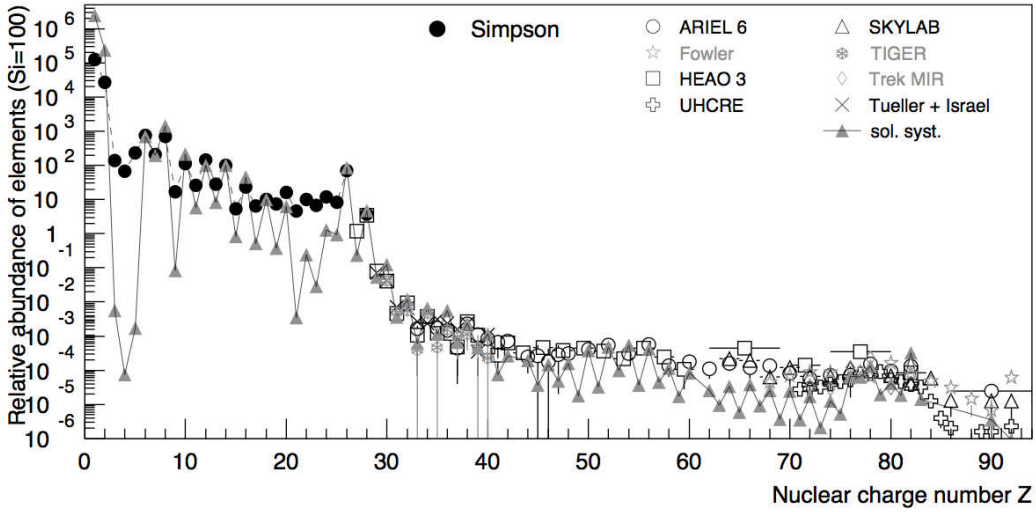


Figure 1.2: Abundance of elements, measured by several experiments, as a function of nuclear charge (at energies around 1 GeV/n, normalized to Si = 100) [26]. The solid line indicates the abundance in the solar system..

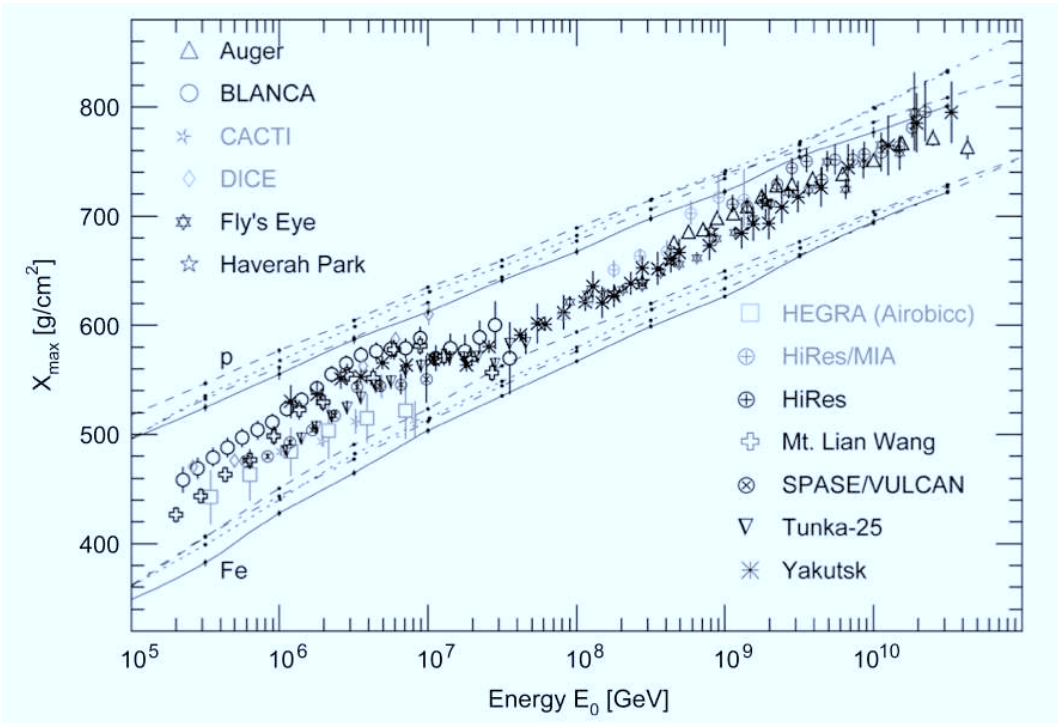


Figure 1.3: Average depth of the shower maximum of $X_{\max}$ from several experiments, as a function of energy. The lines indicate predictions for proton and iron induced showers obtained by using the CORSIKA code with different hadronic models [26].

this chapter. Such structures are particularly evident in the right panel of Fig. 1.4, and are strongly related to the origin of CRs.

CRs with energy up to a few PeV are thought to be of galactic origin, being generated from the explosion of Supernovae. The galactic origin, together with diffusive shock acceleration mechanisms (see further in this chapter), may explain the observed flux up to energies of the order of 10^{17} eV (see the left panel of Fig. 1.5), above which a new component of extragalactic origin may raise, accounting

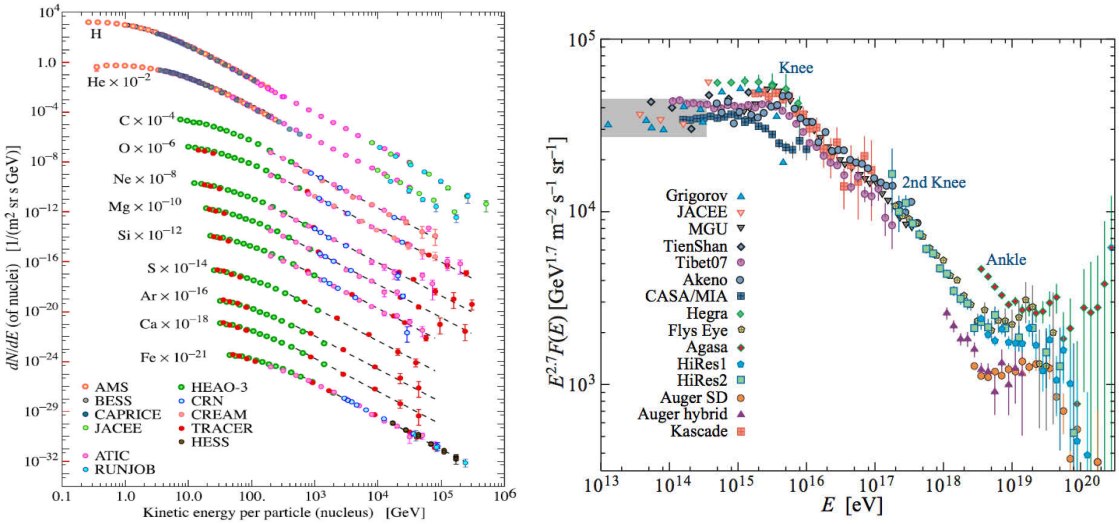


Figure 1.4: Observed flux of CRs [27]. *Left:* observed energy spectrum from GeV to PeV energies, for different nuclei as measured from several experiments. *Right:* observed all-particle (integrated over mass) energy spectrum from PeV to the highest energy, rescaled by a factor $E^{2.7}$ to put in evidence the Knee, the 2nd Knee and the Ankle (see the text for further information).

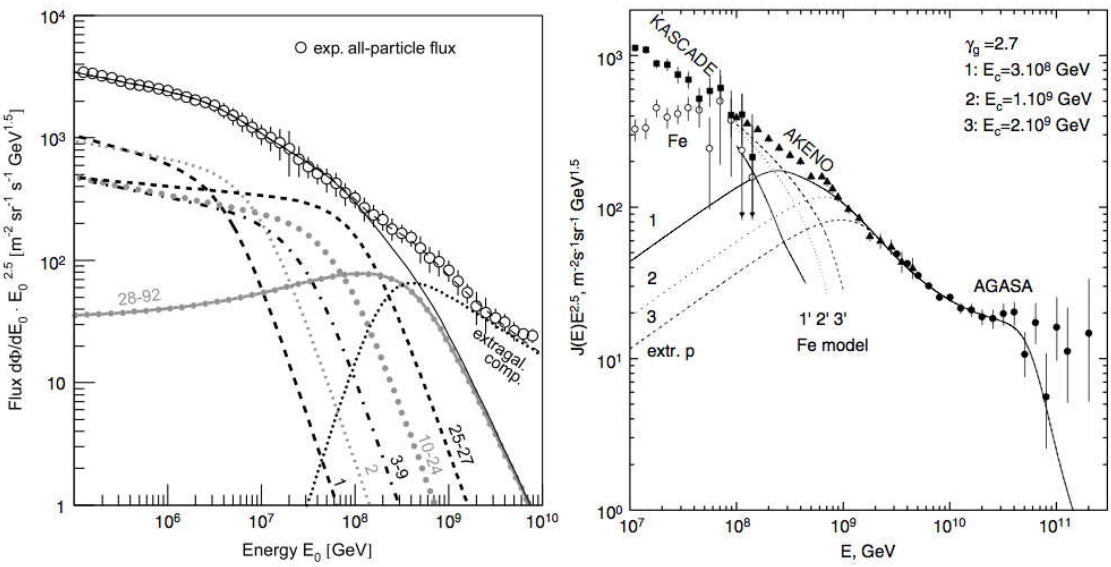


Figure 1.5: Observed flux of CRs from Ref. [26]. *Left:* Cosmic-ray energy spectra according to the poly-gonato model [31]. The spectra for groups of elements are labeled by their respective nuclear charge numbers. The sum of all elements yields the galactic all-particle spectrum (−) which is compared to the average measured flux. In addition, a hypothetical extragalactic component is shown to account for the observed all-particle flux (− − −). *Right:* Transition from galactic to extragalactic cosmic rays according to Berezinsky et al. [32]. Calculated spectra of extragalactic protons (curves 1, 2, 3) and of galactic iron nuclei (curves 1', 2', 3') are compared with the all-particle spectrum from the Akeno and AGASA experiments. KASCADE data are shown as filled squares for the all-particle flux and as open circles for the flux of iron nuclei.

for the flux up to the highest energy (see the right panel of the same figure). The transition, around 10^{18} eV, from galactic to extragalactic component may explain the ankle region, although many models have been proposed for such a purpose and a definitive answer is still missing, whereas the GZK effect could explain the suppression of the flux at the highest energy. When we will discuss the propagation

of UHECRs, we will present the so-called dip model, which attributes the ankle to propagation effects (i.e. the interaction of UHE protons and relic photons of the cosmic microwave background radiation), and we limit here to mention the recent model by Biermann and de Souza, able to explain to observed energy spectrum by assuming only a galactic origin of CRs and one single extragalactic source, namely Cen A [33].

The Ginzburg-Syrovatskii (GS) diffusion equation [34, 35], obtained by combining the Navier-Stokes and Maxwell equations for magnetohydrodynamics (MHD), describes the propagation of *galactic* cosmic rays in our Galaxy under quite general assumptions. The derived leaky box models (see for instance Ref. [36]), are able to account for the observed flux of CRs up to energies around PeV, as well as the observed dependence on the nuclear charge. The origin of higher energy CRs, thought to be of *extragalactic* origin, will be discussed further in this chapter.

In the present thesis work, we focus on the energy range corresponding to the end of the spectrum, and we mainly consider CRs in the trans-GZK and super-GZK regions (i.e. around and above 50 EeV, respectively). In fact, there are strong arguments in favor of the origin of such UHECRs in the nearby Universe. If these particles are light nuclei and the extragalactic magnetic field is weak enough, the astronomy with charged particles of energy above 50 EeV represents an exciting possibility, complementary to the astronomy with $\gamma-$rays and neutrinos (multi-messenger astronomy). The reasons for such a requirements will be discussed in the next chapters.

The modeling of the propagation of such UHECRs and the investigation of their intrinsic clustering, as well as their correlation with candidate astrophysical sources, represents the main subject of the present thesis.

1.3 UHECR acceleration mechanisms

The 1st order Fermi mechanism [37–39] can be used to explain the acceleration of particles in the UHE range. The simplest diffusive shock acceleration mechanism assumes i) a shock structure given *a priori*, which is not affected by the particles being accelerated, ii) a non-relativistically moving plane-parallel shock front, with magnetic field parallel to the shock normal, and that iii) inhomogeneities of the magnetic field can scatter particles efficiently so as to result in a nearly isotropic distribution of the particles. Under such assumptions, a universal power-law energy spectrum of the accelerated particles, $N(E) \propto E^{-s}$ emerges with spectral index $s = (r+2)/(r-1)$, being $r = v_2/v_1$ the shock compression ratio, with v_1 and v_2 the downstream and the upstream velocities of the fluid in the rest-frame of the shock, respectively. The shock-compression ratio r is related to the adiabatic index of the fluid. For typical astrophysical situations, one gets $r < 4$ and hence $s > 2$ [40]. Of course, such assumptions are not realistic, and many issues complicate this simple picture.

In relativistic shocks, the maximum energy at which particles can be accelerated depends on the size of the shock and on the magnetic field. Because we are working in the UHE regime, we can consider particles moving at the speed of light. A condition for such an extreme acceleration is that the Larmor radius of the particle $r_L = E/(ZecB)$, being Ze the charge of the particle, should be smaller than half the accelerator size R [41]. In an astrophysical shock, the acceleration of a particle of energy E is governed by the equation $dE/dt = E/T_{\rm acc}$, where $T_{\rm acc}$ is the energy dependent acceleration time. Moreover, the spectral index is given by $s = 1 + T_{\rm acc}/T_{\rm esc}$, the energy dependent average escape time. For 1st order Fermi acceleration at nonrelativistic shocks, the relation

$$T_{\rm acc} = \frac{3}{v_2 - v_1}\left(\frac{D_1}{v_1} + \frac{D_2}{v_2}\right) \tag{1.1}$$

holds, being D_1 and D_2 the diffusion coefficients corresponding to the two different regions of the stream. As we will see in the next chapter, inhomogeneities in the magnetic field are responsible for the magnetic

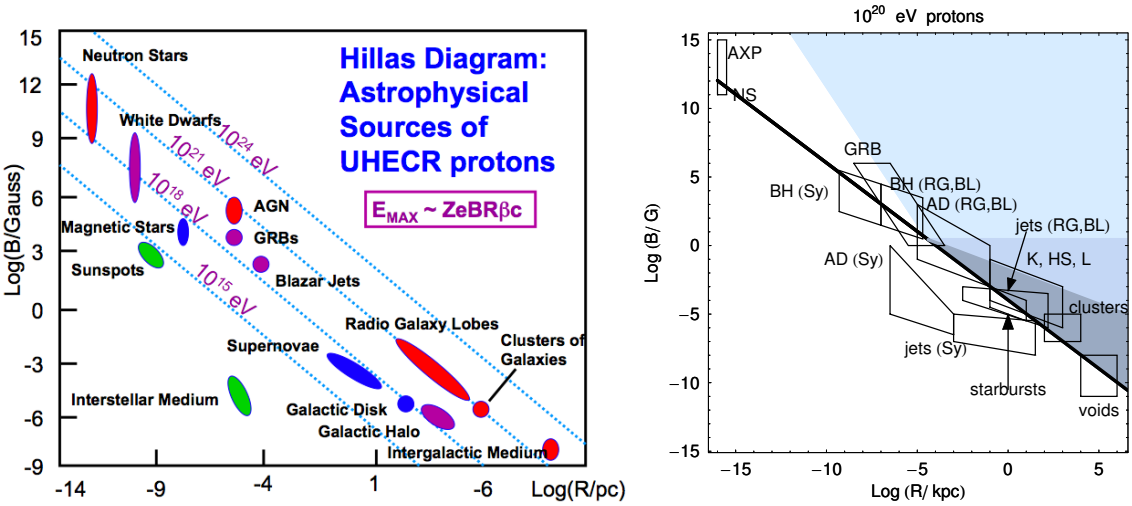


Figure 1.6: The so-called Hillas plot [41] (left) and its recent updated version [42] (right) showing the candidate sources of UHE protons as a function of their size and their magnetic field.

pitch angle scattering that causes diffusion. Such a behavior bounds the mean free path λ from below by the Larmor radius, multiplied by some factor g, from which $D_{1,2} \sim \lambda/3 \gtrsim gE/(3ZecB)$. Monte-Carlo simulations of such relativistic shocks show that g can be as large as 40, whereas it is ≈ 1 in the nonrelativistic case (see [40] and Refs. therein). By minimizing $T_{\rm acc}$ as a function of v_1 and v_2, the relation $T_{\rm acc} \gtrsim gE/(2.25\, ZecB)$ is obtained. Additionally, by requiring that the mean free path is smaller than the accelerator size, the escape time will be given by $T_{\rm esc} = R^2/c\lambda$. By using such information, and by imposing $s = 3$, after some algebra, we obtain the expression

$$E_{\rm max} \sim 10^{17}\ {\rm eV}\ Z \left(\frac{R}{\rm kpc}\right)\left(\frac{B}{\mu{\rm G}}\right) \tag{1.2}$$

for the maximum energy at the accelerator site. In Fig. 1.6, known as Hillas plot [41, 42], the candidate sources of UHE protons are shown as a function of their size and magnetic field. Such a plot can be built by a simple dimensional argument, with no regards for the precise acceleration mechanism. In any statistical acceleration mechanism, there must be a magnetic field B to keep the particles confined within the acceleration site. Thus, the size R of the acceleration region must be larger than the diameter of the orbit of the particle $\sim 2r_L$. Including the effect of the characteristic velocity βc of the magnetic scattering centers one gets the general condition

$$\left(\frac{R}{\rm kpc}\right)\left(\frac{B}{\mu{\rm G}}\right) > \frac{2}{Z\beta}\left(\frac{E}{10^{18}\ {\rm eV}}\right). \tag{1.3}$$

Among active galaxies, only the most powerful ones, radio galaxies and blazars, are able to accelerate protons to UHE, though acceleration of heavier nuclei is possible in much more abundant lower-power Seyfert galaxies [42]. A more detailed description of such candidate sources is provided in the following.

1.4 Candidate sources of UHECR

A complete treatment of the candidate sources of UHECR is beyond the goal of the present work. We will limit to briefly discuss the main candidates and their features: for a complete review about such an interesting topic, we refer to [43] and Refs. therein.

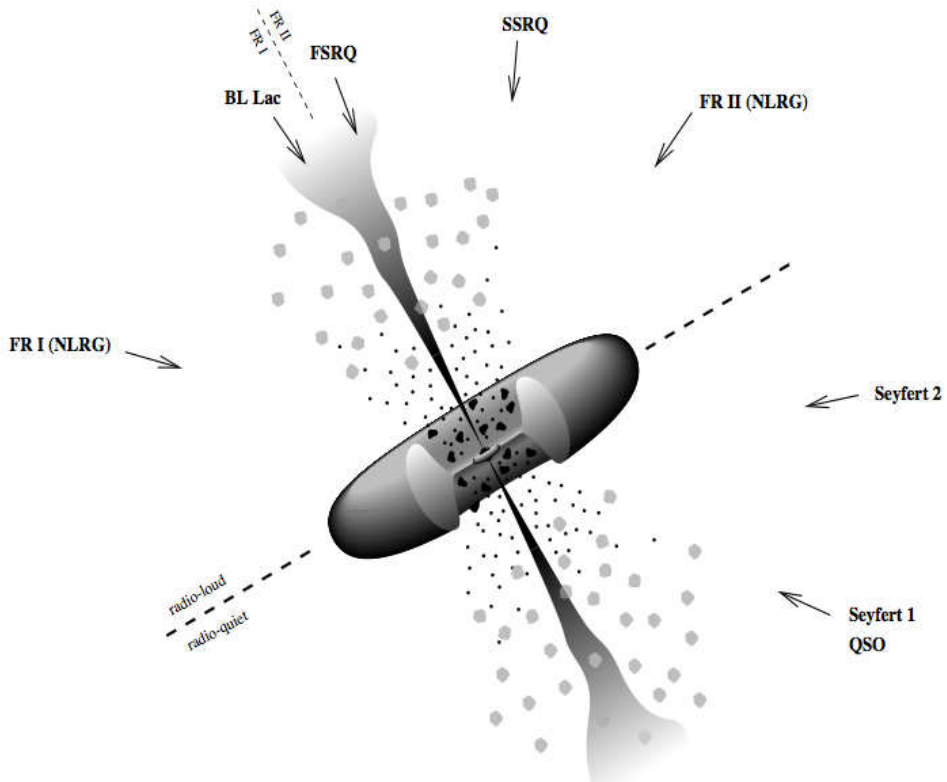


Figure 1.7: The unification model for AGN. Blazars are those AGN for which the jets are close to the line of sight. A regular quasar or a Seyfert 1 galaxy is observed if the orientation angle is $\sim 30°$, where the narrow-line and broad-line regions are visible. At larger angular offsets, the broad-line region will be hidden by the torus, the corresponding class being Seyfert 2 galaxies. Perpendicular to the jet axis, the full extent of the jets may be seen particular at low frequencies, giving rise to a morphology typical of radio galaxies [43].

Neutron stars

Neutron stars represent the final stage of stellar evolution, under certain constraints on the stellar mass (black holes are favorite at the other endpoint). They are characterized by fast rotation, with period of the order of ms, and large surface magnetic field ($B \gtrsim 10^{13}$ G). The acceleration process is magnetohydrodynamic, rather than stochastic as it is at astrophysical shocks. At the light cylinder $R_{lc} = c/\Omega \simeq 10^7 \Omega_{3k}^{-1}$ cm, being $\Omega_{3k} \equiv \Omega/3000$ rad s^{-1} and Ω the angular frequency. By assuming that the surface magnetic field is $B_S \equiv 10^{13} B_{13}$ G, the field at the light cylinder is $B_{lc} = 10^{10} B_{13} \Omega_{3k}^3$ G, from which the maximum energy of particles that can be contained in the wind near the light cylinder is

$$E_{\max} = \frac{ZeB_{lc}R_{lc}}{c} \simeq 8 \times 10^{20} Z_{26} B_{13} \Omega_{3k}^2 \text{ eV}, \tag{1.4}$$

where $Z_{26} \equiv Z/26$. A neutron star is able to accelerate particle up to energy of the order of 100 EeV only if its magnetic field is smaller than $10^{15.4}$ G and its period is smaller than 10 ms, not very restrictive values for a young neutron star [43].

Active galactic nuclei and radio galaxies

The unification model of AGN includes a large variety of astrophysical objects. These include blazars (characterized by strong flat spectrum radio emission, and/or significant optical polarization, and/or

significant flux variability in the optical and in other wavelengths), optically violently variable quasars (blazars with optical variability on short timescales), BL Lacertae (presenting a complete or nearly complete lack of emission lines) and highly polarized quasars.

AGN are galaxies characterized by a supermassive black hole ($10^6 - 10^{10} M_\odot$) at the center, which accretes galactic matter to form an accretion disk. Such a region is surrounded by an extended, dusty, molecular torus, and a corona of electrons populates the inner region generating continuum X-ray emission. Perpendicularly to the plane of the accretion disk, two-sided jets of relativistic particles are present. The unification model of AGN is shown in Fig. 1.7, also including Seyfert galaxies, Fanaroff-Riley objects (FRI and FRII), broad and narrow line galaxies and radio galaxies. In the particular case of radio galaxies, the most energetic protons injected in the intergalactic medium will have an energy

$$E_{\rm max} = 1.4 \times 10^{25} B_{\mu\rm G}^{-5/4} \beta_{\rm jet}^{3/2} u^{3/4} R_{\rm kpc}^{-1/2} (1 + Aa)^{-3/4} \text{ eV}, \tag{1.5}$$

being $B_{\mu\rm G} \equiv B/1\ \mu$G, $R_{\rm kpc} \equiv R/1$ kpc, a the ratio of photon to magnetic energy densities, A the measure of the relative strength of $p\gamma$ interactions versus the synchrotron emission, $\beta_{\rm jet}$ the jet velocity in units of c and u the ratio of turbulent to ambient magnetic energy density in the region of the shock. Cen A is the closest AGN to Earth, with magnetic field of the order of 50 μG and a hot spot of about 2 kpc: in principle, the maximum energy attainable for protons would be $\sim 10^{20.6}$ eV.

Quasars remnants

A spinning supermassive black hole, threaded by magnetic fields generated by currents flowing in a disc or torus, induces an electromotive force which accelerate a particle near the full voltage, if vacuum breakdown is prevented and in the absence of severe energy losses. At the present epoch the number of luminous quasars (with $\mathcal{L} \gtrsim 10^{47}$ erg s^{-1}) is small. Conversely, quasars appearing at large redshifts (the expected local number of dead quasars associated with the same parent population) is expected to be large. The term "quasars remnants" describes the present-epoch population of dead quasars harboring supermassive black hole nuclei. In the compact dynamo model, a poloidal magnetic field B near the hole produce a voltage $V \sim aB$, where a is the hole's specific angular momentum. If $\dot{M}$ is the accretion rate dM/dt in $M_\odot$ yr^{-1} units, the maximum electromotive force is $V \sim 1.2 \times 10^{21} \dot{M}^{1/2}$ V. After taking into account the energy loss rate through curvature radiation, the maximum acceleration energy is

$$E_{\rm max} = 3 \times 10^{19} \mu Z^{1/4} M_9^{1/2} B_4^{1/4} (\rho^2 h / R_g^3) \text{ eV}, \tag{1.6}$$

where ρ is the average curvature radius of an accelerating ion (assumed to be independent of the ion energy), h is the magnetic field curvature, $B_4 \equiv B/(10^4$ G), $M_9 \equiv M/(10^9\ M_\odot)$ and $R_g = GM$ is the gravitational radius. Hence, only a fraction of the potential energy available is released as UHECRs, while the rest is radiated in the form of curvature photons. For a magnetic field curvature of the order of Schwarzschild radius ($h \simeq 2R_g$) and $\dot{M} \approx (0.1 - 10) M_\odot$ yr^{-1}, the acceleration energy of a proton becomes

$$E_{\rm max} = (1 - 1.8) \times 10^{20} M_9^{1/4} \text{ eV}. \tag{1.7}$$

Heavier nuclei would reach higher energies, but are subject to photo-disintegration [43].

Gamma ray bursts

GRBs are short flashes of high energy radiation, brighter than any other gamma ray source in the sky. The bursts are characterized by a large variety of temporal profiles, spectra and timescales, whose origin is still poorly known. GRBs are the most energetic processes that we know about and Lorentz

factors needed to model their jet emission range from 100 to 1000, with timescales ranging from tens of ms up to hundreds of seconds. Generally, the existence of GRB is attributed to the dissipation of the kinetic energy of a relativistic expanding plasma wind, whose primal cause is not yet known (this models is known as "fireball", and we refer to [43] and Refs. therein for further detail). The rapid rise time and short duration of the burst require distant compact sources: the energy necessary to produce the observed events is $\approx 10^{52}$ erg of γ rays released in less than 1 second. However, fairly high current GRB activity is required to explain the observed UHECRs, while most of the GRBs with determined redshifts are at $z > 1$ (i.e. far than 4 Gpc).

Exotics

Up to here, we have described the so-called *bottom-up* mechanisms for the acceleration of CRs to the highest energies. In such mechanisms, particles have an initial energy small if compared to their final energy, reached after acceleration in very strong fields. However, different mechanisms have been proposed for the same purpose. To distinguish them from the bottom-up processes, they were called *top-down*. The basic idea of top-down mechanisms is that massive[2] $\mathcal{X}$ particles may decay to produce the observed UHECR. Such a process is allowed because the observed cosmic rays have energies several orders of magnitude lower than the $\mathcal{X}$ particle mass. These $\mathcal{X}$ particles can be either metastable (superheavy dark matter) or be emitted by topological defects (superconducting cosmic strings, monopoles, monopoles connected by a string) at the present epoch (see [40] and Refs. therein).

The emission of $\mathcal{X}$ particles by topological defects is possible by superconducting cosmic string loops, cusp evaporation in normal cosmic strings, and from intersecting cosmic strings. Successively, $\mathcal{X}$ particles decay in leptons and quarks, the latter hadronizing in baryons and mesons, that decay themselves along their decay chains into nucleons, $\gamma-$rays and neutrinos.

On the other hand, $\mathcal{X}$ particles could be remnants of the early Universe, with a very long lifetime, probably much larger than the age of the Universe. Because of their large mass, and under the hypothesis that they provide a significant fraction of dark matter, such particles would be gravitationally attracted to the Galaxy and to the local supercluster. There, their density could well exceed the average density in the Universe and a significant anisotropic distribution of UHECRs' arrival directions is expected.

At variance with bottom-up models, producing charged nuclei, top-down mechanisms produce neutrinos, $\gamma-$rays and only a small number of protons, with a power-law energy spectrum of index $\gamma = 1.5$. Hybrid models, which include elements of both groups, are also investigated. For instance, in the $Z-$burst model UHE neutrinos are generated in the Universe: they annihilate on the nonrelativistic relic antineutrinos (and vice versa) to produce the $Z-$boson, by $\nu_f + \overline{\nu}_f \longrightarrow Z$, with an enhanced, resonant cross section of $\mathcal{O}(G_F) \sim 10$ nb. The result of the resonant neutrino annihilation is a hadronic $Z-$burst 70% of the time, which contains, on average, thirty photons and 2.7 nucleons with energies near or above the GZK cutoff energy of 50 EeV. These photons and nucleons produced within our Supergalactic halo may easily propagate to Earth and initiate air showers above such an energy threshold [44].

1.5 Extensive air showers

Direct measurements of CRs are possible only up to an energy around 10^{15} eV. Above this energy, the number of particles per unit of time, area and solid angle, namely the *flux*, is very low. In order to increase the number of detected particles, it is necessary to increase both the size of detectors and

[2]The mass of such particles should be larger than 10^{12} GeV. For comparison, the GUT scale ranges from 10^{14} GeV to 10^{17} GeV, whereas the Planck scale is $\approx 1.22 \times 10^{19}$ GeV.

their observation time: ground based experiments are the only ones able to cover surfaces ranging from hundreds of m^2 to thousands of km^2.

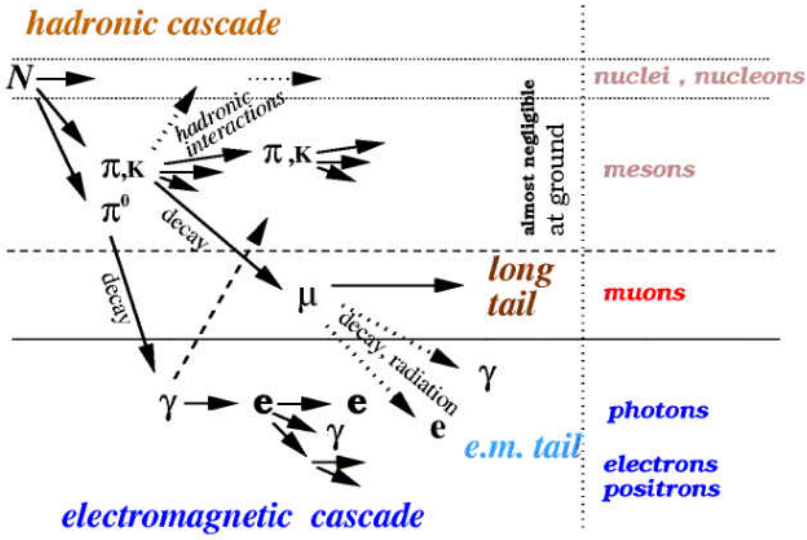


Figure 1.8: Schematic representation of an extensive air shower. A CR interacts with an atom or a molecule in the atmosphere and generates cascades of secondary particles, including nucleons, mesons, muons, electrons, positrons and photons at ground level.

Extensive air showers (EAS) are the particle cascades generated by the interaction of an UHECR with an atom or a molecule in the atmosphere, which acts like a calorimeter of variable density. In Fig. 1.8 we show a schematic representation of an extensive air shower. The cascades of secondary particles mainly develop through hadronic interactions, decays and energy-loss because of the emission of electromagnetic radiation. At ground level nucleons, mesons, muons, electrons, positrons and photons are generally observed. A common approach to the study of EAS is to consider three main components: i) electromagnetic, consisting of electrons, positrons and photons, ii) muonic, consisting of muons generated by the decay of charged mesons, and iii) hadronic, consisting of nucleons and mesons. For instance, a 10^{19} eV primary proton produces at sea level more than 10^{10} particles: 99% of such particles, with energy between 1 and 10 MeV, contributes to the electromagnetic component, transporting $\approx 85\%$ of total energy; the remaining 1% of particles are either muons, with an average energy of about 1 GeV (carrying about 10% of the total energy), few GeV pions (about 4% of the total energy) and, in smaller proportions, neutrinos and baryons [25]. A realistic modeling of EAS is rather difficult, because of the complex dynamics of interactions and decays to which the large number of secondary particle are subjected to. A simplified approach in the case of the electromagnetic cascade is due to Heitler [45], who had previously developed the theoretical background for the stopping of fast particles and positron creation with Bethe [46], and the passage of fast electrons with Bhabha [47]. Recently, a similar approach has been adopted to model hadronic showers [48]. In both cases, the main idea is to describe the cascade as a tree: the path from the sea level to the point where the cascade begins is divided into several slices. For electromagnetic cascades the length of each slice is equal to the radiation length, whereas in the case of hadronic cascades, it is equal to the interaction length. At each step, particles split their energy into two or more particles with smaller energy and this mechanism continues until a certain critical energy is reached. Such models are schematically represented in Fig. 1.9, and we will describe them in more detail, in the following.

Electromagnetic cascade. The electromagnetic cascade is originated by the decay of neutral mesons

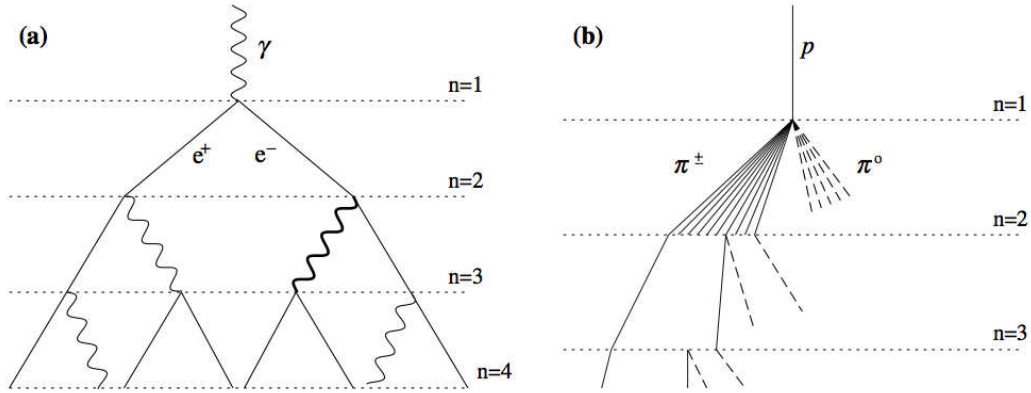


Figure 1.9: Heitler models for development of electromagnetic (a) and hadronic (b) cascades. Picture from [48].

producing high energy photons. The shower path is divided into n steps, and the interaction step-length is given by $d = \lambda_r \log 2$, being λ_r the radiation length λ_r of the medium ($\lambda_r = 37$ g/cm^2 in air). At each step all particles interact, producing two secondary particles of equal energy: each photon produces an electron/positron pair, whereas electrons and positrons lose half of their energy by emitting a photon because of bremsstrahlung. The cross sections of processes are considered as independent of energy and collision energy losses are totally ignored. If E_0 is the energy of the primary photon, after n steps a number $N_n = 2^n$ of secondary particles is expected, each one with energy $E_n = E_0/N_n$. The shower development continues until E_n becomes smaller than a critical energy ϵ_γ ($\approx$ 80 MeV in air), more precisely when the rate of energy loss by electrons via bremsstrahlung is equal to the rate of energy loss by ionization. The number of particles reaches a maximum value $N_{\max} = E_0/\epsilon_\gamma$ and increases no more until reaching the sea level. The atmospherical depth $X_{\max}$ at which $N_{\max}$ is reached depends on the first interaction point X_0, and is defined by

$$X_{\max} = X_0 + \lambda_r \log \frac{E_0}{\epsilon_\gamma}. \tag{1.8}$$

The *elongation rate*, defined as the rate of evolution of $X_{\max}$ with energy, is defined by

$$D_{10}^{\gamma} \equiv \frac{dX_{\max}}{d\log_{10} E_0} = 2.3\lambda_r, \tag{1.9}$$

that is $\approx$ 85 g/cm^2 in air. Monte Carlo simulations show that such predictions, obtained from a very simplified approach, qualitatively describe the electromagnetic shower development, with differences smaller than one order of magnitude.

Hadronic cascade. The hadronic cascade is originated by the decay of charged mesons producing high energy muons (and neutrinos). The approach is similar to that of the electromagnetic shower, although, in this case, the relevant parameter is the hadronic interaction length λ_h instead of λ_r. Another important difference is in the production of particles at each step: in fact, hadronic interactions produce multiple particles at each step. In this recent Heitler model for hadronic showers, it is assumed that $2N_\pi$ charged pions and N_π neutral ones are produced per interaction. Neutral pions contribute to the electromagnetic shower because of the decay $\pi^0 \longrightarrow \gamma\gamma$, whereas π^+ and π^- continue to interact until their decay rate is more likely than a new interaction (the critical energy in this case is $\epsilon_\pi \approx 20$ GeV in air). The decay of charged pions produce muons via $\pi \longrightarrow \mu\nu_\mu$, contributing to the muonic component of the EAS. In this case, the simplified model assume that i) interaction length and pion multiplicity do not depend on energy, ii) one third of the available energy goes into the electromagnetic component

while the remaining 2/3rd continues as hadrons [48]. It is worth remarking that the density profile of the atmosphere is not constant and the critical energy is larger high above ground than at see level and deep showers will produce fewer muons. The number of steps needed to reach the critical energy is simply given by $n_n = \log\left(E_0/\epsilon_\pi\right)/\log(3N_\pi)$.

By assuming that all pions decay into muons at the critical energy, the number of muons in the shower will be $N_\mu = (2N_\pi)^{n_c}$, or, equivalently, $N_\mu = (E_0/\epsilon_\pi)^\beta$, where $\beta = \log(2N_\pi)/\log(3N_\pi)$. For pion energy between 1 GeV and 10 TeV, $N_\pi = 5$ (hence, $\beta \simeq 0.85$) correctly describes the corresponding showers. At variance with the number of electrons in the electromagnetic shower, which grows linearly with the primary energy, the number of muons nonlinearly grows at a slower rate, strongly depending on the value of β. Such a parameter depends on the multiplicity and on the inelasticity of the hadronic interactions. Simulations show that the value of β ranges from 0.9 to 0.95, compatible with the assumption that only half of the available energy goes into the pions at each step (rather than all of it, as previously assumed) [49].

Moreover, at variance with the case of electromagnetic cascades, the position of shower maximum is more difficult to be determined. The larger cross section and the larger multiplicity at each step will reduce the value of $X_{\max}$ while the energy evolution of those quantities will modify the elongation rate. Hence, this simplified model only partially accounts for real observations. However, a more precise result can be obtained by mean of simulations. Assuming a proton-air cross section of 550 mb at 1 EeV and a rate of change of about 50 mb per decade of energy [50], the interaction length becomes

$$\lambda_h \simeq 90 - 9\log\left(\frac{E_0}{\mathrm{EeV}}\right)\ \mathrm{g/cm^2}. \tag{1.10}$$

If we assume that the first interaction initiates $2N_\pi$ electromagnetic cascades of energy $E_0/6N_\pi$, and $N_\pi \propto (E_0/\mathrm{PeV})^{1/5}$ for the evolution of the first interaction multiplicity with energy [48], the elongation rate reads

$$\begin{aligned} D_{10}^p &= \frac{dX_{\max}}{d\log_{10}E_0} = \frac{d\left(\lambda_h\log 2 + \lambda_r\log\left[E_0/(6N_\pi\epsilon_\gamma)\right]\right)}{d\log_{10}E_0} \\ &= \frac{4}{5}D_{10}^\gamma - 9\ln 2 \simeq 62\ \mathrm{g/cm^2}, \end{aligned} \tag{1.11}$$

in agreement with simulations. Such a value has been obtained under the hypothesis that the primary UHECR is a proton. The extension to heavier primary nuclei can be obtained by adopting the superposition model, where the nuclear interaction of a nucleus with atomic number A is simply viewed as the superposition of the interactions of A nucleons of individual energy E_0/A. It follows that showers from heavier nuclei develop faster than those from lightest ones. As a direct consequence, pions in the hadronic cascade reach their critical energy faster, increasing the relative number of muons with respect to the electromagnetic component. Hence, showers induced by nuclei will develop higher in the atmosphere, with $X_{\max}^A = X_{\max}^p - \lambda_r\log A$, and will have a larger number of muons, namely $N_\mu^A = N_\mu^p A^{1-\beta}$.

It is evident that the direct measurement of $X_{\max}$ and N_μ provides information about the mass of the primary cosmic ray. The predicted muon distributions depend on the assumptions on hadron production in air showers. This is evident in Fig. 1.10 where the expected number of muons and electrons is shown for showers initiated by proton, iron, and gamma-ray primaries, as calculated with different interaction models.

EAS detection and reconstruction

In order to infer energy, mass and charge of the primary particle, the EAS produced by a high energy cosmic ray is observed and the main features of the reconstructed shower are measured. The arrival

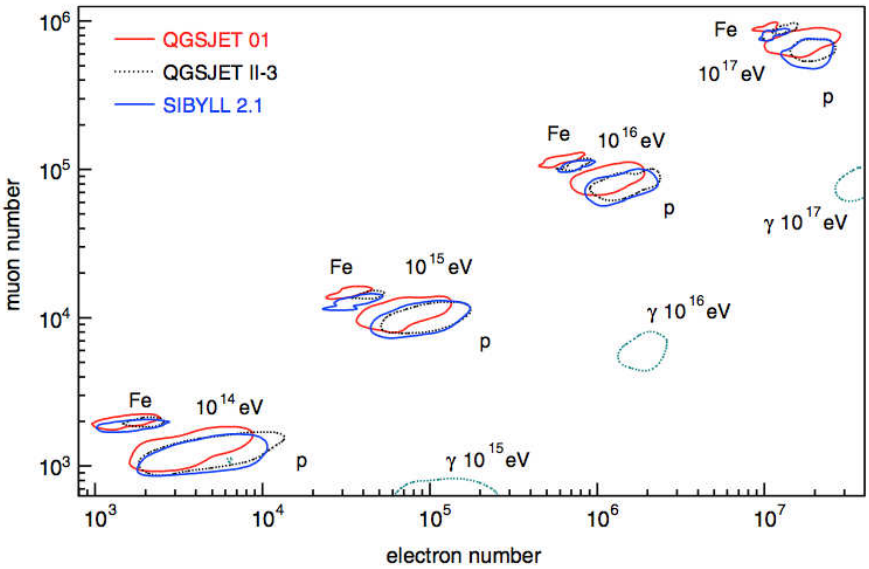


Figure 1.10: Expected number of muons and electrons in vertical showers, at sea level, originated by proton, iron and gamma-ray primaries. The curves show the full with at half maximum of the distributions for different primary particles and energies, as obtained with three different hadronic models [26].

direction of the primary CR generating the EAS is estimated by observing the shower development and, in particular, by reconstructing the *shower axis*. Energy can be estimated by measuring the density of secondary particles at ground level as a function of the distance from the shower axis, i.e. the *lateral distribution function* (LDF), or by measuring the number of particles as a function of the atmospherical depth, i.e. the *longitudinal development* (LD).

For such purposes, two main kind of detectors are used to detect CRs: i) the *surface detector* (SD), consisting of an array of several sensors (scintillators, water Cherenkov tanks, muon detectors, Cherenkov telescopes) spread over a large area; ii) the *fluorescence detector* (FD), for CRs with energy above 10^{17} eV, consisting of one or more telescopes measuring the fluorescence light emitted isotropically by nitrogen molecules in the atmosphere, adopted as a giant calorimeter, excited along the shower trajectory because of their interaction with the electromagnetic cascade.

Surface detector. Reconstruction of the primary particle parameters is based on timing for the geometry and on the distribution of signal densities as a function of the lateral distance to the shower axis for the energy. From the position of the different detectors and from the onset of the shower front signal recorded in each of them, one can reconstruct the shower axis and hence the original cosmic ray direction. Precision of one to three degrees are usually obtained given the large base line of the detector spacing (1 km). For the energy, the detector positions are projected onto the plane transverse to the shower axis and the lateral distribution function (LDF) is adjusted to the measured signals. In the case of UHECRs, the signal at an optimal distance $r_{\rm opt}$, depending on the energy range and the array spacing, can be used for such a purpose. At $r_{\rm opt}$ the sum of the fluctuations from shower to shower and of the statistical fluctuations from particle counting are minimum. Hence, the LDF varies for different experiments. In many cases, it behaves like

$$S(r,\theta,E) = kr^{[\beta(\theta,E)+r/r_0]}, \tag{1.12}$$

where r is in meters, θ is the zenith angle and $\beta(\theta,E) = a + b\sec\theta$. Once the attenuation of the signal due to the zenith angle is accounted for, an estimator of the energy is obtained by $E \propto S(r_{\rm opt},\theta_{\rm ref})^{\alpha}$,

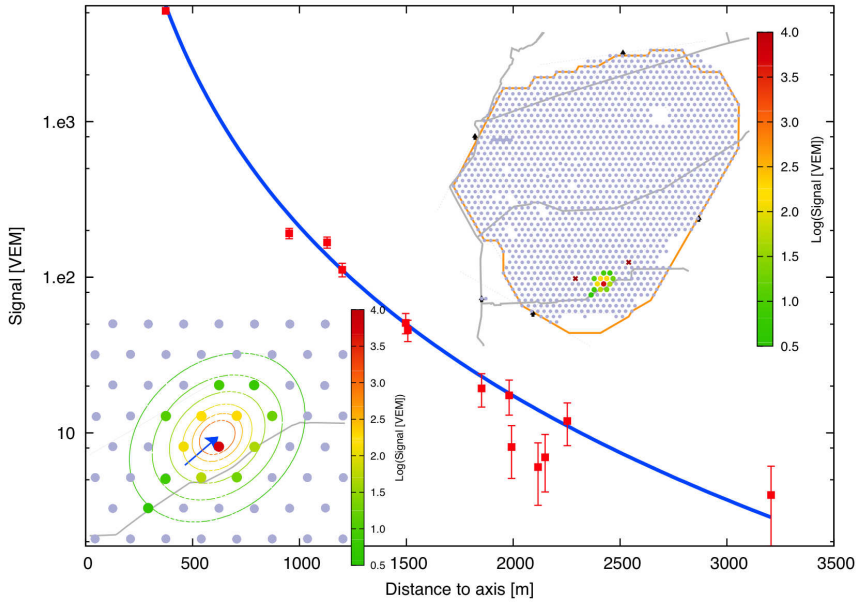


Figure 1.11: Example of EAS detection using a surface array. The upper right insert shows the Auger surface array. Each dot representes a detector the spacing between them is 1.5 km. The lower insert shows details of the footprint of the shower with the estimated contours of the particle density levels. The curve represents the adjusted LDF (lateral distribution function) and the red point the measured densities as a function of the distance to the shower core [25].

with $\alpha \approx 1$, from the corrected density at $r_{\rm opt}$. In Fig. 1.11 we show an example of detection using a surface array.

To reconstruct the primary parameters, a minimum of three detectors with signal is necessary. The spacing between those detectors will determine the array energy threshold. For a vertical shower the 500 m spacing of the trigger stations in Haverah Park corresponds to a threshold of a few 10^{16} eV, while the 1.5 km separation of the Pierre Auger Observatory stations gives a few 10^{18} eV. Unfortunately, the SD does not provide information about $X_{\rm max}$, and this is a strong limitation to this technique for primary identification, which requires the use of Monte Carlo simulations. For a better primary identification, the number of muons can be measured with buried detectors.

Fluorescence detector. A direct measurement of $X_{\rm max}$ can be done with the fluorescence technique, exploiting the light emitted between 300 and 420 nm by nitrogen molecules excited by charged secondary particles in the EAS. Such fluorescence light is emitted isotropically and can be detected with appropriate telescopes. The fluorescence yield is about 4 photons per electron per meter at ground level pressure. Under clear moonless night conditions, using square-meter scale telescopes and sensitive photodetectors, the UV emission from the highest energy air showers can be observed at distances in excess of 20 km from the shower axis. This represents about two attenuation lengths in a standard desert atmosphere at ground level. Such a large aperture, instrumented from a single site, makes this technique a very attractive alternative to ground arrays despite a duty cycle of about 10%.

Fluorescence photons reach the telescopes in a direct line from their source. Thus the collected image reflects exactly the development of the electromagnetic cascade. From the fluorescence profile it is in principle straightforward to obtain the position of the shower maximum and a calorimetric estimate of the primary energy. In practice a number of corrections must be made to account for the scattering and the absorption of the fluorescence light. Also pollution from other sources such as the Cherenkov component which can be emitted directly, or diffused by the atmosphere into the telescope, must be

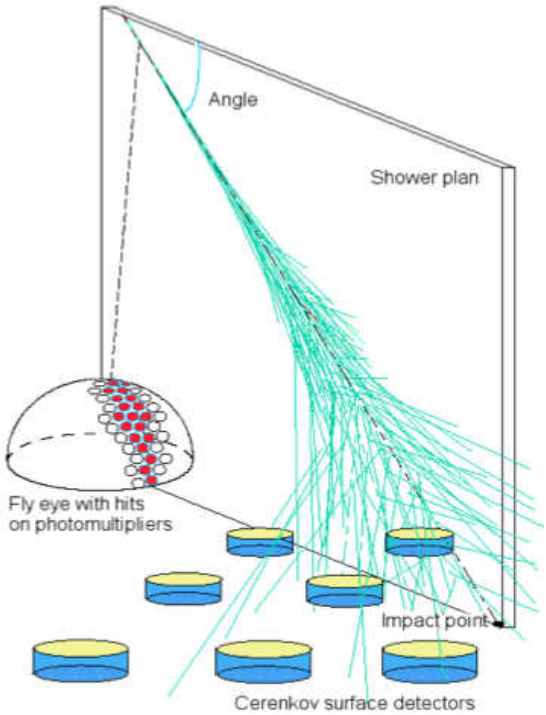


Figure 1.12: Scheme of a hybrid detector of UHECRs. The extended air shower is observed with both surface and fluorescence detectors, improving the reconstruction of the original event.

carefully evaluated and accounted for. A constant monitoring of the atmosphere and of its optical quality is necessary together with a precise knowledge of the shower geometry for a careful account for those corrections. Unfortunately, a single telescope is not sufficient to provide a satisfactory angular resolution on the arrival direction of the primary CR. A standard improvement involves the use of a second telescope, providing, from a different position, a stereo reconstruction of the shower with a precision of the order of few degrees. Such a technique has been adopted, for instance, by the HiRes detector [51].

The amount of fluorescence light emitted along the shower axis is proportional to the number of electrons in the shower. Hence, the total energy of the shower can be estimated by the integral of the longitudinal development (LD), knowing that the average energy loss per particle is $\approx$2.2 MeV/g cm^{-2}. The LD is generally parameterized by the Gaisser-Hillas function [52], depending on 4 parameters:

$$N(X) = N_{\max} \left(\frac{X - X_0}{X_{\max} - X_0} \right)^{(X_{\max} - X_0)/\lambda} \exp \left[\frac{X_{\max} - X}{\lambda} \right], \tag{1.13}$$

being X_0 the depth of first interaction. However, several corrections should be taken into account, such as the missing energy carried by neutrinos, the hadrons interacting with nuclei (whose energy is not converted into fluorescence) and penetrating muons, whose energy is mostly dumped into the Earth. Such a missing energy correction is calculated using detailed simulations and varies with energy, composition and the interaction model used. Additionally, the calibration of telescopes should be done frequently, in order to avoid a wrong estimation of the measured fluorescence light.

Hybrid detector. The two detection techniques described so far are somehow complementary. The SD has the main advantage of allowing a duty cycle of almost 100%, whereas the FD may operate only if the background light (human-made, moon, etc) is negligible. As we have seen the flux of CRs

rapidly decreases for increasing energy. Hence, a SD of a given size, whose sensors are separated by a fixed distance D each other, is able to detect only a fraction of EASs, i.e. those initiated by primary CRs with energy ranging between a minimum and a maximum value: such energy thresholds fix the operational limits of the SD. However, the maximum attainable energy can be increased by increasing the size of the SD, whereas the minimum threshold can be decreased by increasing the number of sensors and reducing their distance. In the case of the FD, the atmospherical depth of the shower maximum increases with energy: in order to vary the operational limits in energy, it is necessary to vary the inclination of telescopes.

Only a combination of both techniques should guarantee precise measurements of the primary arrival direction, its energy and the shower maximum. The Pierre Auger Observatory has been the first experiment to adopt both techniques in a hybrid detector (see Fig. 1.12), taking advantage from both approaches. It will be discussed in more detail in the next section[3].

1.6 UHECR detection: the Pierre Auger Observatory

The first evidence for the existence of UHECR has been provided by Volcano Ranch (USA), the pioneering array of surface detectors operated by a group from MIT, under the leadership of John Linsley and Livio Scarsi [53]. Since then, many other fundamental experiments have been built around the world to observe UHECR, with energy and angular resolution increasing over time. Here, we mention only a few among them, as Haverah Park (UK) [54], Yakutsk (Russia) [55], Fly's eye (USA) [56], SUGAR (Australia) [57], AGASA (Japan) [58] and Hires (USA) [51]. All of such experiments, together with the most recent ones, as the Pierre Auger Observatory (Argentina) [59] and Telescope Array (USA) [60], and the upcoming ones, as Auger North [61] or JEM-EUSO (outer space) [62], allowed (and will allow) us to explore energy regions well beyond those ones available to the most powerful man-made particle accelerator (i.e. the large hadron collider at CERN, Switzerland).

However, although many experimental efforts have been made to explore CRs in the ultra-high energy region, a conclusive answer to their origin, their acceleration mechanism and their mass composition is still missing. In particular, at the end of the 20st century, the two largest UHECR experiments, namely the Akeno Giant Air Shower Array (AGASA) and the High Resolution Fly's Eye (HiRes), provided somehow contradicting measurements of both the energy spectrum and the clustering of the arrival directions.

In the previous sections we have mentioned the longstanding hypothesis which predicts that the flux of the highest energy cosmic rays should be suppressed above $10^{19.6}$ eV, as cosmic rays from distant sources will interact with the cosmic microwave background via photo-pion production until their energy drops below this threshold energy. Such a mechanism is responsible for the so-called GZK suppression. Unfortunately, the detection of such an effect through indirect measurements is not unambiguous, because in fact several cosmic rays with higher energy have been observed over the years. One of the first results, at the end of 90s, disfavoring the GZK hypothesis comes from the AGASA collaboration, whose published energy spectrum showed no indication of a high-energy suppression [63], reporting 11 events above 10^{20} eV. In contrast, the monocular-mode energy spectra measured in the same years by the High Resolution Fly's Eye detectors supported the existence of a GZK feature, with an evident suppression at the highest energy [64, 65].

Another major source of uncertainty about the understanding of UHECRs came from the anisotropy and clustering measurements of the highest energy events, reported by the AGASA collaboration at 10^{18} and 10^{19} eV [66, 67], but not confirmed by the HiRes collaboration [68–70].

[3]This section and the next one have been largely inspired to the two recent and nice reviews about UHECR physics, namely [26] and [25].

The AGASA and HiRes, representing the second generation of UHECR experiments, used radically different techniques to determine the primary energy (AGASA was a ground array and HiRes an air fluorescence detector), and the statistical and systematic errors had fundamentally different sources. Although it is difficult to assign a significance to the discrepancies between the two energy spectra, a direct comparison between the two measurements put in evidence the need for a new independent experiment, able to measure the energy spectrum with the highest achievable accuracy. Moreover, such an ideal experiment should had to be able to measure the arrival direction of UHECRs with the best angular resolution ever, in order to shed light on the possibility of astronomy with charged particles.

The project for such a new experiment was proposed by Jim Cronin and Alan Watson in 1992. The new observatory should extend on a considerable surface and should make use of an hybrid technique, combining the best features of a giant surface array and several different fluorescence detectors. In 2004, the Pierre Auger Observatory, the largest experiment ever built, started data taking, inaugurating the third generation of UHECR detectors.

In this section, we will describe the Pierre Auger Observatory, giving particular attention to the selection and the reconstruction of events observed with the surface detector only, because of interest for the studies presented within this work. However, we will also briefly describe the detection and the reconstruction of events by using the fluorescence detector and the hybrid technique.

1.6.1 The detector

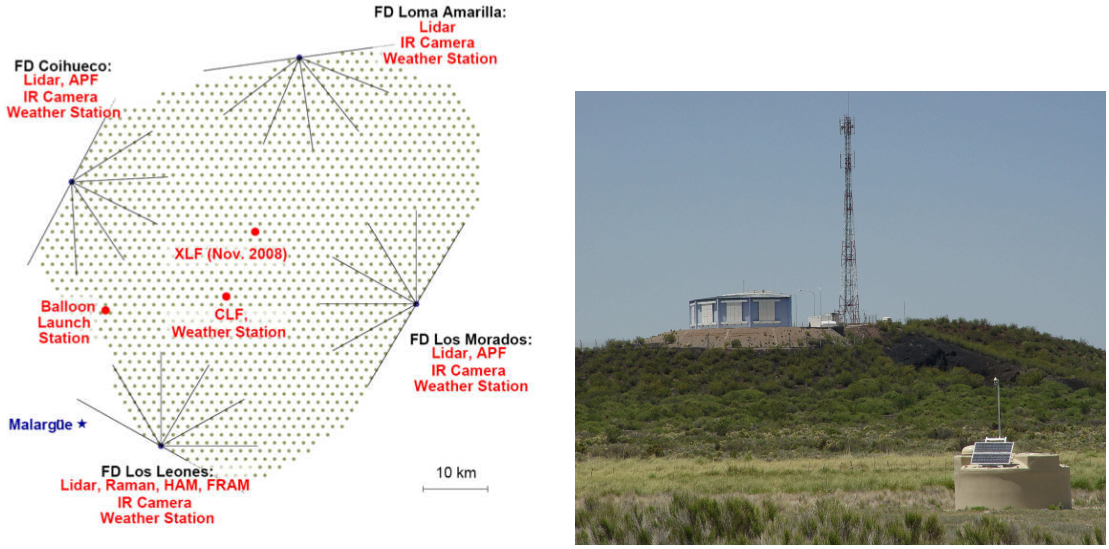


Figure 1.13: **Left panel:** General layout of the Pierre Auger Observatory, extending over 3000 km^2. It consists of an SD array of 1600 water-Cherenkov stations and 4 FDs, each with 6 telescopes. **Right panel:** particular of a FD building and a SD station.

The Pierre Auger Observatory is located at 35.2 S latitude and 69.5 W longitude at 1400 m above sea level, corresponding to the atmospheric depth of $X = 880$ g/cm^2. It is an hybrid detector, i.e. it consists of both a surface (SD) [71] and fluorescence (FD) [72] detector, as shown in Fig. 1.13. The monitoring of atmospherical conditions and cloud coverage, playing a central role in the data analysis, is employed with weather stations placed close to each FD and to the central laser facility (CLF), and with Lidars, whose position is shown in the same figure [73, 74].

The SD consists of 1660 water-Cherenkov stations that detect photons and charged particles in air

showers at ground level. Each station is a plastic tank of 10 m^2 area and 1.2 m height hosting 12000 l of ultra purified water contained in a Tyvek reflective liner. The Cherenkov light generated by charged particles of the shower crossing the tank is registered by three 9" photomultiplier tubes. Each detector operates autonomously with its own front-end electronics and communication system, while a solar energy unit and batteries provide the required 10 W electrical power. The timing is obtained by a GPS unit and the communications achieved via a wireless system. The station electronics is made up of the front-end electronics (FE), a station controller composed of a microprocessor performing local software processing, a slow control module, a GPS receiver and a time tagging unit. On the FE-board, the PMT signals are filtered and fed to a 10-bit ADC which samples at 40 MHz: the signal recorded by the FADC is referred to in units of channels (ch), with a range of 0-1023, corresponding to an input range of 0-2 V. Each FADC bin corresponds to 25 ns. The triangular grids of stations with 1.5 km spacing spans over 3000 km^2, and operates with a duty cycle of almost 100%.

The surface array is overlooked by 4 buildings (called Los Leones, Los Morados, Loma Amarilla and Coihueco) of the fluorescence detector. Each of them includes 6 fluorescence telescopes that observe the longitudinal profile of the shower. Each fluorescence telescope records the fluorescence light isotropically emitted by the nitrogen molecules of the atmosphere excited by the charged particles during the shower development in air. It consists of a Schmidt optical system with a circular diaphragm, positioned at the centre of curvature of a spherical mirror, that defines the aperture of the system. UV transmitting filters are installed in the entrance aperture and just inside it a ring of corrector lenses is placed to increase the effective aperture by about a factor 2 while preserving the quality of the optical aberrations. The light is focused by a large spherical mirror onto a camera of 440 phototubes which allows the 30° azimuth times 28.6° elevation field of view [75]. Since the FD can only operate during clear and moonless nights, its duty cycle is reduced to about 13%.

1.6.2 Event selection of the surface detector array

The SD array, completed in June 2008, is a regular grid of stations continuously taking data since January 2004, whose operational status is monitored every second. The event selection is based on a hierarchical series of conditions, three online triggers (called T1, T2 and T3) and two offline event selection criteria (called T4 and T5).

For each station two levels of local trigger are defined (the *first level* or T1 and the *second level* or T2) both resulting from the OR condition of two different algorithm ("Simple Threshold" and "Time-Over-Threshold", or ToT). At the T2 level the expected rate of random stations is reduced to 1-2 Hz with respect to about 100 Hz of the T1. The trigger strategy and hierarchy are described in details in Ref. [76]. A SD trigger is formed when at least 3 not-aligned T2 stations are in spatial and temporal coincidence (T3). Due to this requirement the SD array is fully efficient for the detection of air showers above $10^{18.5}$ eV ($\approx$ 3 EeV), for all zenith angles between 0° and 60°, independently of the position of the impact point and of the mass of the primary particle.

Once a T3 is formed, all FADC signals from detectors passing the T2 are sent to the central data acquisition system (CDAS), as well as those from detectors passing the T1 but not the T2, provided that they are within 30 μs of the T3. The trigger of the array is realised in two modes. The first T3 mode requires the coincidence of at least three detectors that have passed the ToT condition and that meet the requirement of a minimum of compactness, namely, one of the detectors must have one of its closest neighbours and one of its second closest neighbours triggered. It is called "$ToT2C_1\&3C_2$", where C_n indicates the nth set of neighbours (see Fig. 1.14). Once the spatial coincidence is verified, timing criteria are imposed: each T2 must be within $(6+5C_n)\mu$s of the first one. An example of such T3 configuration is shown in Fig. 1.14, left. Since the ToT as a local trigger has very low background, this trigger selects predominantly physics events. The rate of this T3 with the full array in operation is around 1600 events

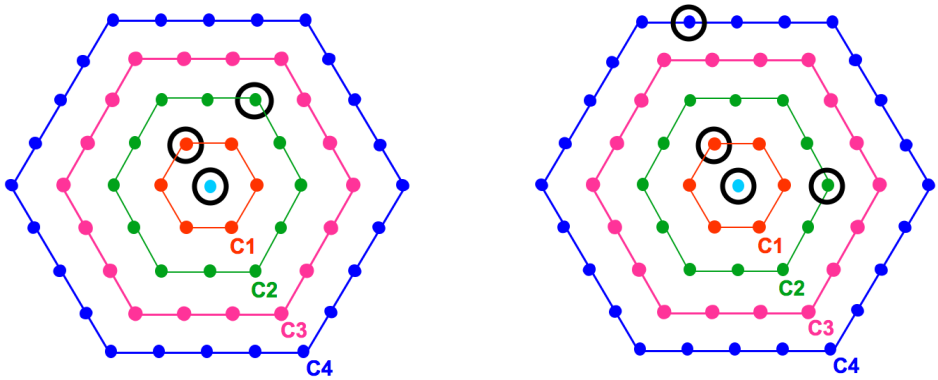


Figure 1.14: Example of T3 configurations: the 3-fold T3 mode $ToT2C_1\&3C_2$ is shown on the left and the 4-fold mode $2C_1\&3C_2\&4C_4$ on the right (see text for the definitions). C1, C2, C3, C4 indicate the first, second, third and fourth sets of neighbours, respectively at 1.5, 3, 4.5 and 6 km from a given detector.

per day, meaning that each detector participates in an event about 3 times per day. This trigger is extremely pure since 90% of the selected events are real showers and it is mostly efficient for showers below 60°. The 10% remaining are caused by chance coincidences due to the permissive timing criteria. The second T3 mode is more permissive. It requires a four-fold coincidence of any T2 with a moderate compactness. Namely, among the four fired detectors, within appropriate time windows, at least one must be in the first set of neighbours from a selected station (C_1), another one must be in the second set (C_2) and the last one can be as far as in the fourth set (C_4). This trigger is called "$2C_1\&3C_2\&4C_4$". Concerning timing criteria, we apply the same logic as for the "$ToT2C_1\&3C_2$". An example of such T3 configuration, is shown in Fig. 1.14, right. Such a trigger is efficient for the detection of horizontal showers that, being rich in muons, generate in the detectors signals that have a narrow time spread, with triggered detectors having wide-spread patterns on the ground. With the full array configuration, this trigger selects about 1200 events per day, out of which about 10% are real showers [76].

Two successive levels of selection are implemented. The first one (physics trigger) is based on space and time configurations of the detector, besides taking into account the kind of trigger in each of them. The second one (fiducial trigger) requires that the shower selected by the physics trigger is contained within the array boundaries, to guarantee the accuracy of the event reconstruction both in terms of arrival direction and energy determination. The logic of this off-line trigger system and its connection to the DAQ triggers is summarized in Fig. 1.15.

Physics trigger

The physics trigger, **T4**, is needed to select real showers from the set of stored T3 data. Two criteria are defined, with different aims. The first T4 criterion, so-called 3ToT, requires 3 nearby stations, passing the T2-ToT, in a triangular pattern. It requires additionally that the times of the signals in the 3 stations fit to a plane shower front moving at the speed of the light. The number of chance coincidence passing the 3 ToT condition over the full array is less than one per day, thanks to the very low rate of the T2-ToT. Due to their compactness, events with zenith angles below 60°are selected with high efficiency, i. e. more than 98%.

The second T4 criterion, so called 4C1, requires 4 nearby stations, with no condition on the kind of T2. In this case also, it is required that the times of the signals in the 4 stations fit to a plane shower front moving at the speed of the light. This 4C1 trigger brings to ≈100% the efficiency for showers below 60°.

Besides disentangling accidental events, there is also the need to identify, and reject, accidental

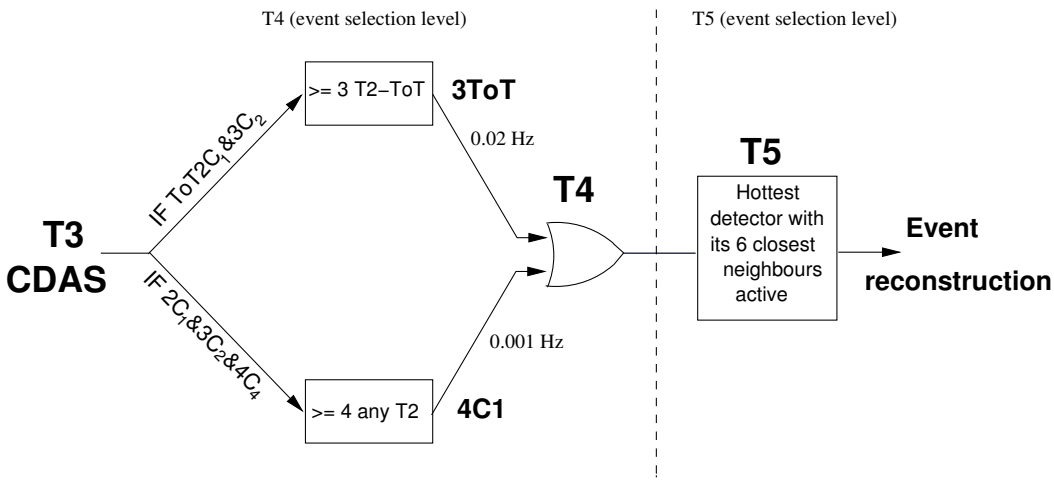


Figure 1.15: Schematics of the hierarchy of the event selection of the Auger surface detector.

detectors in real events, i.e. detectors whose signals are by chance in time with the others, but that in fact are not part of the event. To this aim, we define a "seed" made by 3 neighbouring detectors in a non-aligned configuration. If there is more than one triangle of stations, the seed with the highest total signal is chosen. If the T4 is a 3 ToT, only ToT detectors can be considered to define the seed; if it is a 4C1, also TH detectors can be included. Once the triangle has been determined, the arrival direction is estimated by fitting the arrival times of the signals to a plane shower front moving with the speed of light. Subsequently, all other detectors are examined, and are defined as accidental if their time delay with respect to the front plane is outside a time window of $[-2\,\mu\text{s} : +1\,\mu\text{s}]$. Detectors that have no triggered neighbours within 3 km are always removed.

After the selection chain (both event selection and accidental detectors removal), 99.9% of the selected events pass the full reconstruction procedure, that is arrival direction, core position and $S(1000)$ are determined.

Fiducial trigger

The need for a *fiducial trigger*, **T5**, mainly arises from events falling close to the border of the array, where a part of the shower may be missing. In Fig. 1.16 a hybrid event is shown, that triggered the SD and one of the FD telescopes, where a part of the SD information is missing due to its position on the border of the array. Such events could have wrong core positions, and consequently, incorrect energies, as in this example where the energy derived by SD is more than 4 times larger than the one estimated by FD (1.4×10^{19} eV instead of 3×10^{18} eV). The main task of the fiducial trigger is thus to select only events well contained in the array, ensuring that the shower core is properly reconstructed.

The fiducial trigger should be applied a priori on the events, to be independent of the reconstruction procedure. The T5 adopted requires that the detector with the highest signal has all its 6 closest neighbours working at the time of the event (i.e., it must be surrounded by a working hexagon). This ensures adequate containment of the event inside the array. Even in the case of a high energy event that falls inside, but close to the border of the array, where part of the data may be missing, information from the seven detectors closest to the shower core ensures a proper reconstruction. Applying this condition, the maximum statistical uncertainty in the reconstructed $S(1000)$ (defined in the next section) due to event sampling by the array is $\approx 3\%$ [76]. It has to be noted that this criterion also discards events that, though contained, fall close to a non-working detector: this is an important issue because, due to the large number of detectors distributed over 3000 km^2, about 1% of the detectors are expected to be not

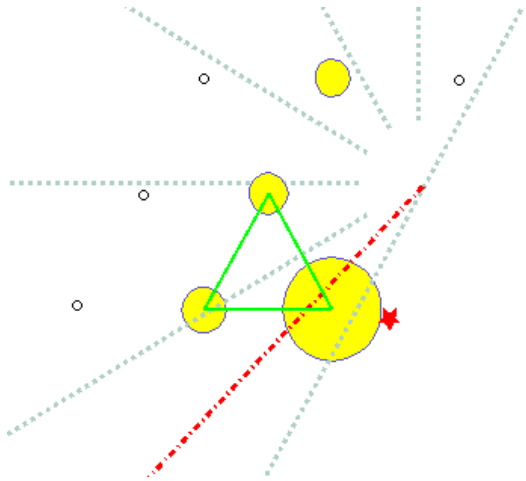


Figure 1.16: Example of a hybrid, non-T5, event: the event falls on the border of the SD array, triggering only four detectors. Filled circles indicate the triggered ones, open circles the non-triggered active ones. The dimensions of the filled circles are proportional to the measured signal. The shower detector plane reconstructed by FD (dash-dotted line) indicates that the core is within the triangle of detectors. The SD only reconstruction places it outside the array (cross), artificially increasing the event energy.

functioning at any moment, even with constant detector maintenance. For the fully completed array, and taking this into account, the application of the T5 condition reduces the effective area by 10% with respect to the nominal one.

The use of the fiducial trigger allows the effective area of the array to saturate to the geometrical one above a certain primary energy. Indeed, with no conditions on event containment, the acceptance would increase with increasing energy, since showers falling outside the borders of the array might still trigger sufficient detectors to be recorded; the higher their energy, the farther the distance [76].

The officially adopted T5 (6T5) requires that the station with the highest signal is surrounded by a working hexagon (i.e. the six closest neighbor station must be working at the time of the event). Such a trigger is rather conservative, ensuring that the maximum statistical uncertainty in the reconstructed $S(1000)$ due to event sampling by the array is 10%. Moreover, the 6T5 eliminates from the data set a large fraction (of the order of 25%) of events passing the T4 requirement: such events are on the border of the array of falls close to a un-operating station.

However, for events with very high energy, the 6T5 criterion is too strict. In fact, the high multiplicity of such events provide enough information for a reliable reconstruction in the majority of cases. A more relaxed criterion can be adopted in these cases, of interest for the studies presented in the last chapter of the present thesis. Such a new criterion, called 5T5-Pos (5T5 in the following), make use of the reconstructed shower core, requiring that 5 of the 6 closest neighbors are active at the time of the event and that its core is contained in an equilateral triangle of active stations.

Integrated exposure of the surface detector

The full efficiency of the SD trigger and event selection is reached at 3×10^{18} eV. Above this energy, the calculation of the exposure is based solely on the determination of the geometrical aperture and of the observation time.

With respect to the aperture, the choice of a fiducial trigger based on hexagons, as explained above, allows us to exploit the regularity of the array very simply. The aperture of the array is obtained as a multiple of the aperture of an elemental hexagon cell, a_{cell}, defined as any active detector with six active neighbours, as shown in Fig. 1.17.

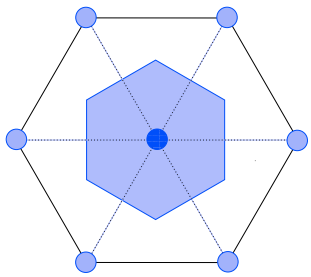


Figure 1.17: Scheme of an hexagon of detectors: the elemental hexagon cell, a_{cell}, is the shaded area around the central detector.

At full efficiency, the detection area per cell is $1.95\,\text{km}^2$. The corresponding aperture for showers with $\theta < 60°$ is then $a_{\text{cell}} \simeq 4.59\,\text{km}^2\,\text{sr}$. The number of cells, $N_{\text{cell}}(t)$, is not constant over time due to temporary problems at the detectors (e. g. failures of electronics, power supply, communication system, etc...). $N_{\text{cell}}(t)$ is monitored second by second. The second-by-second monitoring provides at the same time the aperture of the array per second, $a_{\text{cell}} \times N_{\text{cell}}(t)$, as well as the observation time with high precision. To calculate the integrated exposure over a given period of time, the aperture of the array, $N_{cell}(t) \times a_{cell}$, is integrated over the number of live seconds. This calculation is very precise, since it is based on a purely geometrical aperture and a very good time precision. The uncertainty is 3%. The standard units for the exposure of large detectors as the SD of the Pierre Auger Observatory are km^2 sr yr (sometimes also reported as 1 Linsley). The exposure of some experiments, past and current, as well of future ones, is shown in Fig. 1.18.

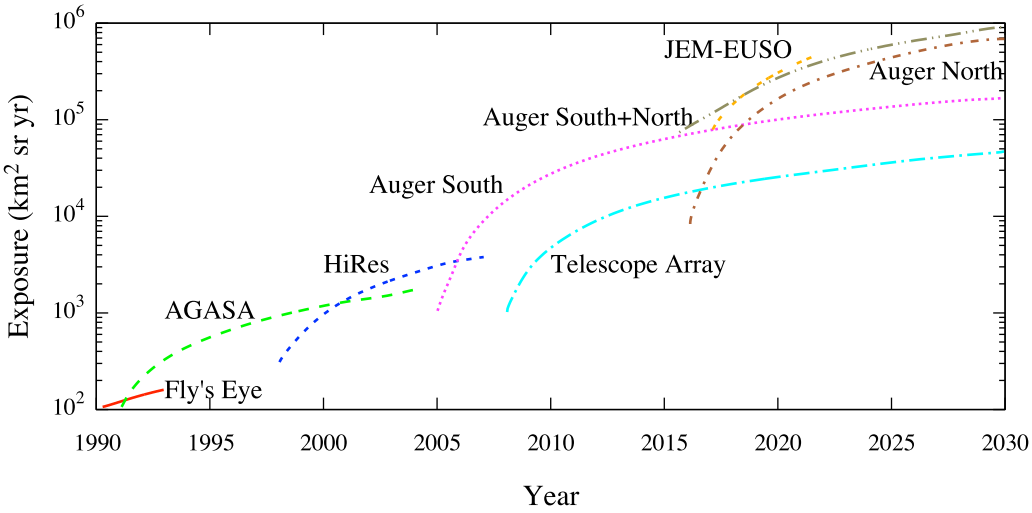


Figure 1.18: Evolution of the exposures of past and current UHECR observatories over time, compared to that of the Pierre Auger Observatory (south). Projected exposures for Auger North (if construction start in 2016) and JEM-EUSO (if launched in 2017, including 20% duty cycle) are also shown. This plot has been readapted from [77] (see Refs. therein for details about the exposure of each observatory).

For the data set considered in our study, the 5T5 exposure is about 20% larger than the 6T5 one.

1.6.3 Event reconstruction

SD event reconstruction

The signal-weighted and time-weighted barycenters of the stations are set as the distance and time origins. A shower track is described by the following equation:

$$-\widehat{a}(\vec{x}(t)-\vec{b}) = c(t-t_0) \tag{1.14}$$

where $\vec{x}(t)$ is a point moving with the speed of light along the straight shower track of normalized axis $\widehat{a}$ ≡(u,v,w), passing the origin at time t_0. The shower plane at this stage is treated as a plane perpendicular to the shower axis (approximate description), therefore the time $t(\vec{x})$ when it passes through a given point at ground $\vec{x}$ ≡(x_i,y_i,z_i) is obtained by projecting to the shower axis:

$$ct(\vec{x}) = ct_0 - (\vec{x}-\vec{b})\widehat{a} \tag{1.15}$$

The shower axis can be determined by minimizing the squares of the time differences between the measured signal start t_i (with variance $\sigma^2_{t_i}$) and the model time prediction $t_{\vec{x}_i}$ for each station:

$$\chi^2 = \sum_i \frac{[t_i - t(\vec{x}_i)]^2}{\sigma^2_{t_i}} = \sum_i \frac{[ct_i - ct_0 + \vec{x}_i\widehat{a}]^2}{c^2\sigma^2_{t_i}} = \sum_i \frac{[ct_i - ct_0 + x_i u + y_i v + z_i w]^2}{\sigma_i^2} \tag{1.16}$$

with $\vec{x}_i = \vec{x}_i - \vec{b}$ the position, $\sigma_i = c\sigma_{t_i}$ and u, v, z satisfying the constraint $u^2+v^2+z^2=1$. The axis determination is thus a non-linear problem. Nevertheless from a simplified linear model an approximate solution can be obtained and used as a starting point to more elaborate fitting procedures. First the estimated core location is projected along the shower axis $\widehat{a}$ towards the plane defined by the barycenter $\vec{b}$ and a normal $\widehat{n}$ and replaced with a new estimation. The lateral dependence of the signal measured in the tanks is modeled by $S(r) = S(1000)f_{LDF}(r)$, where $f_{LDF}(r)$ is a shape parametrization normalized in such a way that $f_{LDF}(1000\text{ m}) = 1$, and the parameter $S(1000)$ is the signal intensity at 1000 m from the shower axis. $S(1000)$ a parameter is a good estimator for the cosmic ray energy, and it depends on the incoming angle because of the shower attenuation in the atmosphere. Two types of LDF functions can be used to describe the data, modified power law or modified NKG functions:

$$f_{LDF}(r) = \begin{cases} \left(\frac{r}{r_{1000}}\right)^{\beta+\gamma\ln(r/r_{1000})} & r \geq r_{300} \\ \left(\frac{r}{r_{1000}}\right)^{\beta+\gamma\ln(r_{300}/r_{1000})} & r < r_{300} \end{cases} \tag{1.17}$$

$$f_{LDF}(r) = \left(\frac{r}{r_{1000}}\right)^{\beta}\left(\frac{r+r_{700}}{r_{1000}+r_{700}}\right)^{\beta+\gamma} \tag{1.18}$$

where r_{300}, r_{700}, r_{1000} are evaluated respectively at 300, 700 and 1000 m, β and γ are slope parameters to be determined from the fit procedure. Using the signal-weighted barycenter as the first estimation for the core position and reasonable initial values for β and γ, the $S(1000)$ parameter is pre-estimated based on the signal in the station closest to the $r = 1000$ m reference. Then a fit with respect to $S(1000)$ and core location is performed. Depending on the number of candidate stations, β and γ are gradually included as variable parameters in the LDF fit. In the next stage a realistic approximation of the shower front is used, including a parabolic term to describe the shower front curvature, and a new estimation of the axis and core location is found. With the new axis the LDF fit is performed again to get the final reconstructed values.

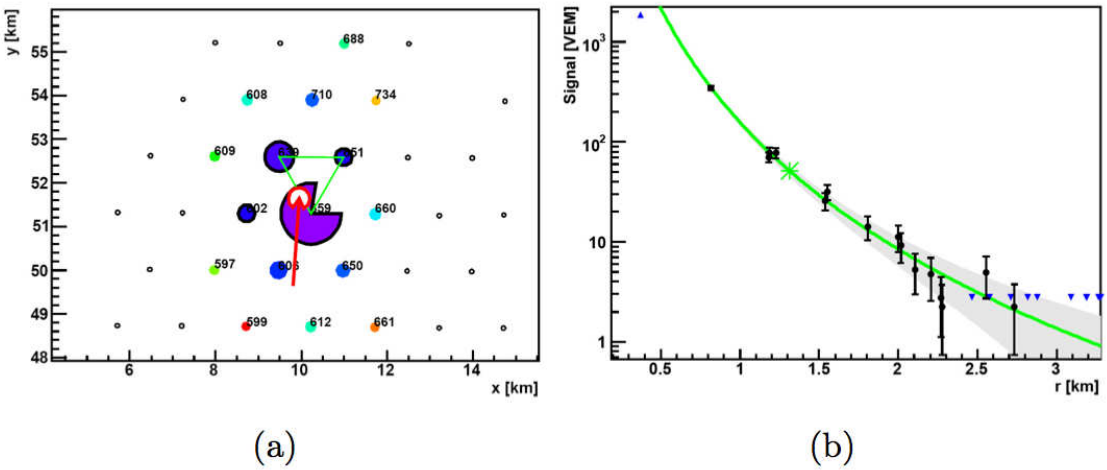


Figure 1.19: **Left panel:** Array layout of a real SD event measured at the Pierre Auger Observatory (ID 1153192). See text for explanations relative to station colors. **Right panel:** Reconstructed lateral distribution function.

In such a way the dependence of the energy measurement on the predictions given by the hadronic interaction models is extremely reduced. Typical results of this procedure are shown in Fig. 1.19 for a real event (SD id=1153192) measured by the surface detector. The final reconstructed parameters of this event are $\theta = 47.69° \pm 0.15°$, $\phi = -93.90° \pm 0.19°$ and $S(1000) = (156.6 \pm 10)$ VEM. The array layout is shown in the left panel of the figure: candidate stations are represented with filled circles (in bluish colors for on-time stations, in reddish colors for delayed stations), with size proportional to the square root of the signal. Zero-signal (or silent) stations are denoted by small black opaque circles. Candidate stations with black-rimmed circles have saturation in the high gain channel. The core position is shown with a red circle with an arrow indicating the shower azimuth. In the right panel of the same figure, the LDF fit of this event is shown: the signal of the candidate stations are plotted with black dots, while the LDF fit and the corresponding 1σ confidence band is shown in green and gray. Zero signal stations are plotted here as downward pointing blue triangles, while saturated-signal stations are represented as upward pointing blue triangles.

The primary energy is calculated according to the constant intensity cut (CIC) procedure. The $S(1000)$ parameter is converted into a reference signal size $S_{38°}$ by

$$S_{38°} = \frac{S(1000)}{1 + 0.92x - 1.13x^2}, \tag{1.19}$$

with $x = \cos^2\theta - \cos^2 38°$. $S_{38°}$ may be regarded as the signal $S(1000)$ the shower would have produced if it had arrived at $\theta = 38°$. The energy conversion is provided by samples of golden hybrid events. For these events the S_{38} parameter is compared to the FD energy estimate and a calibration curve is derived, yielding the desired conversion $E_{\mathrm{FD}} = \alpha\, S_{38}^{\beta}$, where $\alpha = (1.68 \pm 0.05) \times 10^{17}$ eV and $\beta = 1.035 \pm 0.009$ are the values obtained from the fit, as shown in Fig. 1.20. The energy resolution is 15%, with a systematic uncertainty of 22% [79].

Angular reconstruction and resolution. With the SD it is possible to determine the arrival direction of cosmic rays from the secondary particles arriving to the ground. The bulk of these particles is grouped in a thick nearly spherical pancake. The thickness of this spherical cap depends on the distance to the shower axis and ranges from hundreds of nanoseconds at about 100 m from the axis to few microseconds at 1000 m. The radius of curvature of the sphere depends on the altitude of the shower maximum and thus on the cosmic ray composition and can be used to describe the shower front.

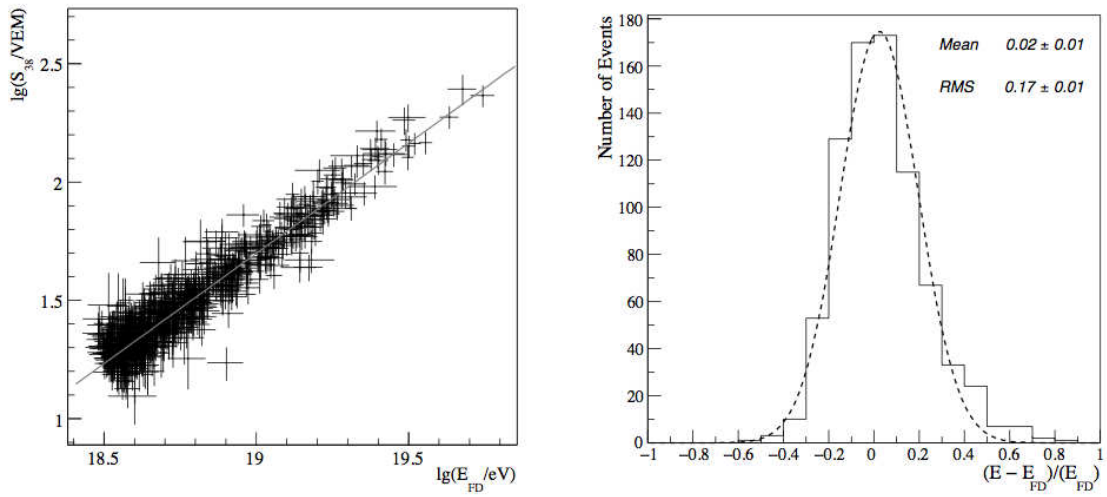


Figure 1.20: **Left panel:** Correlation between $\log S_{38}$ and $\log E_{FD}$ for a given golden hybrid data set used in the fit. The full line is the best fit for the data. **Right panel:** Fractional difference between the calorimetric energy, $E_{\rm FD}$, and the energy estimate of the surface detector, E, obtained by the calibration curve, for the selected events. The signal intensity S_{38} is expressed in Vertical Equivalent Muons (VEM), where 1 VEM is the intensity corresponding to one muon crossing the detector with zenith angle equal to zero [78].

The time of the first particle hitting the SD station is usually used as an estimator of the shower front arrival time (T_s) at each station location. The arrival direction of a SD event is determined by fitting the arrival time of the first particle (T_1) in each station according to a shower front model. Simulations indicate that a good representation of the shower front within 2 km from the shower axis is a spherical cap roughly centered in the muon generation region. Since this region is usually located in the upper atmosphere (about 5 km of altitude for vertical showers) one can consider a paraboloid approximation to simplify the description of the curve. Therefore, the determination of the arrival direction can be achieved by minimizing the expression:

$$\chi^2 = \sum_{\text{stations}} \frac{1}{\sigma_i^2} \left(T_{1i} - T_0 + \frac{u}{c} x_i + \frac{v}{c} y_i - \frac{d_i^2}{2Rc} \right)^2 , \tag{1.20}$$

with respect to T_0,u,v and R. The direction cosines are given by u and v, T_0 is the shower arrival time to the ground, and R is the radius of curvature. Even if at least four stations in the event are required for determining R, a fixed value of R is used for events with less than 5 stations. In these cases R is derived from simulations and is $R[m] = 10^{-5} \sec(\theta)$ where θ is the zenith angle.

To estimate the arrival direction of the primary cosmic ray with the maximum possible precision, it is also necessary to appropriately take into account the time uncertainty (σ_i) of each individual station participating to the event. For this purpose a time variance model was developed in Ref. [80] based on the shower properties at each station location.

The angular resolution, defined as the angular radius that would contain 68% of the reconstructed events, is better than 0.9° above 10 EeV.

FD event reconstruction

First the shower detector plane (SDP) is determined. This plane is defined by the shower axis and the location of the telescope detecting the event. The normal vector $\vec{n}$ of the SDP can be obtained with a

signal-weighted χ^2 fit using the pointing directions r_i of the triggered PMTs:

$$\chi^2 = \sum_i (\vec{n} \cdot \vec{r_i}) w_i \tag{1.21}$$

where the weights w_i are proportional to the signal amplitude $\propto \sqrt{S_i}$. A minimum number of 6 pixels with pulse is considered to reconstruct the event. Initially all triggered pixels are included to provide a first guess of the SDP. For the final SDP fit only the pixels within an interval of 2° around the trial SDP have been used.

In the second phase of geometry reconstruction, the pixel timing information is used to determine the orientation of the shower axis inside the SDP previously recovered. The problem is equivalent to find the closest distance of the shower axis to the observer R_p and the axis inclination in the SDP χ_0 (see left panel of Fig. 1.21). Defining T_0 as the reference time when the shower is observed at a distance R_p, the time difference of a photon arriving at the i-th PMT can be calculated taking into account the time the shower has moved through the atmosphere t_s and the propagation time of the photon to the telescope t_l:

$$t_s = \frac{R_p}{c \tan \theta_i} \quad t_l = \frac{R_p}{c_{air} \sin \theta_i} \tag{1.22}$$

The shower is assumed to propagate with the speed of light in vacuum c, while the fluorescence photons propagate with the corresponding speed of light in the atmospheric layer c_{air}. Usually the simplification $c_{air} \approx c$ is applied and the time difference δt is given by:

$$\delta t = t_l - t_s = \frac{R_p}{c_{air} \sin \theta_i} - \frac{R_p}{c \tan \theta_i} \sim \frac{R_p}{c} \tan \frac{\theta_i}{2} \approx \frac{R_p}{c} \tan \left(\frac{\chi_0 - \chi_i}{2} \right) \tag{1.23}$$

The axis parameters T_0, χ_0 and R_p are determined performing a time fit by minimizing the following χ^2:

$$\chi^2 = \sum_i w_i \left[T_0 + \frac{R_p}{c} \tan \left(\frac{\chi_0 - \chi_i}{2} \right) - t_i \right]^2 \tag{1.24}$$

where w_i are weights including the errors of the time measurements. Typically a minimum number of 8 pixels within an angular distance to the SDP of 2° take part in the time fit. While the SDP can be reconstructed with high accuracy, the deficit of the FD-mono analysis is the exact determination of the shower in the SDP.

A significant improvement in resolution can be achieved with the hybrid reconstruction. In this case the shower's footprint on ground is known from at least one tank information. This can be used as additional data point for the time fit. The expected arrival time of the shower front at the i-th triggered tank can be expressed as:

$$t_i^{\text{tank}} = T_0 + \frac{\vec{r}_j^{\text{tank}} \cdot \vec{a}}{c} \tag{1.25}$$

where $\vec{r}_i^{\text{tank}}$ is the vector pointing from the FD observer to the i-tank and $\vec{a}$ is the shower axis vector (see right panel of Fig. 1.21). This timing information is included in the previous χ^2 minimization, thus representing an important constraint in the shower geometry determination. It is clear that the time offset between the SD and the FD has to be taken into account to obtain correct results. This time offset can be estimated for example by analyzing the central laser facility shots data.

In Fig. 1.22 the 3D view (left panel) and the reconstructed time fit (right panel) of a hybrid event (Auger event id=200513500999 - SD id=1364365 - FD eye 1, run 895, event 981 - date 2005/05/16) are

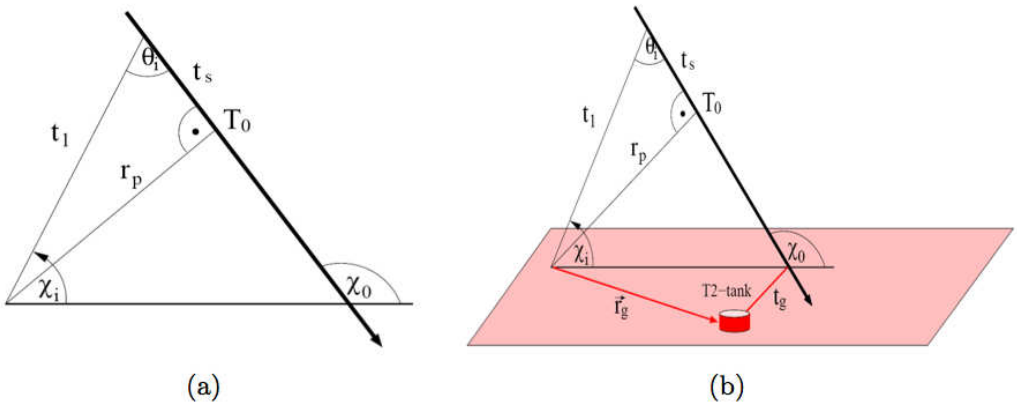


Figure 1.21: The shower geometry in the SDP in the FD-mono scheme and in the FD-hybrid scheme. The shower axis is defined by the parameters R_p, T_0 and χ_0.

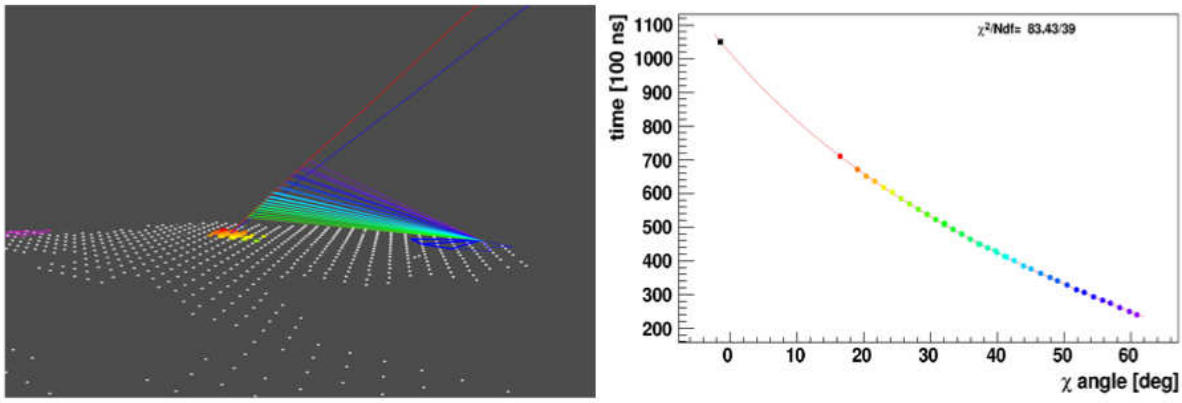


Figure 1.22: Array view and time fit for a hybrid event (Auger event id=200513500999). Note the colored group of tanks involved in the event in the 3D view and the black square in the time fit representing the timing of the hottest station.

shown. Note the group of tanks involved in the event in the 3D view and the black square in the time fit representing the timing of the hottest station, offering an additional constraint in the geometry finding.

An additional possibility to improve the geometry determination, offered by the current detector layout, is the stereo mode: a subset of the available events is measured by two or more FD eyes at the same time. In such cases, the SDP are determined separately for each single FD eye and then the shower axis is found by the intersection of these SDPs, without the timing information. The achieved resolution depends on the accuracy with which the single SDPs are determined and on the intersection angle among the SDPs.

Once the shower geometry is fully reconstructed, it is possible to determine the longitudinal profile of the shower. This is implemented in several steps. A calculation of the light flux reaching the telescope aperture is first carried out. See [81] for a review on different methods adopted to collect light. The FADC pulses in each pixel, expressed in ADC counts, have to be associated with photons arriving at the diaphragm, through the pixel calibration constants. For each FADC time bin (100 ns) it is possible to define a vector $\vec{R}(t)$ lying on the SDP and pointing to the calculated shower position. The signal of a given pixel i is considered in the light collection algorithm if the angular distance $\zeta_i = \arccos(\vec{R}(t) \cdot r_i^{pixel})$ between $\vec{R}(t)$ and the pixel pointing direction vector r_i^{pixel} is less than a given value ζ_{opt}. The optimum

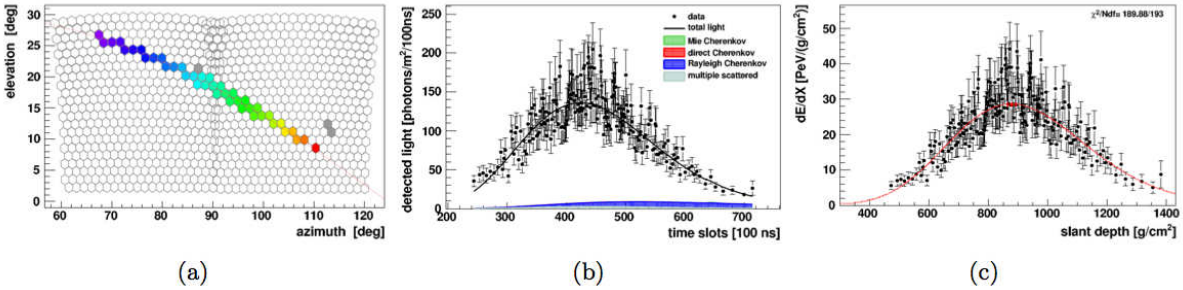


Figure 1.23: a) The camera view of a measured hybrid event (the same of Fig. 1.22). The shower appears as a sequence of hit pixels colored according to their timings from violet to red and crossing in this case two Fd mirrors. b) Reconstructed light at aperture. Different contributions to the light profile (fluorescence, direct and scattered Cherenkov) are shown in different colors. c) Reconstructed longitudinal profile. The red line represents the Gaisser-Hillas fit to data.

value of ζ_{opt} is scanned to maximize the signal to noise ratio S/N:

$$\frac{S}{N} = \frac{\sum_{time\,bins} \sum_{i}^{\zeta_i<\zeta_{opt}} [n_i^{ADC}(t) - n_i^{ped}]}{\sqrt{\sum_{time\,bins} \sum_{i}^{\zeta_i<\zeta_{opt}} [\sigma_i^{ped}]^2}} \tag{1.26}$$

where n_i^{ADC} is the digitized pixel signal, n_i^{ped} is the ADC mean value of the pedestal and σ_i^{ped} is the variance of the pixel pedestal. Once ζ_{opt} has been fixed, the light flux at the diaphragm aperture is calculated as 370 nm equivalent photons in 100 ns time bins t as:

$$n_\gamma^{370}(t) = \sum_{i}^{\zeta_i<\zeta_{opt}} C_i^{370}(n_i^{ADC} - n_i^{ped}) \tag{1.27}$$

In the standard reconstruction only the light detected by pixels within ζ_{opt} around the SDP is used [82] and the fraction of fluorescence light lost outside ζ_{opt} is corrected for using the universal lateral fluorescence light distribution from [83]. An alternative approach, the "spot method", assumes a model for the light distribution at the telescope focal surface (the spot) and uses the light coming from all pixels [81].

In Fig. 1.23a the typical profile structure of a measured hybrid event (the same of Fig. 1.22) is shown. Fig. 1.23b the camera view is represented as a sequence of hit pixels, colored according to their timings from violet to red and crossing two FD mirrors. In Fig. 1.23c the corresponding light at aperture is shown as reconstructed with the above procedure. Different contributions to the light profile (fluorescence, direct and scattered Cherenkov) are also plotted.

The last step is the profile reconstruction, which converts the fluorescence light profile recorded by the telescopes to a determination of the energy deposit at a given atmospheric depth along the shower axis. The profile reconstruction method so far adopted is presented and detailed in [84]. The light arriving at the aperture is actually a combination of direct and scattered components induced by the shower passage through the atmosphere, hence it is necessary to disentangle all different contributions. The direct fluorescence light, emitted isotropically at a certain slant depth X_i and measured at the detector at time t_i, is proportional to the energy deposit. The number of fluorescence photons $N_\gamma^{\text{fluor}}(X_i)$ and

measured light flux y_i^{fluor} produced in a slant depth interval ΔX_i are given by:

$$N_\gamma^{\text{fluor}}(X_i) = Y_i^{\text{fluor}} \frac{dE}{dX_i} \Delta X_i \qquad y_i^{\text{fluor}} = d_i N_\gamma^{\text{fluor}} \tag{1.28}$$

where dE/dX_i is the energy deposited per unit depth at slant depth X_i, Y_i^{fluor} is the fluorescence yield (number of photons expected per unit deposited energy for the atmospheric pressure and temperature at X_i) and $d_i = \varepsilon T_i \frac{A}{4\pi r_i^2}$. The photons are distributed over a sphere with surface $4\pi r_i^2$, where r_i denotes the distance from the detector. Due to the atmospheric attenuation only a fraction T_i of them can be detected, given a light detection efficiency ε.

The total light received at the detector at the time t_i is obtained by adding the fluorescence, the scattered and direct Cherenkov light contributions $y_i = y_i^{\text{fluor}} + y_i^{\text{cher dir}} + y_i^{\text{cher scat}}$. The aim of the profile reconstruction is to estimate the energy deposit and/or the electron profile from the observed light flux. In traditional methods [56] the Cherenkov light is iteratively subtracted from the measured total light to obtain the fluorescence contribution, but a known drawback is that the method does not converge for events with a large amount of Cherenkov light and the propagation of the uncertainties is critical. Instead one can use the fact that the electron energy spectrum is universal in shower age s_i (no dependence on primary mass or energy) [85–87], yielding the relation $dE/dX_i = N_i^e \alpha_i$, with α_i average energy deposit per unit depth per electron at shower age $s_i = 3/(1 + 2X_{\max}/X_i)$.

It has been assumed that the shower induces light emission at a single wavelength λ, but in reality the fluorescence yield shows distinct emission peaks and the number of Cherenkov photons is proportional to $1/\lambda^2$. In a final reconstruction stage also the wavelength dependence of the detector efficiency and the light transmission are taken into account.

A knowledge of the complete profile is required for the calculation of the Cherenkov beam and the shower energy. Due to the limited field of view of the FD only a part of the profile is observed, therefore an appropriate function for the extrapolation to unobserved depths is needed. The adopted choice is the Gaisser-Hillas function [52]:

$$f_{GH}(X) = \frac{dE}{dX}_{max} \left(\frac{X - X_0}{X_{max} - X_0} \right)^{(X_{max} - X_0)/\lambda} e^{(X_{max} - X_0)/\lambda} \tag{1.29}$$

Finally, the calorimetric energy is estimated by integrating the Gaisser-Hillas function previously determined over the energy deposit profile:

$$E_{\text{cal}} = \int_0^\infty f_{GH}(X) dX = \lambda \frac{dE}{dX}_{max} \left(\frac{e}{\xi} \right)^\xi \Gamma(\xi + 1) \tag{1.30}$$

with $\xi = (X_{max} - X_0)/\lambda$. To avoid the analytical propagation of the uncertainty over E_{cal}, the normalization term of the Gaisser-Hillas function is replaced with $E_{\text{cal}}/[\lambda\Gamma(\xi + 1)]$ and the minimization is performed also with respect to E_{cal}. The error of the E_{cal} estimate is in this way directly obtained from the fit procedure. It has to be noted that not all the energy of the shower goes in the electromagnetic part, as it was considered above. Neutrinos escape undetected and muons need long path length to fully release their energy. This is usually accounted for by multiplying the calorimetric energy with a correction factor f_{inv} determined from shower simulations to obtain the total primary energy $E_{\text{tot}} = f_{\text{inv}} E_{\text{cal}}$, where f_{inv} is parameterized as $f_{\text{inv}} = (a + bE_{\text{cal}}^c)^{-1}$ with a, b, c constants depending on the primary composition and interaction model assumed.

In Fig. 1.24 we show a real hybrid event (top-left), together with the corresponding longitudinal development (right) and lateral distribution function (down-left), summarizing the procedure described so far and generally adopted for the reconstruction of the arrival direction, the energy and the elongation rate of the UHE event.

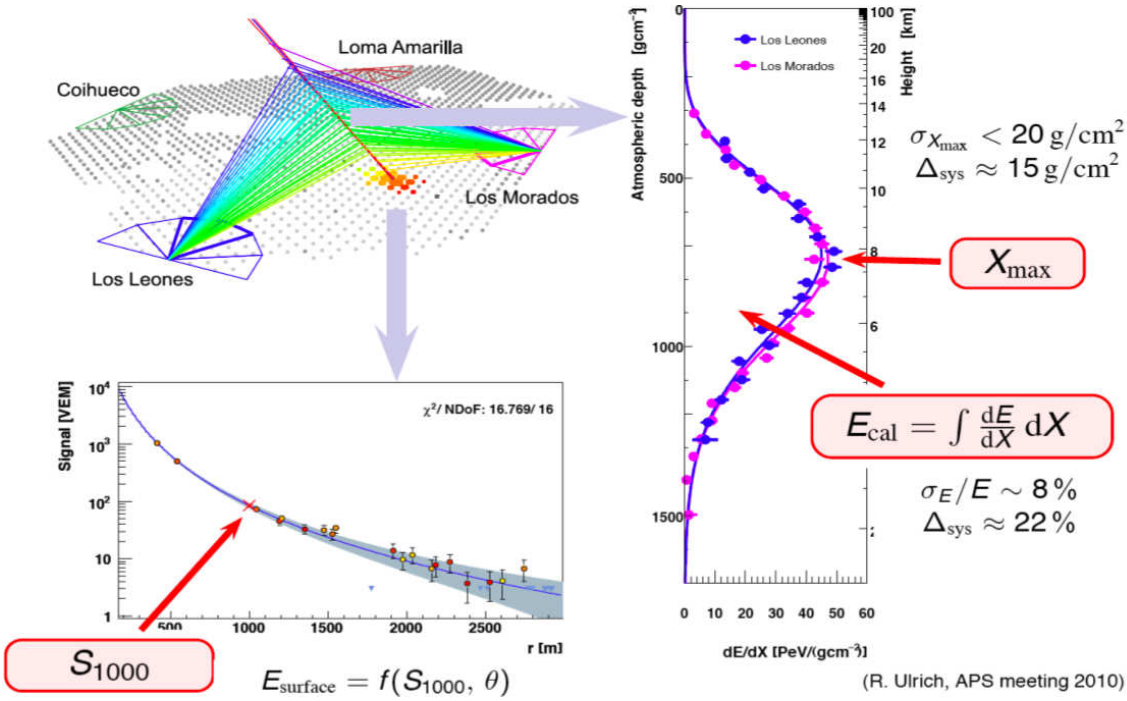


Figure 1.24: A real hybrid event (top-left) detected with the Pierre Auger Observatory. The corresponding longitudinal development measured with the FD (right) and the lateral distribution function measured with the SD (down-left) are shown.

1.6.4 Main results

In the following, we will briefly report some of the main results obtained at the Pierre Auger Observatory. All plots and figures will be referred to such results and measurements, if not specified otherwise.

In 2007 the Pierre Auger Observatory reported a 99% CL correlation between the arrival directions of detected UHECRs with energy above 55 EeV and AGN within 75 Mpc [88]. The recent observations (Fig. 1.25) still show a significant amount of correlation, although weaker than the previous claim but still far from the isotropic expectation (Fig. 1.26) [89]. An apparent excess around the direction of Centaurus A has been reported as well [89]. If the particles responsible for such an excess above 56 EeV are heavy nuclei with charge Z, the proton component of the sources should lead to excesses in the same regions at energies E/Z. However, the lack of anisotropies in these directions, at such energies, has been reported [90].

In Fig. 1.27 we show the measurements of the average maximum shower development $X_{\max}$ (left panel) and its fluctuation (right panel) as a function of the energy [91]. Lines above and below data points indicate the expectation from hadronic models in the case of protons and irons: a clear trend towards a heavy mass composition is found at higher energy. Such results are an update of the recent measurements reported in Ref. [92]. However, the interpretation of such a result and its compatibility with other results is rather difficult and it is still under investigation.

In Fig. 1.28 we show the measured energy spectrum above 10^{18} eV, derived from SD data calibrated with FD measurements (left panel) and that combined with hybrid data (right panel). The strong suppression of the flux above 40 EeV is evident, with a statistical significance larger than 20σ [28, 93], in agreement with the result reported by the HiRes Collaboration [94]. The measurement of a suppression of the flux at the highest energies is consistent with an extragalactic origin of the UHECRs and with the expectation from the GZK effect.

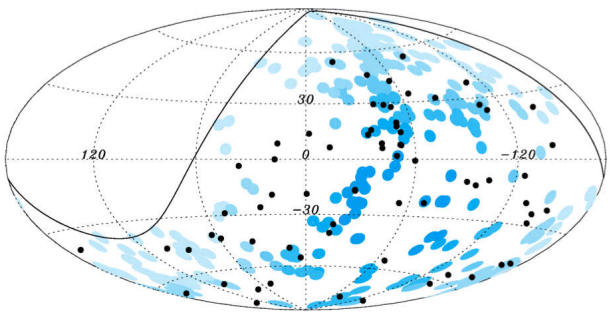


Figure 1.25: Arrival directions, in galactic coordinates, of UHECRs with energy above 56 EeV (black dots) and positions of AGN within 75 Mpc in the Veron-Cetty & Veron catalog. Color codes the relative nonuniform exposure of the Observatory [89].

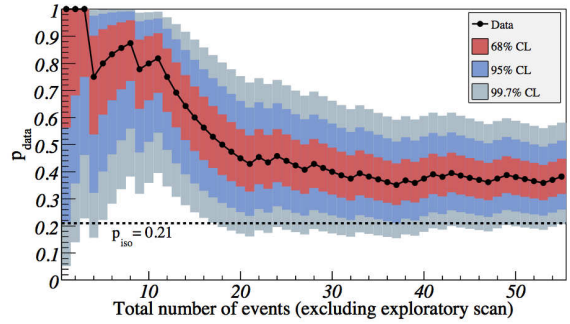


Figure 1.26: The most likely value of the degree of correlation $p_{\rm data} = k/N$ is plotted with black dots as a function of the total number of time-ordered events (excluding those in period I). The 68%, 95% and 99.7% confidence level intervals around the most likely value are shaded. The horizontal dashed line shows the isotropic value $p_{\rm iso} = 0.21$ [89].

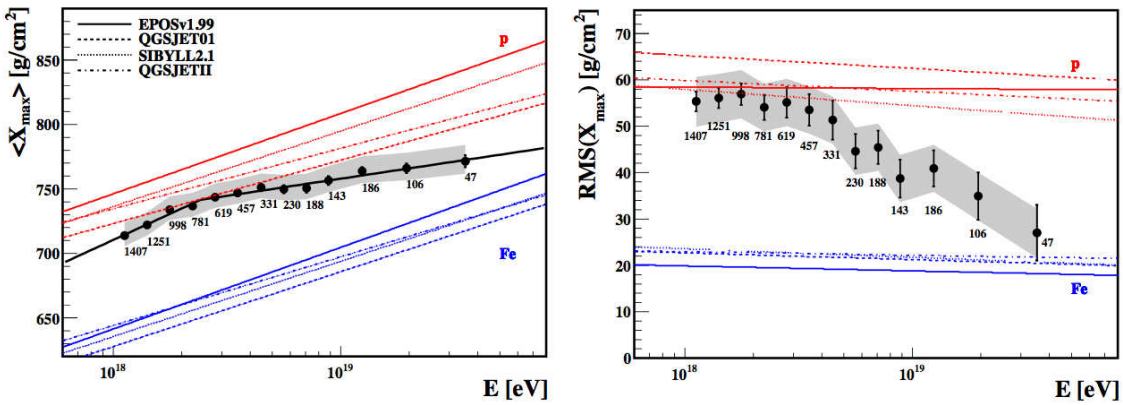


Figure 1.27: Most up-to-date measurements of $\langle X_{\rm max}\rangle$ (left panel) and RMS($X_{\rm max}$) (right panel) as a function of the energy. Data (points) are shown with the predictions for proton and iron for several hadronic interaction models. The number of events in each bin is indicated. Systematic uncertainties are indicated as a band [91].

The energy spectrum derived from hybrid data is combined with the one obtained from surface detector data using a maximum likelihood method. Since the surface detector energy estimator is calibrated with hybrid events, the two spectra have the same systematic uncertainty in the energy scale (22%). On the other hand, the normalization uncertainties are independent. They are taken as 6% for the SD and 10% (6%) for the hybrid flux at 10^{18} eV ($> 10^{19}$ eV). These normalization uncertainties are used as additional constraints in the combination. In the right panel of Fig. 1.28 the resulting combined energy spectrum is shown [95].

Fig. 1.29 shows the up-to-date measurements of the proton-air cross section compared to other measurements and model predictions [96]. The knowledge of such cross section plays a central role in the physics of extended air showers and it has a non-negligible impact on current hadronic models.

In Fig. 1.30 we show upper bounds to the flux of (still unobserved) photons, at different energies [97]. Previous bounds [98] and bounds obtained from AGASA and Yakutsk are shown, as well as predictions from top-down models and GZK photon flux. The current observations of the Pierre Auger Observatory rule out several exotics model, whereas $Z-$bursts and GZK models are still compatible with the data.

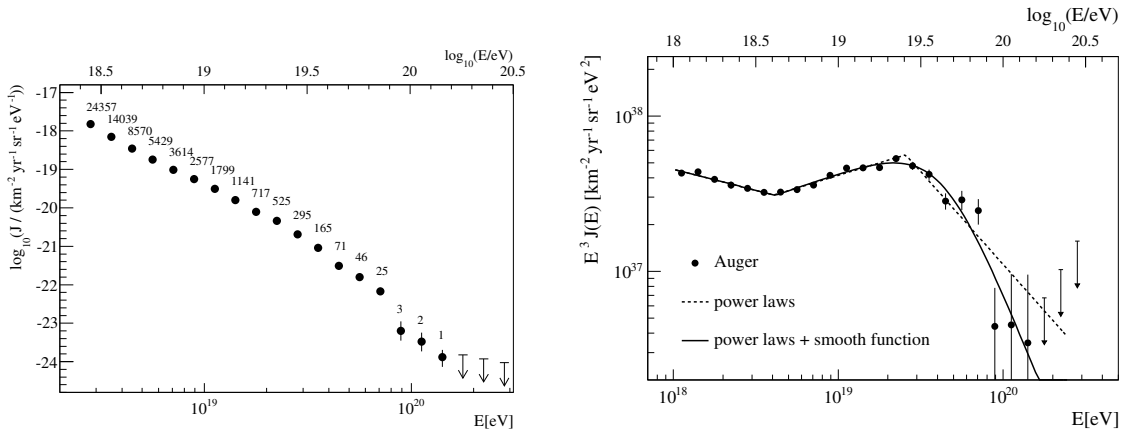


Figure 1.28: Up-to-date measurements of the energy spectrum as a function of the energy. **Left panel**: Energy spectrum derived from surface detector data calibrated with fluorescence detector measurements. The spectrum has been corrected for the energy resolution of the detector. Only statistical uncertainties are shown. Upper limits correspond to 68% CL. **Right panel:** Combined energy spectrum fitted with two functions (see [95]). Only statistical uncertainties are shown. The systematic uncertainty in the energy scale is 22%.

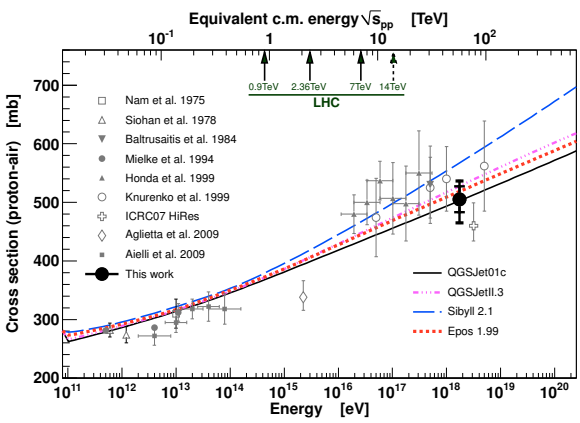


Figure 1.29: Up-to-date proton-air cross section compared to other measurements and model predictions. The inner error bars are statistical only, while the outer include all systematic uncertainties for a helium fraction of 25% and 10 mb photon systematics (see Ref. [96] and Refs. therein for further details).

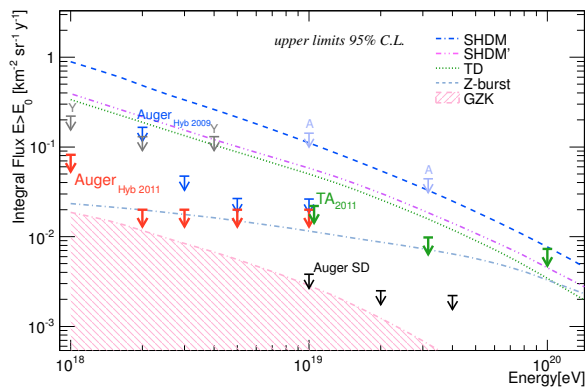


Figure 1.30: Up-to-date upper limits on the photon flux (red arrows) compared to previous limits from Auger, AGASA (A) and Yakutsk (Y). The shaded region and the lines give the predictions for the GZK photon flux and for top-down models (see Ref. [97] and Refs. therein for further details).

With the SD of the Pierre Auger Observatory, we can also detect ultra-high energy neutrinos in the sub-EeV energy range and above. Neutrinos of all flavours can interact in the atmosphere and induce inclined showers close to the ground (down-going). The sensitivity of the SD to tau neutrinos is further enhanced through the "Earth-skimming" mechanism (up-going). Both types of neutrino interactions can be identified through the broad time structure of the signals induced in the stations of the SD.

In the left panel of Fig. 1.31 we show the exposure of the SD for Earth-skimming neutrino initiated showers and for down-going neutrino initiated showers for all the considered channels as a function of neutrino energy. In the right panel of the same figure, differential and integrated upper limits (90% C.L.) for a diffuse flux of down-going neutrinos and Earth-skimming ν_τ are shown, together with expected fluxes from exotic and GZK models [99].

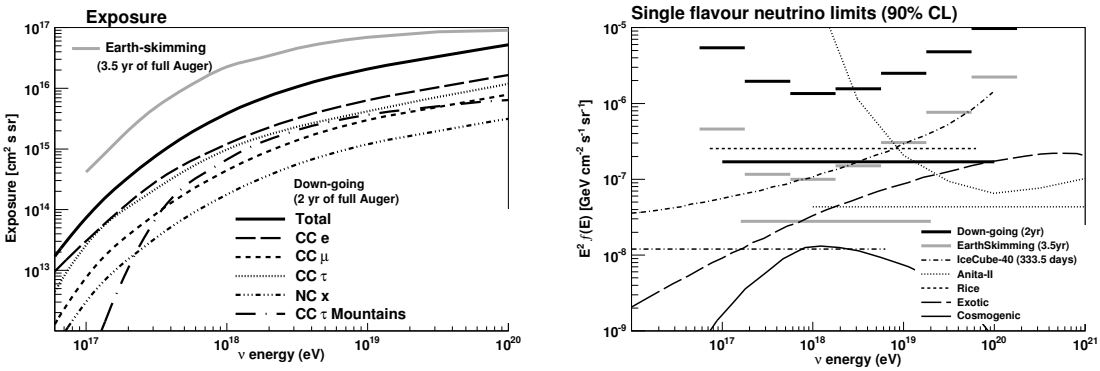


Figure 1.31: **Left panel:** Exposure of the surface detector of the Pierre Auger Observatory for Earth-skimming neutrino initiated showers (3.5 yr of full Auger) and for down-going neutrino initiated showers for all the considered channels as a function of neutrino energy (2 yr of full Auger). **Right panel:** Differential and integrated upper limits (90% C.L.) from the Pierre Auger Observatory for a diffuse flux of down-going ν (2 yr of full Auger) and Earth-skimming ν_τ (3.5 yr of full Auger). Limits from other experiments are also plotted. Expected fluxes are shown for cosmogenic neutrinos and for a theoretical exotic model (see Ref. [99] and Refs. therein for further details).

To myself I am only a child playing on the beach, while vast oceans of truth lie undiscovered before me.

I. Newton

The wise man looks into space and does not regard the small as too little, nor the great as too big, for he knows that, there is no limit to dimensions.

Chuang Tzu

Chapter 2

Simulating the propagation of UHECRs I: magnetic fields

At the end of the previous chapter, we have discussed the most likely origin of UHECR, although it is still unknown, by describing candidate sources able to accelerate particles up to the highest energies. Within the present and the next chapter, separated for simplicity, we will present the Monte Carlo we developed to simulate of nuclei propagation in the Universe. In this chapter, we will describe the effects of magnetic fields as included in the propagation code. In the next one, we will instead describe the included relevant energy-loss processes, due to the interaction of UHECR with photons of the extragalactic background radiation. Also, by mean of the described Monte Carlo, we will explore the impact of such propagation effects on physical observables as the energy spectrum, mass composition and arrival directions. In the following we will refer to such a code as "HERMES".

In the first section, the general structure of HERMES is presented. Successively, the origin of cosmic magnetic fields and their modeling for what concerns our Galaxy and the nearby extragalactic space, are briefly described. Recent models for the regular component of the galactic field (Stanev, Harari-Mollerach-Roulet, Tinyakov-Tkachev with ASS or BSS patterns) and, in general, for the turbulent component which can be simulated with HERMES, are discussed. The diffusion in the irregular (turbulent) magnetic field is briefly treated, together with the theoretical and numerical estimation of the components of the diffusion tensor. Preliminary checks of consistency between results obtained by HERMES and those by other codes available in literature are reported.

2.1 General structure of the HERMES code

UHECRs, produced by powerful astrophysical objects, propagate in the Universe experiencing deflections due to magnetic fields and energy-loss processes due to the interaction with photons of the extragalactic background radiation (EBR). In Fig. 2.1 we show a simplified sketch of the propagation, whose simulation will be presented in some detail within this chapter and the next one. In fact, the contribution of many factors should be taken into account; to cite the most relevant: i) geometry of the (expanding) Universe; ii) interactions with EBR, responsible for the creation of secondary UHE photons and neutrinos for any charged nucleus, and, in the particular case of heavy nuclei, secondary nucleons and lighter nuclei produced because of photo-disintegration processes; iii) regular magnetic fields responsible of bending particles; iv) irregular ones responsible of randomly scattering them. The simulation of cosmology and energy-loss will be treated in the next chapter, whereas magnetic fields will be treated in the next sections.

In this section we outline the preliminary version of HERMES, our Monte Carlo code simulating the

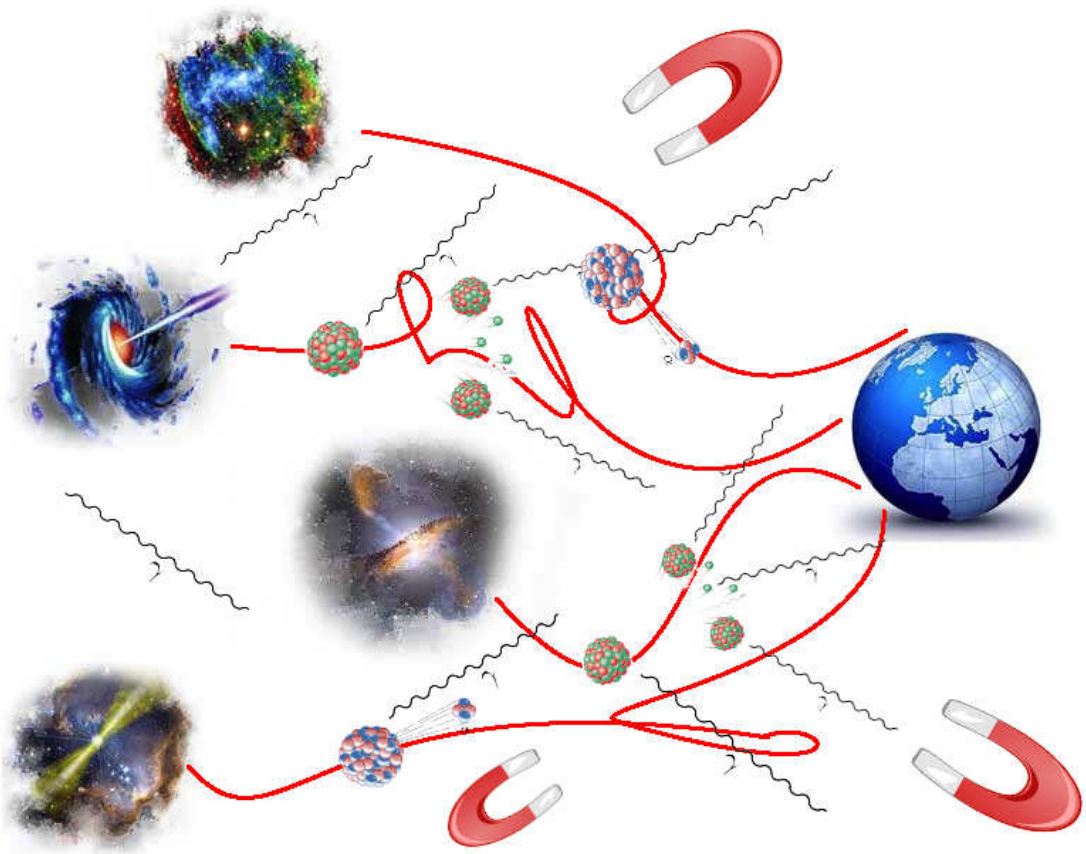


Figure 2.1: Sketch of nuclei propagation in a magnetized Universe.

propagation of UHECRs. It is composed by different modules, each one performing a specific task. A sketch of such modules and their interconnections is shown in Fig. 2.2, including:

- **MagProp:** responsible for i) the parameterization and the simulation of galactic magnetic fields; ii) the simulation of the irregular (turbulent) component of the extragalactic magnetic field; iii) the propagation of nuclei in the simulated magnetic fields;
- **xHERMES:** responsible for i) the parameterization of the EBR; ii) the parameterization of the cross sections regarding the relevant interactions between charged nuclei and photons of the EBR; iii) the numerical estimation of the corresponding mean free paths and energy-loss lengths;
- **EleCa:** responsible for the full propagation of the electromagnetic cascade generated by nuclei interactions with photons of the EBR.

The kernel of HERMES is responsible for interconnecting the different modules, allowing the simulation of several astrophysical scenarios. The propagation of charged nuclei, as well as secondary photons and neutrinos produced during their propagation, is included. Any model that will be discussed farther in the text, regarding the EBR, the magnetic fields or the interactions with EBR, can be easily switched off or taken into account in the simulation.

It is worth remarking here that such a code will be finalized in collaboration with Dr. O. Deligny and Dr. H. Lyberis (Institut de Physique Nucléaire d'Orsay), and with Dr. M. Settimo (University of Siegen), who is developing the component for the propagation of the electromagnetic cascade.

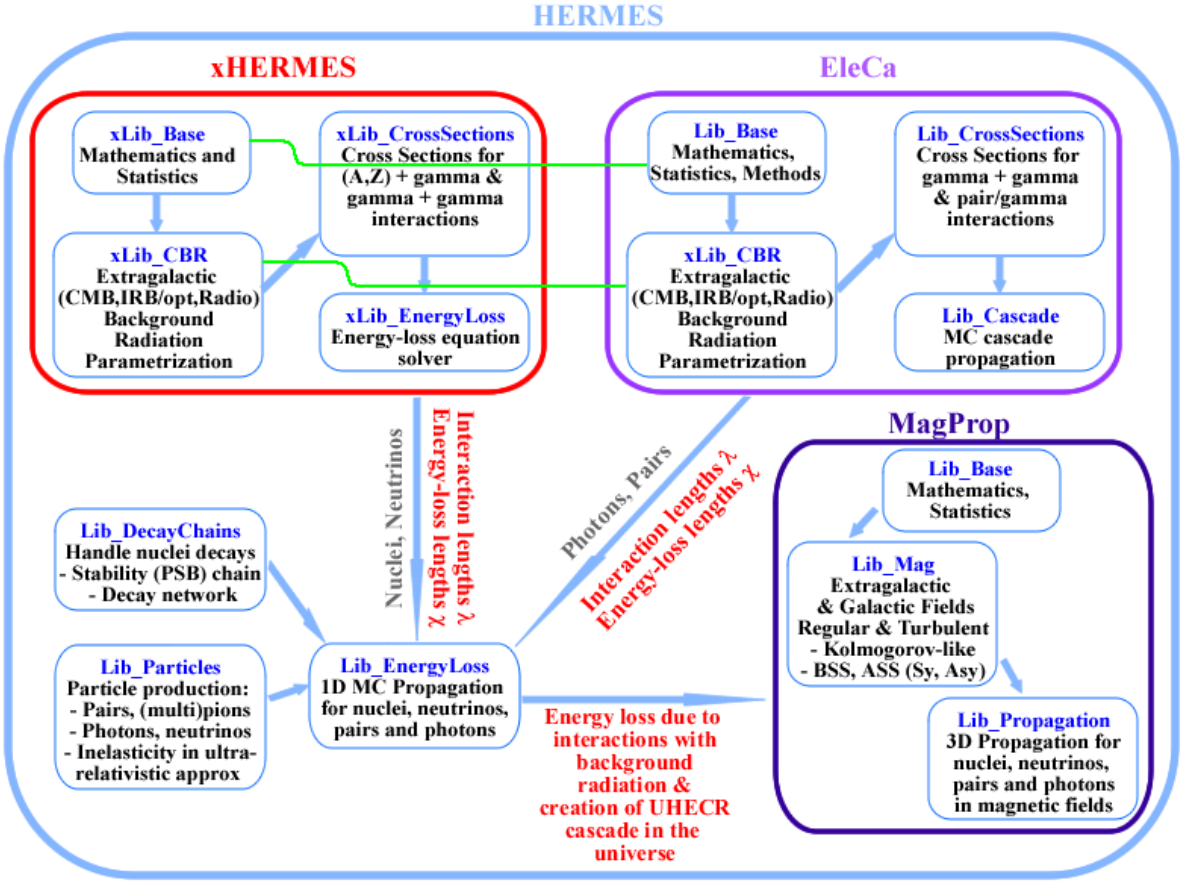


Figure 2.2: Sketch of the simulation modules which build HERMES.

2.2 Origin of cosmic magnetic fields

The presence of magnetic fields, both in the intergalactic space and in our galaxy, has a non-negligible impact on the propagation of charged nuclei. It is thus of fundamental importance to investigate the structure of galactic and extragalactic magnetic field (GMF and EMF, respectively), that have a direct impact on the energy spectrum, the strength of the anisotropy signal and the correlation with candidate sources.

Note. It is worth remarking here that the presence of magnetic fields does not introduce a spurious anisotropy in the arrival directions of UHECR, if such anisotropy is not *intrinsic*, i.e. if it is not due to the anisotropic distribution of the candidate sources. In fact, the magnetic field does not change the magnitude of the cosmic-ray momentum: the phase space is preserved and the flux is constant along each particle trajectory. Such a behavior is ensured by the Liouville's theorem. Such statement is true as long as particles are not lost to a volume element by collision or radiation processes, conditions satisfied in large part for cosmic-ray ions in the Galaxy spiraling around the interstellar magnetic fields. If there exists a trajectory for any direction at Earth that can be followed back to the outer sphere of sources, the flux at Earth has to be the same as it is just outside the Galactic magnetosphere. If the flux is

isotropic outside the GMF, then it is also isotropic at Earth [100].

In stars and galaxies, and indeed in many other astrophysical settings, the gas is partially or fully ionized and can carry electric currents that, in turn, produce magnetic fields. The associated Lorentz force exerted on the ionized gas (also called plasma) can in general no longer be neglected in the momentum equation for the gas. Magneto-hydrodynamics (MHD) is the study of the interaction of the magnetic field and the plasma treated as a fluid. In MHD, Maxwell's equations of electrodynamics are combined with the fluid equations, including also the Lorentz forces due to electromagnetic fields [101]. The obtained evolution equation describes the variation of the magnetic field $\vec{B}$ with time in a given point of the space[1]

$$\frac{\partial \vec{B}}{\partial t} = \vec{\nabla} \wedge \left(\vec{v} \wedge \vec{B} - \eta \vec{\nabla} \wedge \vec{B} \right), \tag{2.1}$$

where $\vec{v}$ is the velocity of the plasma relative to the magnetic field in some fixed reference frame and $\eta = c^2/(4\pi\sigma)$ is the magnetic diffusivity and σ is the electric conductivity.

One can argue that $\vec{B} = 0$ is a valid solution of the evolution equation, with the direct consequence that no magnetic field would be generated if one were to start with zero magnetic field, the most quite plausible assumption on the initial condition of the early Universe. In order to drive currents from a state with initially no current, we need some way of violating the induction equation to produce a cosmic battery effect, as the Biermann battery [102] or other mechanisms [103–106], to cite some of them.

The solution of Eq. (2.1) is able to describe, for instance, the magnetic field inside a galaxy. If the field is not allowed to undergo stochastic fluctuations, the resulting magnetic field provides the *regular* component. Conversely, if Eq. (2.1) is adopted to describe the evolution of the initial fluctuations, the resulting magnetic field provides the *irregular* component (generally referred to as turbulent). Recently, it has been proposed a scenario in which turbulent flow motions are induced via the cascade of the vorticity generated at cosmological shocks during the formation of the large scale structure. Simulations have shown that the turbulence inside clusters and groups of galaxies is subsonic, whereas it is transonic or mildly supersonic in filaments. By mean of a turbulence dynamo model, the average magnetic field strength has been estimated to be of the order of few μG inside clusters/groups, approximately 0.1 μG around clusters/groups, and approximately 10 nG in filaments. Moreover, such model presents a physical mechanism that transfers the gravitation energy to the turbulence and magnetic field energies in the large scale structure of the Universe [107].

2.3 Simulating the motion of charged particles

In order to simulate the diffusion of charged particles in magnetic fields, we adopt in HERMES a standard approach, based on the numerical integration of the equation of motions obtained in the ultra-relativistic approximation, in the case of nuclei.

In general, the motion of a particle with charge $q = Ze$ in an electromagnetic field defined by the field strength tensor $F^{\mu\nu}$ is described by

$$\partial_0 p^\mu = q v_\nu F^{\mu\nu}, \tag{2.2}$$

being the contravariant counterpart of the classical Lorentz equation. Here, p^μ represents the particle 4-momentum, v_ν indicates its 4-velocity and the derivative should be intended with respect to the proper

[1] The formal derivation of the induction equation is not within the scope of the present work: we refer to [101] and Refs. therein for a detailed discussion of this point.

time. If the electric field is absent (or negligible) and we assume the case of a particle in ultra-relativistic regime, i.e. the particle travels at the speed of light in the direction $\hat{v}(t)$ at time t subjected to a magnetic field $\vec{B}(\vec{r})$ along the trajectory $\vec{r}(t)$, the Lorentz equation reduces to the set of six ordinary differential equations defined by

$$\begin{aligned} \frac{d\vec{r}(t)}{dt} &= c\hat{v}(t) \\ \frac{d\hat{v}(t)}{dt} &= \frac{qc^2}{E}\hat{v}(t) \wedge \vec{B}(\vec{r}). \end{aligned} \tag{2.3}$$

In practice, charged particles accelerating in a magnetic field lose energy because of the emission of synchrotron radiation: in the case of light particles as electrons or positrons, such energy loss should be taken into account during the propagation, whereas for heavier particles as protons it is negligible.

The helical path of a charged particles propagating in an uniform regular magnetic field $\vec{B}_0$ is characterized by a pitch angle θ and a Larmor radius

$$r_L = \frac{pc}{ZeB_0} \simeq \frac{E/Z}{10^{18}\ \text{eV}} \left(\frac{\mu\text{G}}{B_0}\right)\ \text{kpc}, \tag{2.4}$$

The radius of the helical trajectory is $r_L \sin\theta$, whereas $v_\parallel = c\cos\theta$ is the component of the velocity parallel to $\vec{B}_0$.

2.4 Modeling the diffusion in the irregular field

For the simulation of the irregular component of the magnetic field, we adopt in HERMES the approach proposed by Giacalone and Jokipii, based on a local step-by-step simulation of the turbulent field.

Although some techniques have been developed to detect magnetic fields in cosmic objects, our knowledge of such fields is still poor. The magnetic field component $B_\perp$ perpendicular to the line of sight can be measured from the polarization of optical, infrared and radio synchrotron emission [108], whereas the parallel component $B_\parallel$ can be obtained from the Faraday effect. The synchrotron process is responsible of creating linearly polarized radio waves whose polarization planes are rotated when such waves experience a uniform magnetic field. From the measurements of such rotations (so-called RM), being clockwise if the magnetic field points away from the observer and counterclockwise otherwise, it is possible to estimate $B_\parallel$. Other techniques exploit the Zeeman effect, i.e. the splitting of spectral lines into some distinct components because of the degeneracy breaking of quantum states of electrons in atoms. Zeeman splitting and Faraday rotation measurements provide information about the magnetic field component along the line of sight and its sign, whereas synchrotron radiation and polarimetry (of starlight dust) are limited to the component perpendicular to the line of sight.

Magnetic fields in galaxies are mainly probed using radio observations of their synchrotron emission. It has been shown that the mean equipartition strength of the total magnetic field for a sample of 74 spiral galaxies is $\langle B_{\text{tot}} \rangle \approx 9\ \mu$G [109], with the total field strength ranging from $\approx 4\ \mu$G, in radio faint galaxies like M31 and M33, to $\approx 15\ \mu$G, in grand design spiral galaxies like M51, M83 and NGC 6946 [110].

Unfortunately, we have no exact knowledge of both galactic and extragalactic magnetic fields and, as a consequence, the investigation of charged particles propagation through our galaxy and intergalactic space, respectively, should be based on empirical (or theoretical) models and numerical simulations.

While the trajectory of a charged particle along the regular field is deterministic, i.e. for a given initial condition only one solution to the equations of motion exists, the trajectory of a particle through the turbulent field is stochastic, thus not unique, and it depends on the features of the irregular field as

its r.m.s. strength and its coherence length. The irregular motion in d dimension can be macroscopically described in terms of the particle density $\varrho(\vec{r},t)$ by the standard diffusion equation

$$\frac{\partial \varrho}{\partial t} = \sum_{i,j=1}^{d} \mathbb{D}_{ij} \frac{\partial^2 \varrho}{\partial x_i \partial x_j}, \tag{2.5}$$

where the components of the diffusion tensor $\mathbb{D}$ are the transport coefficients that have to be calculated from the microscopic dynamics.

As any other tensor, the diffusion tensor can be decomposed into a symmetric and an antisymmetric component. In particular, if $\vec{b} = \vec{B}_0/B_0$ is a unit vector along the regular magnetic field, we have

$$\mathbb{D}_{ij} = \left(D_\parallel - D_\perp\right) b_i b_j + D_\perp \delta_{ij} + D_A \epsilon_{ijk} b_k, \tag{2.6}$$

where δ_{ij} is the Kronecker delta symbol, and ϵ_{ijk} is the Levi-Civita fully antisymmetric tensor. The symmetric terms contain the diffusion coefficients parallel and perpendicular to the mean field, $D_\parallel$ and $D_\perp$, which describe diffusion due to small-scale turbulence, while the antisymmetric term contains the Hall diffusion coefficient D_A describing the macroscopic drift associated to the gradient of the CR density.

In the following we will discuss our modeling of turbulent component of the magnetic field and the numerical estimation of the diffusion coefficients as a function of the turbulence level, providing important information about the propagation of cosmic rays of all energies in a variety of astrophysical environments, as well as the efficiency of Fermi acceleration processes.

Modeling of the turbulent field. The randomness of the irregular component of the magnetic field is probably due to the evolution of stochastic fluctuations which are correlated up to a given correlation scale. In fact, such an irregular component should show the features typical of correlated flows undergoing turbulent evolution, characterized by a minimum and a maximum scale of turbulence, $\ell_{\rm min}$ and $\ell_{\rm max}$, respectively. The particles scatter off the magnetic irregularities and change their pitch angle θ, but not their velocity. The pitch angle scattering is principally dominated by the inhomogeneities with scales of the order of the Larmor radius, i.e. by resonance, providing an effective mechanism of isotropization as long as $r_L < \ell_{\rm max}$.

Our simulation of such an irregular behavior is based on the following approach. The turbulent magnetic field $\vec{B}(\vec{r})$ satisfies two main requirements: i) it is a zero-mean field $\langle \vec{B}(\vec{r})\rangle = 0$ with ii) non-vanishing fluctuations $\langle \vec{B}^2(\vec{r})\rangle = B_{\rm rms} > 0$. Let $\vec{k}$ be the wave vectors with modulus k, power spectrum $\mathcal{P}(k) \propto k^{-5/3}$ and amplitudes $\vec{B}(\vec{k})$ of its Fourier modes following the Kolmogorov spectrum $|\vec{B}(\vec{k})|^2 \propto k^{-11/3}$: such a field defines a turbulent Kolmogorov 3D magnetic field[2]. In the Fourier space the wave vectors satisfy $\frac{2\pi}{\ell_{\rm max}} \leq k \leq \frac{2\pi}{\ell_{\rm min}}$, where the correlation length of the field is equal to [111]

$$\Lambda_c = \frac{1}{2}\ell_{\rm max} \frac{\gamma - 1}{\gamma} \frac{1 - (\ell_{\rm min}/\ell_{\rm max})^\gamma}{1 - (\ell_{\rm min}/\ell_{\rm max})^{\gamma-1}}, \tag{2.7}$$

where γ is the spectral index of the Kolmogorov spectrum. Such a turbulent field can be simulated by mean of two different approaches. The first approach, proposed by Giacalone and Jokipii [112, 113] considers the field as the sum of N_m modes, physically corresponding to the superposition of a finite number of plane waves:

$$\vec{B}(\vec{r}) = \sum_{n=1}^{N_m} A_n \hat{\varepsilon}_n e^{i\vec{k}_n \cdot \vec{r} + i\beta_n}, \tag{2.8}$$

[2]The spectral index is 8/3 and 5/3 for 2D and 1D magnetic fields, respectively.

where $\hat{\varepsilon}_n = \cos\alpha_n \hat{x}_n + i \sin\alpha_n \hat{y}_n$ and the amplitude A_n of the $n-$th plane wave is given by

$$A_n^2 = \mathcal{A} B_{\rm irr}^2 G(\vec{k}), \tag{2.9}$$

with

$$G(\vec{k}) = \frac{\Delta V_n}{1 + (k\Lambda_c)^\gamma}, \quad \Delta V_n = 4\pi k^2 \Delta k, \quad \mathcal{A} = \left(\sum_{n=1}^{N_m} G(\vec{k}_n) \right)^{-1}. \tag{2.10}$$

In this last equation, the index γ is equal to 11/3, 8/3 and 5/3 for 3D, 2D and 1D turbulent magnetic fields, respectively. The direction of the $n-$th wave vector $\hat{k}_n$ is randomly chosen: the unit vectors $\hat{x}_n$ and $\hat{y}_n$ are chosen in order to form an orthogonal basis with $\hat{k}_n$ and the real numbers α_n and β_n represent random polarizations and phases, respectively. For practical applications, the spacing Δk between $k_{\rm min} = \frac{2\pi}{\ell_{\rm max}}$ and $k_{\rm max} = \frac{2\pi}{\ell_{\rm min}}$ should be constant in logarithmic scale and the number of modes N_m should be large enough to obtain the expected results in the small-angle regime. The main advantage of such an approach is the definition of the turbulent field at any point in space with arbitrary precision at the price of a much slower computation than other methods.

The second method, due to Casse, Lemoine and Pelletier [114], makes an extensive use of fast Fourier transforms (FFT) and it is briefly described here for the sake of completeness. The field is precomputed on a three dimensional lattice in the real space and the coordinates of its wave vectors correspond to the positions of vertices on the reciprocal grid in the Fourier space. The resulting field is therefore computed with FFT and can be interpolated at any point of the space. A lattice of 2^8 vertices per side is sufficient for applications [114, 115] and it can be periodically repeated in real space to cover regions of the order of hundred of Mpc. The main advantage of this method is the quick computation of the magnetic field, although it suffers of i) limitation in the allowed dynamic range, ii) interpolation of the magnetic field on scales smaller than the cell size and iii) the periodicity on the scale $\ell_{\rm max}$. Such limitations imply that the results of the FFT method obtained for Larmor radii much smaller than the cell size, i.e. $r_L << 1/N_{\rm cell}$, or much bigger than the periodicity scale, $r_L >> 1$, cannot be trusted, since these regimes are likely to be dominated by systematic effects related to the discreteness or to the periodicity [114].

Where not otherwise specified, in the following we will make use of the isotropic model, although the simulation of the composite model is also allowed by our code. Moreover, we will consider a total magnetic field $\vec{B} = \vec{B}_{\rm tot} = \vec{B}_{\rm reg} + \vec{B}_{\rm irr}$ and, following Ref. [114], we define the turbulence level by

$$\eta = \frac{\langle \vec{B}^2 \rangle}{B_{\rm reg}^2 + \langle \vec{B}^2 \rangle}. \tag{2.11}$$

Such a quantity is useful to investigate the diffusion properties of particles in magnetic fields and it is strictly related to the values of the diffusion coefficients, as we will see further in this section.

Note. Giacalone and Jokipii have introduced two models for the turbulent magnetic field, namely isotropic and composite. In the isotropic model, random polarizations α_n and phases β_n are chosen in the interval $[0, 2\pi]$. Conversely, in the composite model, representing the superposition of a slab component associated with Alfvén waves and a two-dimensional component generally known as "structures", the field is given by $\vec{B}_{\rm irr}(x, y, z) = \vec{B}_{\rm irr}^{\rm 1D}(z) + \vec{B}_{\rm irr}^{\rm 2D}(x, y)$, where $\gamma = 5/3$ and $\Delta V_n = \Delta k$ for the 1D component, and $\alpha_n = \pi/2$ is kept fixed, $\Delta V_n = 2\pi k \Delta k$ and $\gamma = 8/3$ for the 2D component.

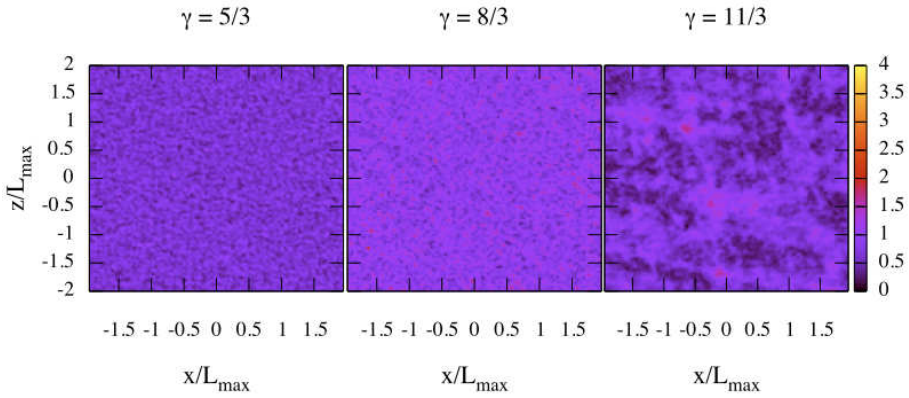


Figure 2.3: 2D slices of three random realizations of a Kolmogorov 3D turbulent magnetic field, for three different values of the scaling index, namely $\gamma = 5/3$, $\gamma = 8/3$ and $\gamma = 11/3$. We have simulated the turbulent field according to the Giacalone-Jokipii 3D isotropic approach.

In Fig. 2.3 we show the HERMES simulation of a random 3D realization of such a turbulent field for three different values of the spectral index in the Kolmogorov spectrum, namely $\gamma = 5/3$, $\gamma = 8/3$ and $\gamma = 11/3$. The formation of large-scale structures is evident for increasing values of γ.

As a consistency check between the goodness of our simulations and those reported in literature, we have followed the trajectories of several thousands of particles with different magnetic rigidity $\rho = r_L k_{\min}$ and we have recorded the evolution of their pitch angles with time. Let $\mu(t)$ be the cosine of the pitch angle $\theta(t)$, i.e. the direction of the velocity with respect to the mean field, and let us investigate the properties of $\mu(t)$. The most simple correlation study can be performed by mean of the autocorrelation function

$$C(\tau) \equiv \frac{\langle \mu(t+\tau)\mu(t) \rangle}{\langle \mu^2(t) \rangle}. \tag{2.12}$$

The average depends on the assumptions about the motion: in chaos theory, it is performed over the phase space subset of chaotic motions, whereas by assuming ergodicity, it corresponds to the temporal average. In Fig. 2.4 we show such an autocorrelation function for different values of the rigidity, ranging from 0.01 to 1 (color coding in each panel), and different turbulence levels, namely 0.1 (left panel) and 0.5 (right panel). When $C(\tau)$ approaches zero, the underlying dynamics is defined to be uncorrelated[3] and the value τ_c such that $C(\tau_c) = 0$ is defined as decorrelation time. It is evident that the decorrelation time increases for decreasing rigidity and decreases for increasing turbulence level, as intuitively expected. Moreover, for time larger than the decorrelation time, the autocorrelation is zero and the dynamics is stationary and totally uncorrelated, i.e. equivalent to the motion in a Gaussian field. Such results are in agreement with those reported, for instance, in Ref. [114]. If we define the scattering time

$$\tau_s \equiv \int_0^\infty C(\tau) d\tau, \tag{2.13}$$

it can be shown that $D_\parallel = \frac{1}{3} v \tau_s^2$ and $D_\perp = D_\parallel / \left(1 + (\omega_L \tau_s)^2\right)$, being ω_L the Larmor frequency [114].

[3]This is true only if the dynamical system generating the time series is linear, otherwise the autocorrelation analysis may give rise to spurious results. A more general approach to the study of correlations, valid also for nonlinear time series, is the estimation of the average mutual information [116].

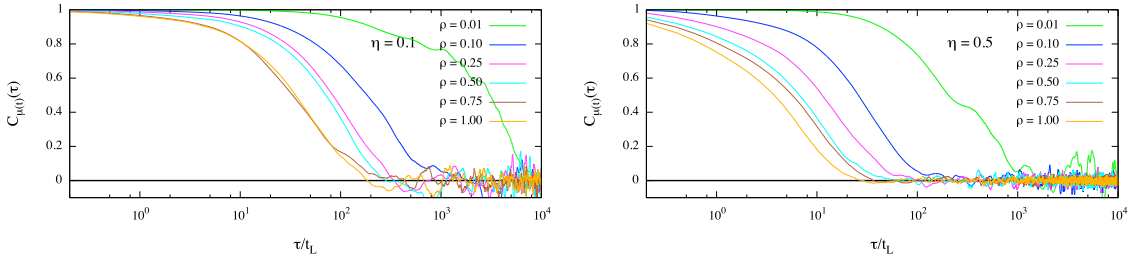


Figure 2.4: Consistency check of the HERMES simulation code. The autocorrelation function of the pitch angle cosine is shown as a function of time τ (in units of Larmor time $t_L = r_L/c$) for increasing values of the rigidity $\rho = r_L k_{\min}$ and two different values of the turbulence level, $\eta = 0.1$ (left panel) and $\eta = 0.5$ (right panel). Results are in agreement with those reported, for instance, in Ref. [114].

Estimation of diffusion coefficients. By starting from the diffusion equation, the Green-Kubo formalism allows an analytical estimation of the coefficients $\mathbb{D}_{ij}$ [117]. This formalism enables us to express transport coefficients in terms of time integrals of the velocity autocorrelation functions $R_{ij}(t) = \langle v_i(0) v_j(t) \rangle$, where $\langle ... \rangle$ denotes the ensemble average taken over an isotropic distribution of many particles:

$$\mathbb{D}_{ij} = \int_0^\infty R_{ij}(t) dt. \tag{2.14}$$

Let us assume that the decorrelations of velocity are modulated by exponential factors [118, 119]

$$R_{xx}(t) = R_{yy}(t) = \frac{c^2}{3} \cos(\omega_L t) e^{-t/\tau_\perp} \tag{2.15}$$

$$R_{yx}(t) = -R_{xy}(t) = \frac{c^2}{3} \sin(\omega_L t) e^{-t/\tau_A} \tag{2.16}$$

$$R_{zz}(t) = \frac{c^2}{3} e^{-t/\tau_\parallel}, \tag{2.17}$$

where $\omega_L = c/r_L$ is the Larmor frequency and $\tau_\perp, \tau_\parallel$ and τ_A are the decorrelation timescales corresponding to the different diffusion components. The resulting diffusion coefficients read

$$D_\perp = \frac{c r_L}{3} \frac{\omega_L \tau_\perp}{1 + (\omega_L \tau_\perp)^2} \tag{2.18}$$

$$D_\parallel = \frac{c r_L}{3} \omega_L \tau_\parallel \tag{2.19}$$

$$D_A = \frac{c r_L}{3} \frac{\omega_L \tau_A}{1 + (\omega_L \tau_A)^2}. \tag{2.20}$$

However, although such an approach is very elegant, i) there is no general theory providing the decorrelation timescales, and ii) the implicit assumption of such calculations is the small departure from the helical trajectories, which is no longer adequate for high turbulence levels [119]. A valid alternative is to estimate $D_\perp$ and $D_\parallel$ by means of numerical simulations, estimating the asymptotic rate of increase of the mean squared displacements in each direction, namely

$$D_\parallel = \lim_{\Delta t \longrightarrow \infty} \frac{\langle (\Delta z)^2 \rangle}{2\Delta t}, \quad D_\perp = \lim_{\Delta t \longrightarrow \infty} \frac{\langle (\Delta x)^2 \rangle}{2\Delta t} = \lim_{\Delta t \longrightarrow \infty} \frac{\langle (\Delta y)^2 \rangle}{2\Delta t}.$$

In Fig. 2.5 two important results obtained in Ref. [114] are shown. In the left panel, the ratio $D_\perp/D_\parallel$ is shown as a function of rigidity, for different values of the turbulence level η, putting in evidence

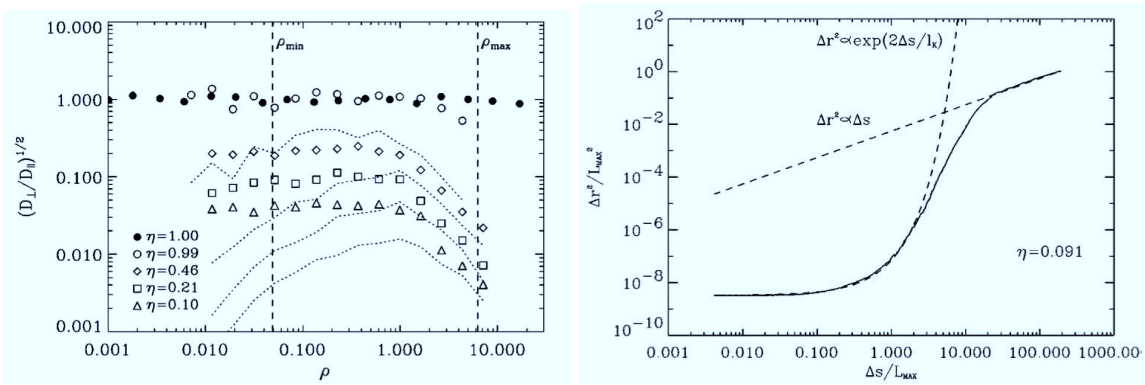


Figure 2.5: **Left:** The square root of $D_\perp/D_\parallel$ as a function of rigidity, for different values of the turbulence level η [114]. Symbols indicate the values obtained from MC simulations, vertical dashed lines indicate the range of validity of FFT approach and the dotted curves correspond to the classical scattering results, the latter diverging significantly from numerical results. **Right:** The square of the separation distance between two field lines as a function of the curvilinear abscissa s along the field line. The exponential divergence followed by the diffusion regime is clearly identified. The transition between these two regimes occurs when s is of the order of $\ell_{\max}$.

the disagreement between the numerical simulations and expected classical result, as well as the strong dependence on the value of η. In the right panel is shown the average displacement $\langle \Delta r^2 \rangle$ of two field lines as a function of the curvilinear abscissa s: the transition from a chaotic dynamics, clearly identified by the exponential divergence of $\langle \Delta r^2 \rangle$ characterized by a Lyapunov exponent l_K, to the diffusive regime, identified by the scaling $\langle \Delta r^2 \rangle \propto \Delta s$.

From the scaling analysis of the diffusion coefficients and the Lyapunov exponents with the turbulence level, it has been conjectured that the relation

$$D_\perp = \eta^{2.3\pm0.2} D_\parallel \tag{2.21}$$

holds for almost one order of magnitude over η, [114]. For the classical estimation of the ratio $D_\perp/D_\parallel$, or other parameterizations as a function of the turbulence level or the r.m.s. strength of the turbulent field, we refer to Refs. [113, 119] and Refs. therein. The impact of such results on the Fermi acceleration mechanisms in the case of galactic supernovae remnants, superbubbles and jets in extragalactic sources is described in detail in Ref. [114].

2.5 Simulating the Galactic magnetic field

In the following, we will present the models included in HERMES for the simulation of nuclei propagation in our Galaxy.

In spiral galaxies, the turbulent component of the magnetic field is almost always strongest within the spiral arms and thus follows the distribution of cool gas and dust, whereas the regular component is generally weak within spiral arms, except for rare cases like M51 with strong density waves. However, the regular field also extends far into the inter-arm regions. The regular field in M31 is nearly aligned with the spiral arms forming the bright "ring" of emission seen in this galaxy [101]. The total field is expected to be more intense within the spiral arms where the random field dominates.

Observations suggest that the large-scale spiral field produce an halo, extending outside the galactic disks. In cylindrical coordinates, the distribution of the magnetic field $B(\rho, \phi, z)$ in the galaxy can be described by the product of three separated components, related to pure radial dependence $R(\rho)$, spiral "winding" modulation $S(\rho, \phi)$, and halo extinction $H(z)$, respectively. Several models have been

proposed to describe the regular component of the magnetic field in our galaxy. In the following we will briefly describe the most common ones, which are also the models included in HERMES.

The structure of the magnetic field obtained by dynamo mechanisms can be described by modes of different azimuthal symmetry in the disk, and vertical symmetry perpendicular to the disk plane: bisymmetric (BSS) or axisymmetric (ASS), depending on π or 2π symmetry, respectively. Along the vertical dimension, the field can change direction while traversing the disk plane (odd or A-parity) or keep it fixed (even or S-parity). Thus, the possible patterns of the spiral field are four, and we will indicate them with the notation BSS-S, BSS-A, ASS-S and ASS-A.

Let us assume a Cartesian space centered at the Galactic Center (GC) of coordinates (0,0,0), and let $(-R_{\rm sun}, 0, 0)$ the position of the sun, with $R_{\rm sun} = 8.5$ kpc. The cylindrical coordinates are $(\rho, \phi, z) \equiv (\sqrt{x^2+y^2}, \tan y/x, z)$. Below, we will briefly describe[4] some common galactic magnetic field models proposed by Stanev [121], Harari, Mollerach and Roulet (HMR) [122], and Tinyakov and Tkachev [123]. For the sake of completeness, we refer to Refs. [124–128] for other models describing the galactic magnetic field.

Stanev model

From the investigation of the current data available in the early 1990s, Stanev proposed the following functions to model the field:

- Radial dependence:

$$R(\rho) = \begin{cases} B_0 \frac{R_{\rm sun}}{\rho} & \rho > \rho_0 \\ B_0 \frac{R_{\rm sun}}{\rho_0} & \rho \leq \rho_0 \end{cases}, \tag{2.22}$$

  where $B_0 = 3\ \mu$G and $\rho_0 = 4$ kpc. The magnetic field is 6.4 μG in the central region of the Galaxy (at $\rho < 4$ kpc), and decreases like $1/\rho$ at larger ρ values.

- Spiral modulation:

$$S_{\rm BSS}(\rho, \phi) = \cos\left(\phi - \beta \ln \frac{\rho}{\rho_1} - \phi_{\rm sun}\right), \tag{2.23}$$

$$S_{\rm ASS}(\rho, \phi) = \left|\cos\left(\phi - \beta \ln \frac{\rho}{\rho_1} - \phi_{\rm sun}\right)\right|, \tag{2.24}$$

  where $\beta = 1/\tan p$, being p the pitch angle. The parameters have been chosen according to observations, with $p = -170°$, $\phi_{\rm sun} = \pi$ and $\rho_1 = 10.55$ kpc, the latter related to the distance to the two possible close reversals (in the BSS model) or locations of the field zeroing (in the ASS model), which occur in the Galactic Center direction at 0.5 kpc and ∼3 kpc.

- Halo extinction:

$$H_{\rm S}(z) = \begin{cases} \exp\left(-\frac{|z|}{z_1}\right) & |z| < z_0 \\ \exp\left(-\frac{|z|}{z_2} + \frac{z_0}{z_2} - \frac{z_0}{z_1}\right) & |z| \geq z_0 \end{cases}, \tag{2.25}$$

$$H_{\rm A\pm}(z) = \pm {\rm sign}(z) H_{\rm S}(z), \tag{2.26}$$

  for the even and odd parity, respectively, where $z_0 = 0.5$ kpc, $z_1 = 1$ kpc and $z_2 = 4$ kpc.

[4] We would like to remark that such a description is inspired by the nice review written by Serguei Vorobiov, Mustafa Hussain and Darko Veberic [120].

HMR model

The model proposed by Harari, Mollerach and Roulet (HMR), is a modification of the model proposed by Stanev, smoothed out in order to avoid the discontinuities of the field and its derivatives:

- Radial dependence:

$$R(\rho) = B_0 \frac{R_{\rm sun}}{\rho} \tanh^3 \frac{\rho}{\rho_2}, \tag{2.27}$$

  where $B_0 = 3\ \mu$G, $\rho_2 = 2$ kpc and $R(\rho)$ goes smoothly to zero at the Galactic Center.

- Spiral modulation:

$$S_{\rm ASS}(\rho, \phi) = \cos^2 \left(\phi - \beta \ln \frac{\rho}{\rho_1} - \phi_{\rm sun} \right), \tag{2.28}$$

  where $\rho_1 = 10.55$ kpc.

- Halo extinction:

$$H_{\rm S}(z) = \frac{1}{2} \left(\frac{1}{\cosh(z/z_1)} + \frac{1}{\cosh(z/z_2)} \right), \tag{2.29}$$

$$H_{\rm A\pm}(z) = H_{\rm S}(z) \tanh(z/z_3), \tag{2.30}$$

  for the even and odd parity, respectively, where $z_1 = 0.3$ kpc, $z_2 = 4$ kpc and $z_3 = 20$ pc.

Tinyakov and Tkachev model

In their model, Tinyakov and Tkachev have considered only the spiral disk component with bisymmetric field and different parity with respect to the $z-$axis. Moreover, they have provided a new expression for the in-plane field $R(\rho)S(\rho, \phi)$ in terms of parameters related to the local field, including an additional parameter related to the distance d to the nearest field reversal (being negative if it occurs in the direction of the Galactic Center, being positive otherwise):

- Radial dependence:

$$R(\rho) = B_0 \frac{R_{\rm sun}}{\rho \cos \phi_{\rm add}}, \tag{2.31}$$

  where $B_0 = 1.4\ \mu$G and $d = -0.5$ kpc and $\phi_{\rm add} = \beta \ln (1 + d/R_{\rm sun}) - \pi/2$.

- Spiral modulation:

$$S_{\rm BSS}(\rho, \phi) = \cos \left(\phi - \beta \ln \frac{\rho}{R_{\rm sun}} + \phi_{\rm add} - \phi_{\rm sun} \right). \tag{2.32}$$

- Halo extinction:

$$H_{\rm S}(z) = \exp\left(-|z|/z_1\right), \tag{2.33}$$

$$H_{\rm A\pm}(z) = \pm {\rm sign} H_{\rm S}(z), \tag{2.34}$$

  for the even and odd parity, respectively, where $z_1 = 1.5$ kpc.

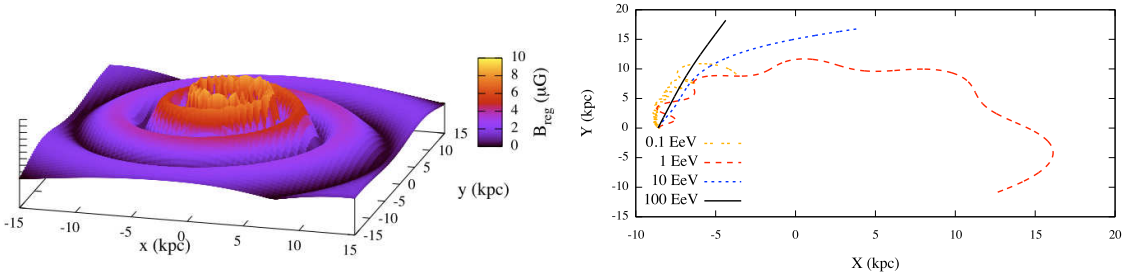


Figure 2.6: **Left:** HERMES simulation of the HMR model for the regular component of the magnetic field in our galaxy (at $z = 0$), where the color indicates the intensity of the field. **Right:** Two-dimensional projection of the corresponding backtracked trajectories of UHECR for different values of the rigidity E/Z, ranging from 10^{17} eV (0.1 EeV) to 10^{20} eV (100 EeV)

In the left panel Fig. 2.6 we show the HERMES simulation of the HMR model for the regular component of the magnetic field in our galaxy (at $z = 0$). In the right panel of the same figure, the two-dimensional projection of the corresponding backtracked trajectories of UHECR are shown for different values of the rigidity E/Z, ranging from 10^{17} eV (0.1 EeV) to 10^{20} eV (100 EeV). It is evident that at the lower energy particles tend to move along helical trajectories around the field lines, whereas for increasing energy particle tend to be less deflected.

Note. In the forward tracking approach, a particle with charge Z is propagated forward in time and its trajectory is followed. However, if we are interested in investigating the original direction of the particle, i.e. the direction that the particle had at the beginning of its travel from the source to the Earth, we should have to follow the path of the particle back in time. A backtracked trajectory is the path traveled by the antiparticle, and it is obtained by substituting the charge Z with the charge $-Z$ in the equations of motion.

By mean of the backtracking technique, it is possible to build the maps of the average deflections experienced by UHECR by estimating the angular distance between the original direction of the particle and the direction observed at Earth. In Fig. 2.7 we show the result of such a procedure for HMR (left) and Tinyakov-Tkachev (right) models obtained by recent simulations [125]. It is evident that deflections are strongly model-dependent and that different regions of the sky behave differently, making more difficult the source detection for events with energy smaller than 10^{20} eV.

For what concerns the irregular component of the GMF, as previously discussed, observations suggest a r.m.s. intensity of the order of the regular one, although no precise information is currently available. In Fig. 2.8 we show the HERMES simulation of the trajectory (forward in time) of a particle with E/Z ranging from 10^{17} eV to 10^{20} eV (left, middle and right panel, respectively), propagating in a magnetic field with an uniform component of intensity $B_0 = 3$ μG, parallel to the $z-$axis and a 3D turbulent component, characterized by maximum coherence length $\ell_{\max} = 100$ pc, r.m.s. strength $\langle B^2_{\rm irr}\rangle = 1$ μG and Kolmogorov index $\gamma = 11/3$. The turbulence level is $\eta = 0.1$. At the lowest energy the particle undergoes a brownian motion, being the Larmor radius of the order of turbulence scale $\ell_{\max}$, whereas for increasing energy the particle only partially "feels" the turbulent component. At the highest energy the particle is subjected to the regular component only.

In Fig. 2.9 we show some realizations of backtracked trajectories of nuclei with random direction in the GMF, including both regular (HMR BSS-S) and irregular (Kolmogorov 3D) components, obtained from

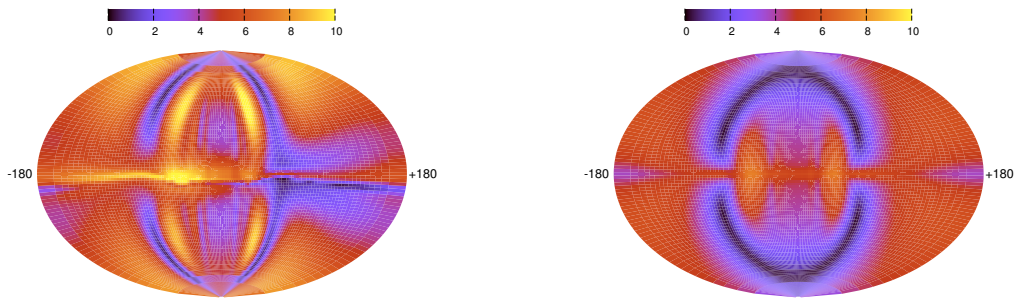


Figure 2.7: Deflection maps for HMR (left panel) and Tinyakov-Tkachev (right panel) models of the GMF, for a rigidity of 4×10^{19} V. The deflection scale is in degree, and the maps refer to the direction as observed at the Earth [125].

our simulations. Nuclei have been backtracked from the Earth (black diamond) up to a galactocentric distance of 20 kpc. Results for three different values of the rigidity E/Z (10^{18} eV, 10^{19} eV and 10^{20} eV) and two different values of the r.m.s. strength of the irregular magnetic field (1 μG and 6 μG, the latter of the order of the regular component) are shown. At the lowest energy, particles suffers of many turbulent deflections, whereas at the highest energy both fields have almost no impact on the trajectories. At $E = 10^{19}$ eV the irregular field has a small impact on the trajectories for $\langle B_{\rm irr}^2 \rangle = 1$ μG but a non-negligible impact for $\langle B_{\rm irr}^2 \rangle = 6$ μG, where the reconstruction of the original direction is impossible, in practice. For the application of such a procedure to real events observed by AGASA (as well as by other experiments as Haverah Park, Volcano Ranch and Yakutsk) we refer to Refs. [123–125, 129].

The deflection $\delta_{\rm reg}$ due to the regular component of the magnetic field can be parametrized [130] by

$$\delta_{\rm reg} \simeq 3.2^\circ \frac{10^{20}\ {\rm eV}}{E/Z} \frac{D}{3\ {\rm kpc}} \frac{B_{\rm reg}}{2\ \mu{\rm G}}, \tag{2.35}$$

where D is the distance traversed by the particle across the field. The deflection $\delta_{\rm irr}$ due to the irregular component of the magnetic field can be estimated by assuming that the particle undergoes a brownian motion at the scale of the coherence length Λ of the field and that the ratio D/Λ provides an estimation of the number of magnetic regions traversed [115, 130]:

$$\delta_{\rm irr} = \frac{1}{\sqrt{2}} \frac{ZeB_{\rm rms}}{E} (D\Lambda)^{\frac{1}{2}} \simeq 0.6^\circ \frac{10^{20}\ {\rm eV}}{E/Z} \frac{B_{\rm rms}}{4\ \mu{\rm G}} \left(\frac{D}{3\ {\rm kpc}}\right)^{\frac{1}{2}} \left(\frac{\Lambda}{50\ {\rm pc}}\right)^{\frac{1}{2}}, \tag{2.36}$$

being $B_{\rm rms} = \langle B_{\rm irr}^2 \rangle$.

2.6 Simulating the extragalactic magnetic field

The current knowledge of the extragalactic magnetic field is even lower than for the galactic case. The use of extensive numerical simulations, based on MHD evolution of the magnetic field, represents a fundamental tool for the comparison with observations. It is generally assumed that EMF is coherent for scales above few Mpc, and it is turbulent for scales below this limit, with a distribution following that of matter.

We simulate the propagation of charged nuclei in the turbulent component of the EMF with the same approach adopted in the case of the random component of the GMF, by considering, of course,

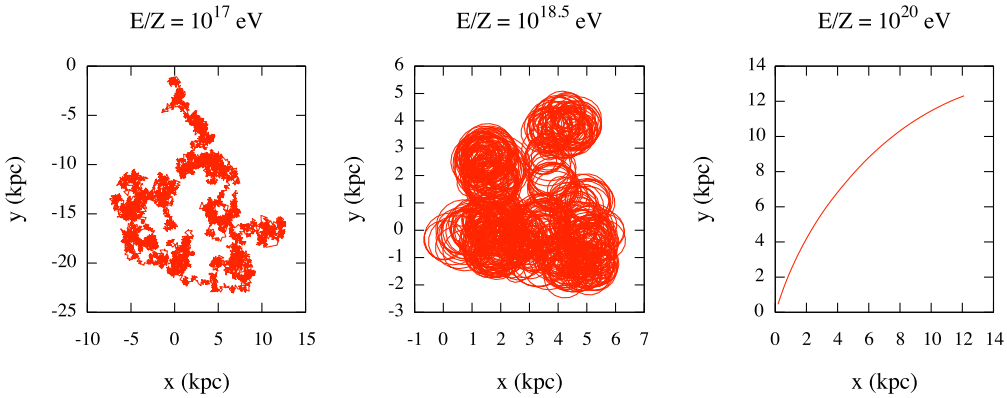


Figure 2.8: A random realization of nuclei trajectories in a uniform magnetic field ($B_0 = 3\ \mu$G, parallel to the $z-$axis) plus a Kolmogorov 3D turbulent field ($\gamma = 11/3$), for three different values of the ratio E/Z, namely 10^{17} eV, $10^{18.5}$ eV and 10^{20} eV. We have simulated the turbulent field ($\ell_{\rm max} = 100$ pc, $\langle B^2_{\rm irr}\rangle = 1\ \mu$G, $\eta = 0.1$) according to the Giacalone-Jokipii 3D isotropic approach.

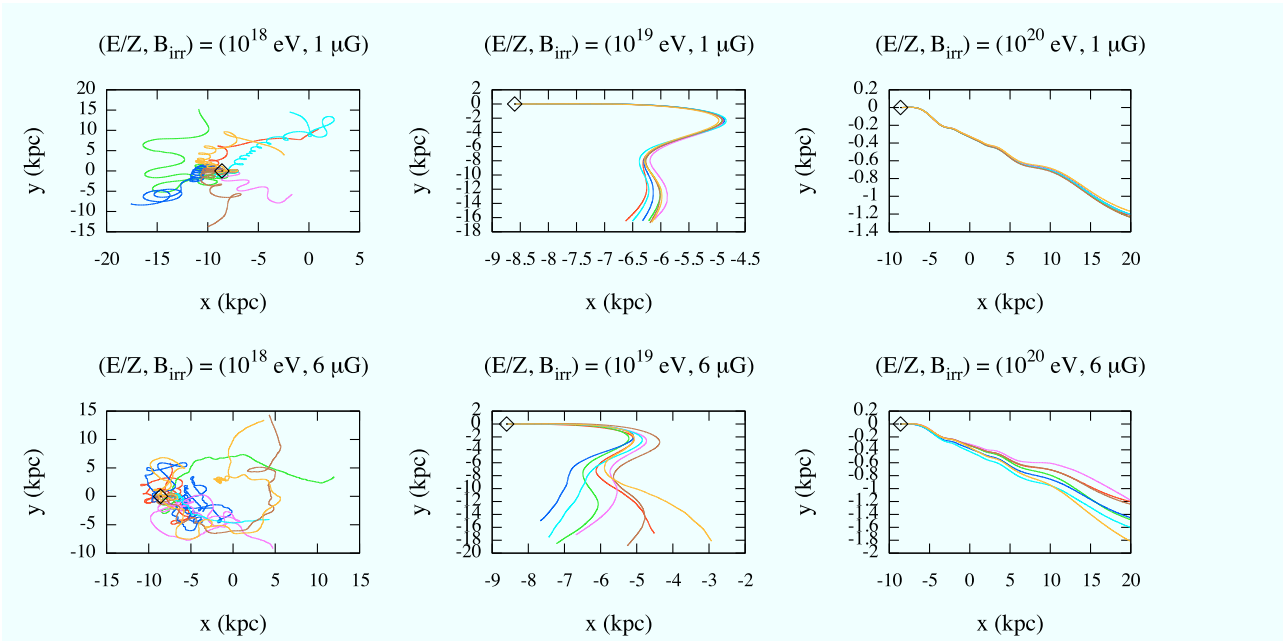


Figure 2.9: Backtracked trajectories (color coded) of nuclei in the GMF, obtained with HERMES, including both regular (HMR BSS-S) and random realizations of the irregular (Kolmogorov 3D) components. Nuclei have been backtracked from the Earth (black diamond) up to a galactocentric distance of 20 kpc. Results for three different values of the rigidity E/Z (10^{18} eV, 10^{19} eV and 10^{20} eV) and two different values of the r.m.s. strength of the irregular magnetic field (1 μG and 6 μG) are shown.

that different scales are involved. The deflection $\delta_{\rm irr}$ due to the irregular component of the EMF can be estimated similarly to the case of the GMF, by considering the appropriate coherence length and by neglecting energy loss processes [40]:

$$\delta_{\rm irr} =\simeq 0.8^\circ \frac{10^{20}\ {\rm eV}}{E/Z}\frac{B_{\rm rms}}{1\ {\rm nG}}\left(\frac{D}{10\ {\rm Mpc}}\right)^{\frac{1}{2}}\left(\frac{\Lambda}{1\ {\rm Mpc}}\right)^{\frac{1}{2}}. \tag{2.37}$$

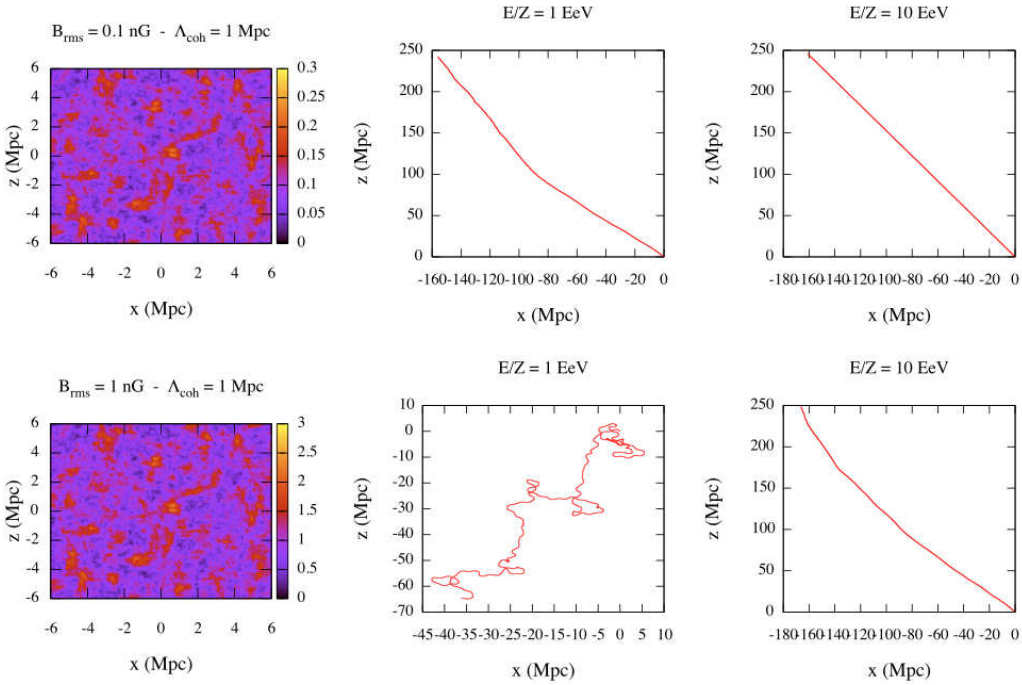


Figure 2.10: HERMES simulations for the diffusion of charged nuclei in the EMF (energy loss processes have been switched off). **Left:** Small region of the two-dimensional projection of a 3D turbulent magnetic field characterized by a coherence length $\Lambda = 1$ Mpc, $\gamma = 11/3$, $B_{\rm rms} = 0.1$ nG (top) and $B_{\rm rms} = 1$ nG (bottom); the random realization is the same in both cases. **Middle**: Trajectories of a particle with $E/Z = 10^{18}$ eV for $B_{\rm rms} = 0.1$ nG (top) and $B_{\rm rms} = 1$ nG (bottom). **Right**: Trajectories of a particle with $E/Z = 10^{19}$ eV for $B_{\rm rms} = 0.1$ nG (top) and $B_{\rm rms} = 1$ nG (bottom).

In Fig. 2.10 we show two-dimensional projections for random realizations of the trajectory of a charged particle starting from the center $(0, 0, 0)$ of the reference frame, for different values of the r.m.s. strength $B_{\rm rms} = \langle B^2_{\rm irr} \rangle$ of the magnetic field (coherence length $\Lambda = 1$ Mpc) and different rigidities, obtained from our simulations. For $E/Z = 10^{18}$ eV and $B_{\rm rms} = 0.1$ nG, the particle traverses distances of the order of tens Mpc before experiencing significant deflections, whereas for $E/Z = 10^{19}$ eV the propagation is rectilinear, in practice. For increasing values of $B_{\rm rms}$, for instance of an order of magnitude, an energy of $E/Z = 10^{19}$ eV is no more sufficient to ensure a rectilinear propagation over hundreds Mpc.

Results discussed so far have been obtained by neglecting the energy loss processes involved during the propagation in the turbulent magnetic field. However, such an approximation introduces an underestimation of the instantaneous diffusion coefficients, as shown in Fig. 2.11 [131]. In the left panel, the propagation of protons in a turbulent field characterized by $B_{\rm rms} = 10$ nG and $\Lambda = 1$ Mpc is considered, and the instantaneous diffusion coefficients are estimated as a function of elapsed time, for several values of the energy, ranging 10^{16} to 10^{20} eV: here energy loss have been switched off. In the top-right panel of the same figure, the same results are shown after switching on the energy loss processes: the difference with the previous case is evident, because, in this second case, protons may not survive long enough to enter the diffusion regime. In the bottom-right panel of the same figure, the diffusion coefficient is shown as a function of energy, under the same physical assumptions about the propagation: it is evident a transition between two asymptotic behaviors, $E^{1/3}$ and E^2, for increasing energy scale. It is worth remarking that the resulting diffusion coefficients are not in agreement with Bohm's theory, describing the diffusion of plasma across a magnetic field.

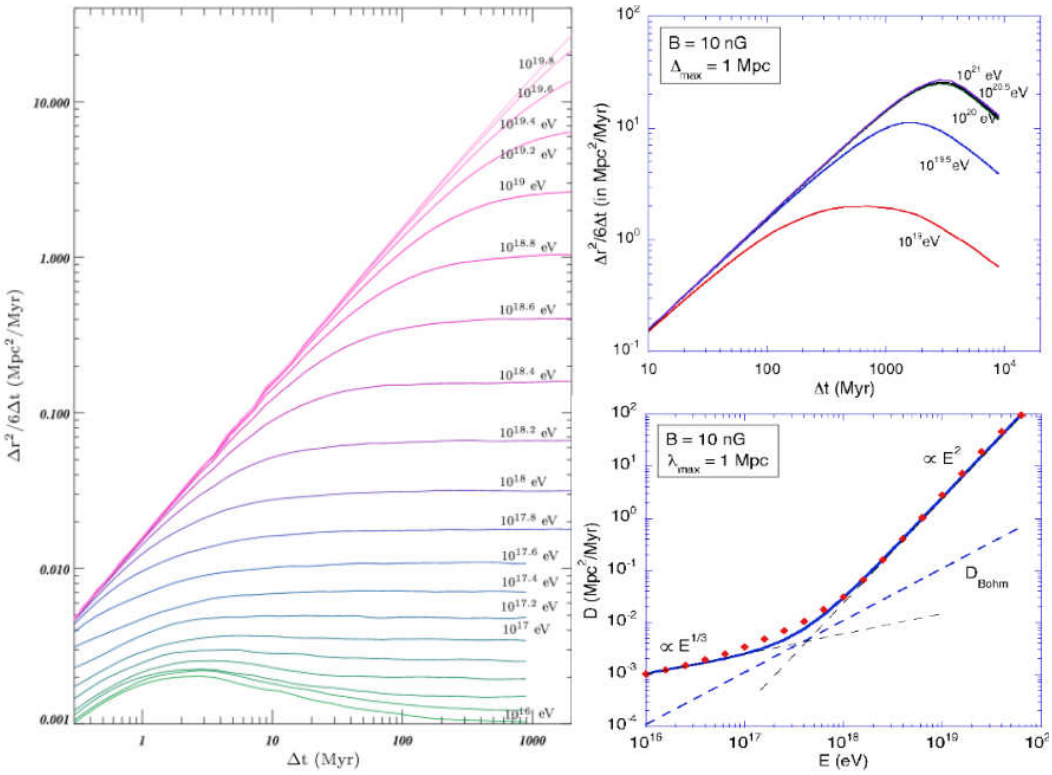


Figure 2.11: **Left:** Instantaneous effective diffusion coefficient, $\Delta r^2/(6\Delta t)$, as a function of elapsed time, for protons of various energies, from 10^{16} to 10^{20} eV. The turbulent magnetic field is characterized by $B_{\rm rms} = 10$ nG and $\Lambda = 1$ Mpc. Energy losses have been switched off [131]. **Top-right:** Same of left panel, but with energy losses switched on. **Bottom-right:** Diffusion coefficient of protons as a function of energy, for the same turbulent magnetic field. The red dots are the computation results, with asymptotic behaviors indicated by the thin dashed lines. The Bohm diffusion coefficient is also shown for comparison (dashed line) [131].

Impact on the spectrum of UHECR. In this small section we will briefly discuss the impact of EMF on the flux of UHECR measured at Earth. Recently, Berezinsky and Gazizov presented an analytic solution to diffusion equation for high energy cosmic rays in the expanding Universe [132], generalizing the Syrovatsky solution [133], valid for the case when energy losses and diffusion coefficient are time-independent.

In another recent paper, they have shown that the flux of UHECR at Earth at the lowest and at the highest energy depends on the assumptions about the magnetic field [134]. In particular, by assuming an evolution of the magnetic field, with the intensity of the field decreasing with time because of magnetic flux conservation, they have shown that some differences are present if a static of an expanding Universe is considered[5].

In Fig. 2.12 the energy spectra of extragalactic UHECR propagated to the Earth are shown, by varying the spectral index at the source, the density of sources and their evolution, the composition and the EMF as reported by [131]. In agreement with the result of Berezinsky and Gazizov, the overall effect of the EMF is to reduce the flux of UHECR at the lowest (below 10^{18} eV) and at the highest energy (above $10^{19.8}$ eV). For further details, we refer to the original papers.

[5]It is worth noticing that in the expanding Universe the source term in the diffusion equation, as well as the density of particles, depend explicitly on time because they are related to the comoving volume.

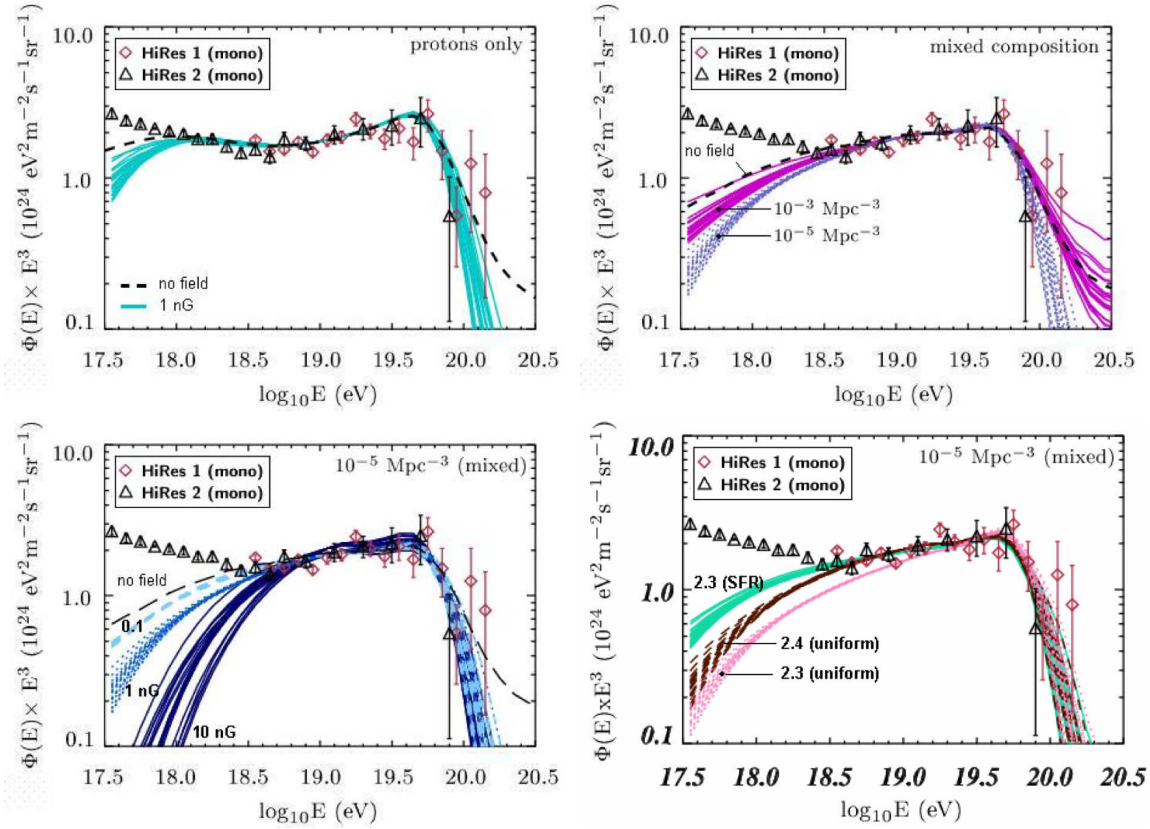


Figure 2.12: Spectra of extragalactic UHECR propagated to the Earth, compared with the flux measured by the Hires Collaboration. **Top-left:** Pure proton sources with injection spectrum $E^{-2.6}$ and source density 10^{-5} Mpc^{-3}. **Top-right:** Mixed source composition with injection spectrum $E^{-2.4}$, and EMF defined by $B_{\rm rms} = 1$ nG and $\Lambda = 1$ Mpc; two different source density, namely 10^{-5} Mpc^{-3} and 10^{-3} Mpc^{-3}, are considered. **Bottom-left:** Mixed source composition, source density 10^{-5} Mpc^{-3} and different r.m.s strength of EMF (injection index corresponding to a configuration of EMF is varied in order to better fit the data). **Bottom-right:** Mixed source composition, source density 10^{-5} Mpc^{-3}, $B_{\rm rms} = 1$ nG and different values of the injection index and source power evolution. For further details we refer to the original paper [131].

2.7 Conclusions

In this chapter we have introduced the Monte Carlo code (HERMES) that we have originally developed to simulate the propagation of nuclei in the Universe. We have shown the different modules composing HERMES and we have briefly presented their role in the framework.

Successively, we have discussed how HERMES simulates magnetic fields. In particular, we have treated the simulation of: i) the regular component of the Galactic magnetic field, showing the models included in our code (Stanev, Harari-Mollerach-Roulet, Tinyakov-Tkachev with ASS or BSS patterns); ii) the irregular component of both the Galactic and the extragalactic magnetic fields by using the Giacalone-Jokipii approach. Several simulations of nuclei propagating in a magnetized Universe and in our Galaxy have been provided, by varying the relevant parameters.

We have shown the impact of magnetic fields on the propagation of UHE nuclei, without considering the energy-loss processes relevant for a complete study. This is the main topic of the next chapter.

Imagination is more important than knowledge. Knowledge is limited. Imagination encircles the world.

A. Einstein

A frog in a well cannot conceive of the ocean.

Chuang Tzu

Chapter 3

Simulating the propagation of UHECRs II: energy-loss in the EBR

In the previous chapter, the general structure of HERMES, our Monte Carlo code for simulating the propagation of nuclei in a magnetized Universe, has been presented. In particular, we have discussed how HERMES simulates magnetic fields and the diffusion of UHE particles in the absence of energy-loss.

Within the present chapter, which should be considered complementary to the previous one, we will show the impact of energy-loss processes on the propagation of UHE nuclei, photons and neutrinos. First, we describe the cosmological framework chosen for HERMES and the parameterizations we have adopted for the extragalactic background radiations relevant to our study, namely the cosmic microwave, infrared/optical and radio backgrounds. Successively, we will define the parameterizations we have chosen for the cross sections of the interactions between propagating UHECRs and photons of the background radiation. We will discuss all the relevant energy-loss processes included in our simulator as the adiabatic loss (due to the expansion of the Universe), the pair and photo-pion production, and, in the particular case of heavy nuclei, the photo-disintegration processes.

The creation of secondary particles, produced by UHE nuclei undergoing pair and photo-pion production during their propagation, is also described. Complementary information can be found in Appendix A, where the relativistic kinematics of collisions and decays is treated with some detail. The development of the resulting UHECR cascade, including neutrinos and photons, will be briefly described to underline the complexity of simulating a realistic propagation. The comparison between HERMES and other codes available in the UHECR community will be performed and discussed.

Finally, we will conclude the chapter by presenting our result about the impact of cosmology and magnetic fields on the GZK horizon of UHE protons.

3.1 Modeling the cosmological framework

Motivated by up-to-date observations, we have chosen a general Friedmann's Universe, defined by a Friedmann-Robertson-Walker metric, to be the cosmological framework in HERMES. In the following, we will briefly describe such a framework, to provide the reader with the necessary tools to understand the parameterizations and the energy-loss equation adopted in our Monte Carlo code.

Friedmann's Universes

Let us consider the Einstein equation

$$R_{\mu\nu} - \frac{1}{2} g_{\mu\nu} R + \Lambda g_{\mu\nu} = \frac{8\pi G}{c^4} T_{\mu\nu}, \tag{3.1}$$

in the classical General Relativity framework to describe the gravitational field, where $R_{\mu\nu}$ and R are the Ricci tensor and scalar, respectively, $T_{\mu\nu}$ is the stress-energy tensor, $g_{\mu\nu}$ denotes the metric, G is the Newton gravitational constant and Λ is the cosmological constant. Under the assumptions of an isotropic and homogeneous Universe, we consider the Friedmann-Robertson-Walker (FRW) metric

$$ds^2 = c^2 dt^2 - a^2(t)\left[\frac{dr^2}{1-\kappa r^2} + r^2\left(d\theta^2 + \sin^2\theta d\phi^2\right)\right], \tag{3.2}$$

where $a(t)$ is the scale factor, such that $a(0) = 1$ is its present value, while the parameter κ accounts for the spatial curvature: $\kappa = -1$ denotes an open metric, $\kappa = 0$ a flat metric and $\kappa = 1$ a closed metric. Indeed, we consider the Universe as a perfect fluid with energy density ϱ and pressure p, described by the stress-energy tensor

$$T_{\mu\nu} = \left(\varrho + \frac{p}{c^2}\right) u_\mu u_\nu + p g_{\mu\nu},$$

where u_μ denotes the 4-velocity. Within the above assumptions, the Einstein equation leads to Friedmann equations

$$H^2 = \left(\frac{\dot{a}}{a}\right)^2 = \frac{8\pi G}{3}\varrho - \frac{\kappa c^2}{a^2} + \frac{\Lambda c^2}{3} \tag{3.3}$$

$$\dot{H} + H^2 = \frac{\ddot{a}}{a} = -\frac{4\pi G}{3}\left(\varrho + \frac{3}{c^2}p\right) + \frac{\Lambda c^2}{3} \tag{3.4}$$

where H is the time-dependent Hubble parameter. By assuming $\Lambda = 0$ and the spatial curvature $\kappa = 0$, the critical density is defined as $\varrho_c = 3H^2/8\pi G$ and it is used to obtain the well known dimensionless density parameter

$$\Omega = \frac{\varrho}{\varrho_c} = \frac{8\pi G}{3H^2}\varrho,$$

a suitable quantity adopted for the comparison of different cosmological models. The density parameter accounts for the matter and the energy in the Universe, and can be parametrized as the sum of different contributions. In the standard ΛCDM model, there are some contributions to Ω: Ω_b due to baryonic matter, Ω_c due to cold dark matter, Ω_Λ due to dark energy, Ω_r due to radiation and Ω_κ for the spatial curvature. If we define the redshift z by $1 + z = a^{-1}(t)$, the first Friedmann equation can be written in terms of z and of density parameters as

$$\frac{H^2(z)}{H_0^2} = \Omega_r(1+z)^4 + \Omega_M(1+z)^3 + \Omega_k(1+z)^2 + \Omega_\Lambda, \tag{3.5}$$

where $\Omega_M = \Omega_b + \Omega_c$ is the total density of matter and H_0 is the Hubble parameter at the present time. By taking into account that the radiation density contributes only in the early Universe, i.e. at high redshifts, whereas in practice it is negligible in the late Universe, the constraint $\Omega_M + \Omega_\kappa + \Omega_\Lambda = 1$ for the density parameters can be obtained from very general considerations. Finally, changes in the expansion rate of the Universe are described by the deceleration parameter

$$q(z) = -\frac{\ddot{a}}{aH^2} = \frac{H'(z)}{H(z)}(1+z) - 1, \tag{3.6}$$

from which

$$q(z=0) = q_0 = \frac{1}{2}\Omega_M - \Omega_\Lambda \tag{3.7}$$

at the present time. The parameter q_0 and the density parameters described above can be varied to reproduce very different cosmological models: in the next chapter, we will show the impact of cosmology on the GZK horizon of UHECR protons. However, it is worth remarking that we are under the assumptions of an isotropic and homogeneous Universe in the approximation of perfect fluid. For practical applications, it is worth noticing that a particle with energy $E(z)$ at redshift z, propagating through the Universe and not subjected to energy loss processes, will adiabatically lose its energy because of the expansion of the Universe (of course, by assuming a cosmological model where the Universe is expanding), and it will be observed with energy $E_0 = E/(1+z)$ at the Earth.

The values of all relevant parameters discussed so far, as the Hubble constant, the density of matter and energy, can be freely fixed in our simulator. In the following, we will briefly present the notation widely adopted in this chapter and in the next ones.

Cosmological units

A common notation to indicate the Hubble parameter at the present time is $H_0 = 100h$ km s^{-1} Mpc^{-1}, being h a dimensionless number which parametrizes our "ignorance", whose value ranges from 0.6 to 0.9. The inverse of H_0 defines the Hubble time $t_H \equiv H_0^{-1} = 3.09 \times 10^{17} h^{-1}$ s $= 9.78 \times 10^9 h^{-1}$ yr, whereas the Hubble distance is given by $D_H \equiv ct_H = 9.26 \times 10^{25} h^{-1}$ m $= 3000 h^{-1}$ Mpc.

Comoving distance and volume

If the ratio v/c is small enough, the approximation $z \approx d/D_H$ is allowed for an object at distance d from the observer, otherwise a more general formula should be used.

Following Ref. [135], we define the infinitesimal *comoving distance* δD_C as the distance (between two nearby objects in the Universe) which would be measured with rulers at the time they are being observed (the proper distance) divided by the ratio of the scale factor of the Universe then to now; it is the proper distance multiplied by $(1+z)$. As a consequence, the comoving distance between two objects remains constant with time if they are moving with the Hubble flow. The total comoving distance D_C is obtained by integrating over all the δD_C contributions between nearby events along the radial ray from $z = 0$ to the object:

$$D_C = D_H H_0 \int_0^z \frac{dz'}{H(z')}, \tag{3.8}$$

where $D_H H_0 = c$ because of our previous definitions. The comoving distance between two events at the same redshift but separated by an angle $\delta\theta$ on the sky, is $D_C = D_M \delta\theta$, where the *transverse comoving distance* D_M is defined by

$$D_M = \begin{cases} D_H \frac{1}{\sqrt{\Omega_k}} \sinh\left(\sqrt{\Omega_k} D_C/D_H\right) & \Omega_k > 0 \\ D_C & \Omega_k = 0 \\ D_H \frac{1}{\sqrt{|\Omega_k|}} \sin\left(\sqrt{|\Omega_k|} D_C/D_H\right) & \Omega_k < 0 \end{cases} . \tag{3.9}$$

The transverse comoving distance is equivalent to the proper motion distance, defined as the ratio of the actual transverse velocity (in distance over time) of an object to its proper motion (in radians per unit time).

Another important quantity for the study of number densities is the *comoving volume* V_C, the volume in which the number densities of non-evolving objects are constant with respect to the redshift, and it is calculated by multiplying the proper volume times $(1+z)^3$. The infinitesimal comoving volume in a solid angle $d\Omega$ and redshift interval dz is given by

$$dV_C = D_H H_0 \frac{(1+z)^2 D_A^2}{H(z)} d\Omega dz, \tag{3.10}$$

where $D_A = (1+z)^{-1} D_M = (1+z)^{-2} D_L$ is the angular diameter distance.

Luminosity distance

We define the *luminosity distance* D_L by the relationship between bolometric (i.e., integrated over all frequencies) flux S and bolometric luminosity L:

$$D_L \equiv \sqrt{\frac{L}{4\pi S}} = (1+z) D_M. \tag{3.11}$$

If we have to deal with differential flux S_ν and luminosity L_ν instead of bolometric quantities, a correction should be applied to the flux or luminosity because the redshifted object is emitting flux in a different band than that in which the observation is being performed. Such a correction depends on the spectrum of the observed object, and it can be neglected only if the object has spectrum $\nu L_\nu = \text{const}$. For any other spectrum, the differential flux S_ν is related to the differential luminosity L_ν by

$$S_\nu = (1+z) \frac{L_{(1+z)\nu}}{L_\nu} \frac{L_\nu}{4\pi D_L^2}, \tag{3.12}$$

where the ratio of luminosities equalizes the difference in flux between the observed and emitted bands, and the factor of $(1+z)$ accounts for the redshifting of the bandwidth.

Spectrum of UHECR and Evolution of sources

Let $Q(E)$ indicate the contribution of a CR source to the spectrum of UHECRs. Such a contribution is due to the spectrum of UHECR at the source, also known as *injection spectrum*, representing the number of particles injected per unit energy and time. The other key quantity is the *source luminosity* defined by

$$\mathcal{L} = \int_{E_{\min}}^{E_{\max}} Q(E) E dE, \tag{3.13}$$

quantifying the energy emitted from the source in terms of UHECR per unit time. Here, we are assuming that UHECR at the source can be produced from a minimum energy $E_{\min}$ to a maximum energy $E_{\max}$. There are some arguments predicting a power-law injection spectrum of both Galactic and extragalactic CRs [37, 43, 57]. Under such an assumption, we can rewrite the injection spectrum as a function of the source luminosity by $Q(E) = \mathcal{L}\mathcal{N}E^{-\gamma}$, being γ the injection index and $\mathcal{N}$ a normalization factor. The source luminosity may increase with redshift, as well as the comoving density of sources: in general such a *cosmological source evolution* depends on several factors, related to the class of astrophysical sources under consideration. If the source evolution is present, the luminosity should include an additional factor $\mathcal{H}(z) = (1+z)^m$, giving $\mathcal{L}(z) = \mathcal{H}(z)\mathcal{L}$. As we will see in the following, the source evolution factor can play a significant role for the study of the energy spectrum of UHECR at Earth. Thus, the injection spectrum becomes $Q(z, E) = \mathcal{H}(z) Q(0, E)$. We report in this section some common evolution factors useful for successive applications:

- Star formation rate (SFR) [136]:

$$\mathcal{H}_{\rm SFR}(z) = \begin{cases} (1+z)^{3.4} & z < 1, \\ 2^{3.7}(1+z)^{-0.3} & 1 < z < 4, \\ 2^{3.7} \times 5^{3.2}(1+z)^{-3.5} & z > 4; \end{cases} \tag{3.14}$$

- Gamma-ray burst (GRB) [137]: $\mathcal{H}_{\rm GRB}(z) = (1+z)^{1.4} \mathcal{H}_{\rm SFR}(z)$;

- Active galactic nuclei (AGN) [138, 139]:

$$\mathcal{H}_{\rm AGN}(z) = \begin{cases} (1+z)^5 & z < 1.7, \\ 2.7^5 & 1.7 < z < 2.7, \\ 2.7^5 \times 10^{0.43(2.7-z)} & z > 2.7; \end{cases} \tag{3.15}$$

- Quasi-stellar object (QSO) [140]:

$$\mathcal{H}_{\rm QSO}(z) = \begin{cases} (1+z)^3 & z < 1.9, \\ (1+1.9)^3 & 1.9 < z < 2.7, \\ (1+1.9)^3 e^{1-z/2.7} & z > 2.7. \end{cases} \tag{3.16}$$

In the case of a uniform evolution $\mathcal{H}_{\rm unif}(z) = (1+z)^3$, whereas in the case of no evolution $\mathcal{H}(z) = 1$ can be assumed.

3.2 Modeling the extragalactic background radiation

The modeling of the EBR is rather difficult, if the well known cosmic microwave background is excluded. In this section we will briefly describe the backgrounds simulated in HERMES and we will only define the parameterizations that we have adopted for our simulations.

In the last decades many effort have been made to investigate the extragalactic background light (EBL). Such a radiation should be produced by the assembly of matter into stars and galaxies, as well as by the evolution of such systems which releases radiant energy powered by gravitational and nuclear processes. Absorption of large frequency radiation by dust and re-emission at small frequency considerably increase the infrared component of the background light, whose investigation should shed light on structure formations processes. In the top-left panel of Fig. 3.4 we show the spectral power per unit surface and solid angle νI_ν, for each radiation, in units of nW m^{-2} sr^{-1}, as a function of the wavelength (energy) of background photons. The conversion to the energy density, in units of eV cm^{-3}, is easily obtained by the relation

$$\epsilon^2 n_\epsilon = 2.62 \times 10^{-4}\, \nu I_\nu \tag{3.17}$$

where ϵ is the photon energy in eV and n_ϵ is the photon spectral number density in units of photons cm^{-3} eV^{-1}.

EBL spans over almost 20 decades, according to observations and models, from radio waves around 10^{-7} eV up to the high energy $\gamma-$ray photons of several GeV, with cosmic microwave background (CMB), the relic blackbody radiation from the Big Bang, being the dominant form of electromagnetic energy followed by ultraviolet/optical (CUVOB) and infrared backgrounds (CIB). In Fig. 3.1 we show the measurements of the multi-wavelength extragalactic background radiation (Kneiske, 2006).

3.2.1 Cosmic microwave background

CMB is an electromagnetic radiation filling the observable Universe and represents a landmark test of the Big Bang model of the Universe. In the early stages of the Universe, hot plasma and radiation dominated, permeating the cosmos with a uniform fog making it opaque at the highest redshifts. Successively, the Universe became cooler by expanding in space-time and when it cooled enough, stable atoms could form: such atoms were able no more to absorb the thermal photons and the Universe became transparent. Relic photons of CMB produced at that stage, the so-called surface of last scattering, have been propagating through the whole Universe, with an energy decreasing with the expansion of the Universe. The energy

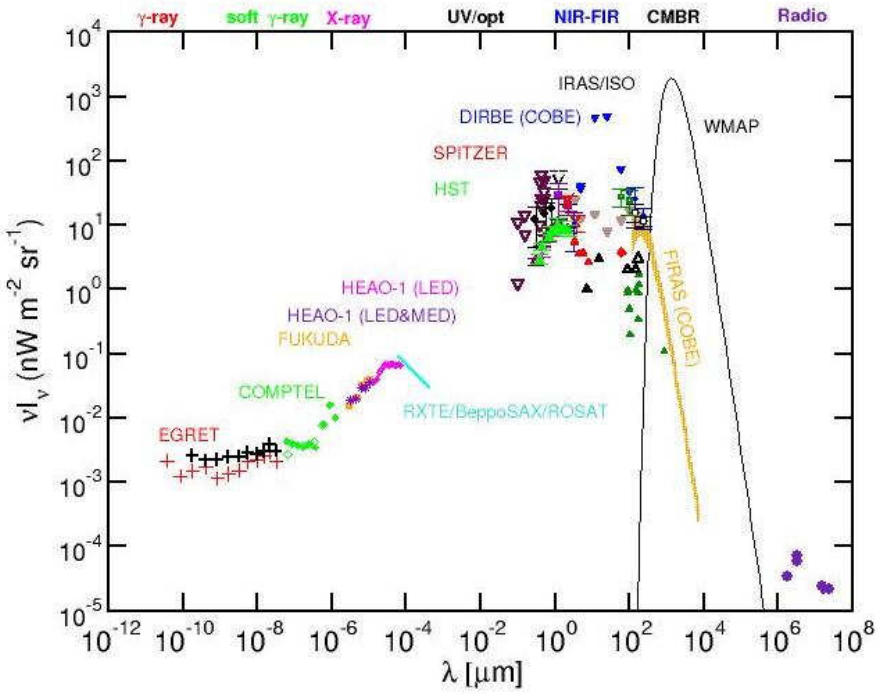


Figure 3.1: Measurements of the multi-wavelength extragalactic background radiation (T. Kneiske, 2006).

distribution of CMB photons can be nicely modeled by a blackbody radiation, described by the Planck's law

$$I_\nu(T) = \frac{2h\nu^3}{c^2}\left(e^{\frac{h\nu}{k_B T}} - 1\right)^{-1} \tag{3.18}$$

with temperature $T_0 \simeq 2.725$ K, being $k_B \simeq 8.6 \times 10^{-5}$ eV K^{-1} the Boltzmann constant, $h \simeq 4.1 \times 10^{-15}$ eV s the Planck constant and $c \simeq 3.0 \times 10^8$ m s^{-1} the speed of light in the vacuum. By assuming that the cosmological model of gravitation is described by general relativity and electromagnetism by Maxwell theory, a theoretical consequence of the adiabatic expansion of the Universe is that photons should propagate along null geodesics and that the CMB temperature should evolve with redshift as $T(z) = T_0(1+z)^{1-\beta}$, with $\beta = 0$. From the same arguments, it can be shown that the energy of CMB photons evolve as $E(z) = E_0(1+z)$, whereas their number density evolve as $n_\epsilon(z) = n_\epsilon(z=0)(1+z)^3$.

In the left panel of Fig. 3.2 we show the skymap of CMB obtained from seven years observations of WMAP (Wilkinson Microwave Anisotropy Probe) satellite. The anisotropic distribution of the radiation is even more evident in the right panel of the same figure, where the corresponding power spectrum is shown as a function of multipole moment l. Such observations allow the estimation of cosmological parameters with unprecedented accuracy. In Tab. 3.1 we show the obtained values for cosmological parameters which play a significant role in the propagation of UHECR, including protons, heavier nuclei and photons.

3.2.2 Cosmic infrared/optical background

The infrared background is a radiation of extragalactic origin, and it is therefore expected to be isotropic on large scales. The spectrum will depend in a complex way on the characteristics of the luminosity sources, on their cosmic history, and on the history of dust formation and the distribution of dust

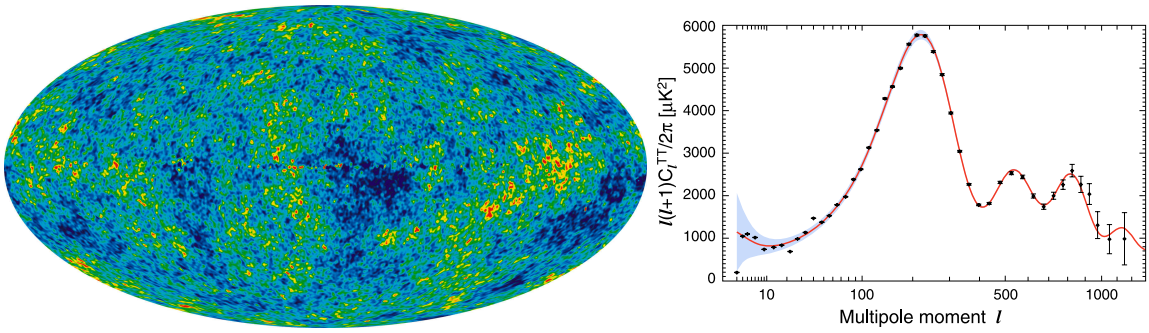


Figure 3.2: **Left:** Skymap of CMB radiation from 7-years WMAP measurements obtained from a weighted linear combination of the five WMAP frequency maps (23, 33, 41, 61 and 94 GHz), minimizing the galactic foreground contribution to the sky signal (credits: NASA Goddard Space Flight Center, WMAP Science Team). **Right:** Corresponding power spectrum as a function of multipole moment l from observations of WMAP and other experiments (dots) and from predictions of ΛCDM cosmological model [141].

Parameter	**7-year Fit**	**5-year Fit**
$10^2\Omega_b h^2$	$2.258^{+0.057}_{-0.056}$	2.273 ± 0.062
$\Omega_c h^2$	0.1109 ± 0.0056	0.1099 ± 0.0062
Ω_Λ	0.734 ± 0.029	0.742 ± 0.030
Derived parameters		
t_0	13.75 ± 0.13 Gyr	13.69 ± 0.13 Gyr
H_0	71.0 ± 2.5 km/s/Mpc	$71.9^{+2.6}_{-2.7}$ km/s/Mpc
Ω_b	0.0449 ± 0.0028	0.0441 ± 0.0030
Ω_c	0.222 ± 0.026	0.214 ± 0.027

Table 3.1: Cosmological parameters directly involved in the propagation of UHECR: six-Parameter ΛCDM fit to *WMAP* data only [141].

relative to the luminosity sources. Experimentally, direct measurements of the CIB are very difficult. From a technical point of view the major challenge is to make absolute sky brightness measurements relative to a well-established zero flux level, where emission from telescope and instrument components and the Earth's atmosphere must be eliminated. Scattered and diffracted light from the very bright local sources (Sun, Earth, and Moon) must also be strongly rejected. In practice, this requires that observations be conducted with carefully designed, cryogenically cooled instruments located above the Earth's atmosphere. Confident measurement of the CIB requires sufficient observation time to identify and eliminate potential sources of systematic measurement errors [142]. In Fig. 3.3 we show three skymaps of CIB radiation, measured from DIRBE satellite, for three different values of the wavelength, namely 100, 140 and 240 μm.

In the bottom-left panel of Fig. 3.4 we show the predictions of EBL from backward evolution models, which extrapolate the spectral properties of local galaxies to higher redshifts using some parametric form for their evolution. In their simplest form, commonly referred to as no evolution models, these models assume that neither the spectral energy distribution nor the comoving number density of galaxies evolve with time. The spectral luminosity density is therefore explicitly independent of redshift. In forward evolution models (bottom-right panel of Fig. 3.4) the spectral evolution of stellar populations is taken into account, as well as the stellar, gas, and metallicity content and spectral energy density of a galaxy as a function of time starting at the onset of star formation. These models rely on a wide range of computational and observational data sets, such as stellar evolutionary tracks, libraries of

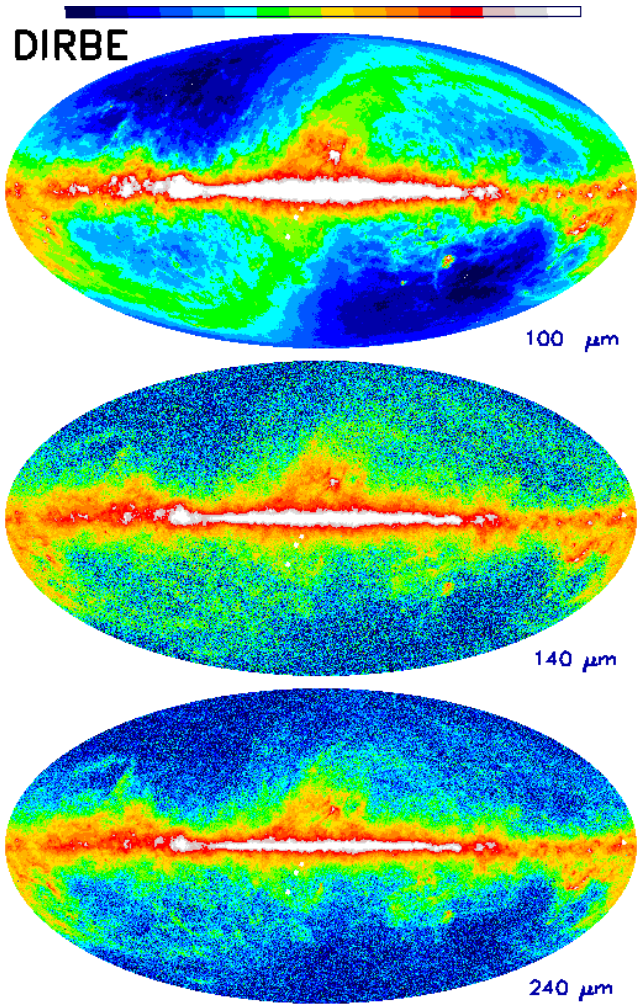


Figure 3.3: Skymaps of CIB radiation from DIRBE measurements in the wavelength range 100-240 μm (credits: NASA Goddard Space Flight Center, Diffuse Infrared Background Experiment (DIRBE) Project).

observed and calculated stellar atmospheres, stellar nucleosynthesis yields, and the observed luminosity functions of galaxies. Assuming an initial formation epoch, cosmological parameters are used to map the temporal evolution of galaxies into redshift space, allowing direct comparison of the model predictions with observations. Model parameters are adjusted to match the galaxy number counts, spectral energy distribution, colors, and metallicity as a function of redshift [142].

Because the extragalactic background light is an integrated measure of cosmic activity, summed over time and over the wide variety of processes and systems that have populated the Universe, it can only inform us about global characteristics of cosmic history. A modeling approach which deals with average properties of the Universe rather than the many complex details involved would most naturally, and perhaps most informatively, relate to the background radiation. Cosmic chemical evolution models, here in the top-right panel of Fig. 3.4, use just such an approach, relating in a self-consistent way the time history of a few globally averaged properties of the Universe. Such models provide a picture of the evolution of the mean density of stars, interstellar gas, metals and radiation averaged over the entire population of galaxies in a large comoving volume element. Inputs to the models are tracers of stellar activity (emitted light) and tracers of the (interstellar medium) ISM in galaxies (absorbed

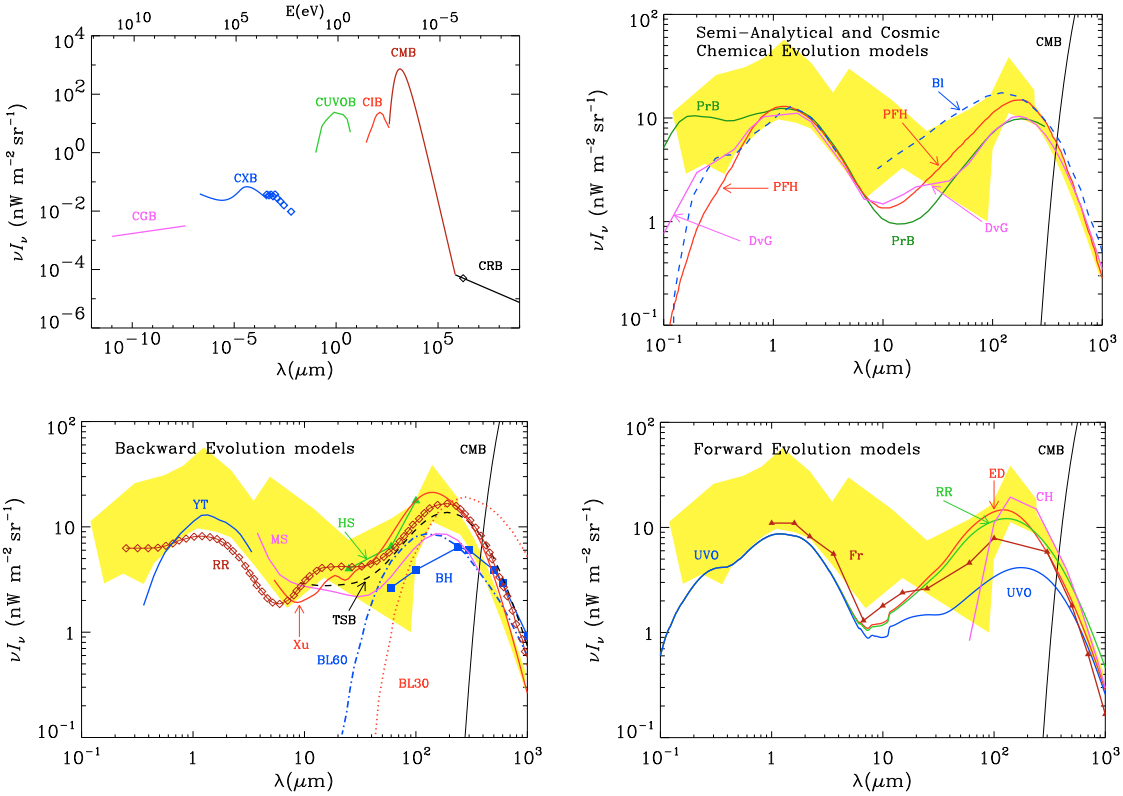


Figure 3.4: **Top-left:** Spectrum of the cosmic background radiations, including Radio (CRB, $\nu I_\nu \propto \nu^{0.3}$, Microwave (CMB, blackbody spectrum at $T_0 = 2.725$ K, UV/Optical (CUVOB) and Infrared (CIB) [142], X-ray (CXB) [143, 144] and $\gamma-$ray (CGB) [145]. **Top-right:** EBL predicted by semi-analytic and cosmic chemical evolution models. **Bottom-left:** EBL predicted by backward evolution models. **Bottom-right:** EBL predicted by forward evolution models. For detailed information about each figure, see Ref. [142] and refs. therein.

light). In the same panel we show some semi-analytical models, considering a large number of physical processes as the cooling of the gas that falls into the halos, a prescription for the formation of stars, a feedback mechanism that modulates the star formation efficiency, a stellar IMF, and a star formation efficiency during merger events, in order to reproduce observable galaxy properties. The main advantage of these models is that, in spite of the many adjustable parameters, they provide a physical approach to the formation and evolution of galaxies [142]. For a more detailed treatment of infrared and optical background radiations we refer to [142, 146–151] and Refs. therein.

The evolution of the density of CMB photons is simply given by $\mathcal{E}(z) = (1+z)^3$, whereas the evolution of the density of CIB photons is still debated and depends on the adopted scenario for the luminosity evolution. Two models, included in HERMES, have been recently suggested by Stecker et al [148]:

1. **Base-line model**:

$$\mathcal{E}(z) = \begin{cases} (1+z)^{3.1} & z \leq 1.3 \\ (1+1.3)^{3.1} & 1.3 < z \leq 6 \\ 0 & z > 6 \end{cases} \tag{3.19}$$

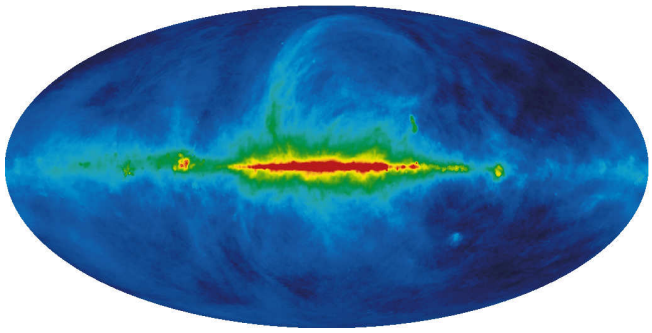


Figure 3.5: Haslam 408 MHz skymap of CRB radiation, derived from 4 separate surveys (credits: NASA Goddard Space Flight Center, LAMBDA).

2. **Fast model**:

$$\mathcal{E}(z) = \begin{cases} (1+z)^4 & z \leq 1 \\ (1+1)^4 & 1.3 < z \leq 6 \\ 0 & z > 6 \end{cases} \tag{3.20}$$

In the current cosmological epoch and at the IRB maximum epoch, which is around $z = 2$, the fast evolution model provides an higher density than base-line model. In any case, it is worth remarking that the cosmological evolution of the infrared background density is much slower than that of CMB.

We will see further in this chapter that the infrared background is negligible when the propagation of high energy protons is considered: conversely, it plays an important role during the propagation of high energy nuclei.

3.2.3 Cosmic radio background

The real intensity of the CRB radiation is poorly known, as in the case of the CIB, because of the lack of experimental observations due to our galaxy, being responsible of emitting and absorbing in the corresponding frequency range.

The main contributions to the radio background are provided by normal galaxies and radio galaxies, with a significant correlation with the evolution of the infrared background [152]. It has been speculated that the CRB signal could be dominated by the contribution of star-forming galaxies at high redshift if the far infrared-radio correlation evolved in time, even if it has been recently shown that star-forming galaxies can only account for $\approx 13\%$ of the observed intensity of the CRB [152].

One of the first models of the CRB, in agreement with current observations at that time, is due to Protheroe and Biermann, who have provided a new estimation of extragalactic radio background radiation down to kHz frequencies [153]. The intensity of the CRB at any given frequency is given by

$$I_\nu = \frac{c}{4\pi H_0} \int_0^\infty \frac{\varepsilon_{\nu'}(z)}{(1+z)H(z)} dz, \tag{3.21}$$

where $H(z)$ reflects the cosmological expansion, $\nu' = \nu(1+z)$ is the initial frequency of photons at which the photons observed today with a frequency ν were emitted and

$$\varepsilon_{\nu'}(z) = \int_0^\infty n(\omega, z) L_{\nu'}(\omega) d\omega \tag{3.22}$$

is the average emissivity per unit volume. Here, ω denotes the star formation rate (SFR) of the galaxy, $L_{\nu'}(\omega)$ is the luminosity function, $n(\omega, z)$ indicates the number density of galaxies with SFR between ω

and $\omega + d\omega$, and the contributions of all the galaxy at a redshift z are taken into account through the sum over any z.

The main contribution to the CRB is due to normal galaxies and is uncertain because of uncertainties in their evolution. As long as luminosity linearly depends on the instantaneous star formation rate, the average emissivity can be expressed in terms of the variation rate of the cosmic SFR density [154]. The brightness temperature of the CRB has been recently estimated to be

$$\mathcal{T} = 0.0137 \text{ K } \left(\frac{\nu}{\text{GHz}}\right)^{-2.1} \tag{3.23}$$

for young, massive stars producing intense UV radiation that ionizes the surrounding medium, and thermal bremsstrahlung from these free electrons, and

$$\mathcal{T} = 0.1402 \text{ K } \left(\frac{\nu}{\text{GHz}}\right)^{-2.7} \tag{3.24}$$

for synchrotron emission, assuming the evolution of the far infrared-radio correlation [152].

The evolution of the density of CRB photons included in HERMES is the one proposed by Protheroe and Biermann, who modified the luminosity evolution to fit the source counts [153]:

$$\mathcal{E}(z) = \begin{cases} (1+z)^4 & z < 0.8 \\ (1+0.8)^4 & z \geq 0.8 \end{cases} \tag{3.25}$$

where the value $z_0 = 0.8$ has been obtained from the best fit for both normal galaxies and radio galaxies.

In Fig. 3.5 we show the skymap for the CRB radiation obtained by combining 4 separate surveys. We will see further in this chapter that the radio background is negligible when the propagation of high energy nuclei is considered: conversely, it plays an important role during the propagation of high energy photons.

Note. For a blackbody, the brightness $I_\nu d\nu$ is defined as the amount of energy per unit surface, time and solid angle, emitted between ν and $\nu + d\nu$. The fraction of the blackbody radiance determined by the emissivity ε_ν defines the inverse of the *brightness temperature* by

$$\mathcal{T}_\nu^{-1} = \frac{k_B}{h\nu} \ln\left(1 + \frac{e^{\frac{h\nu}{k_B T}} - 1}{\varepsilon_\nu}\right)$$

where T is the temperature of the blackbody. In the case of low frequency, or, equivalently, at very high temperatures, we can use the Rayleigh-Jean approximation of the Planck's law to estimate the temperature

$$T_\nu = \frac{c^2 I_\nu}{2 k_B \nu^2}$$

as a function of the frequency ν. Within such approximation, the brightness temperature is defined by the simple relation $\mathcal{T}_\nu = \varepsilon_\nu T_\nu$.

3.2.4 Summary of the parameterizations included in HERMES

For the propagation of UHECR nuclei, in HERMES we adopt the blackbody model for CMB and the semi-analytical "model D" proposed by Finke et al [151], modeling the star formation rate recently

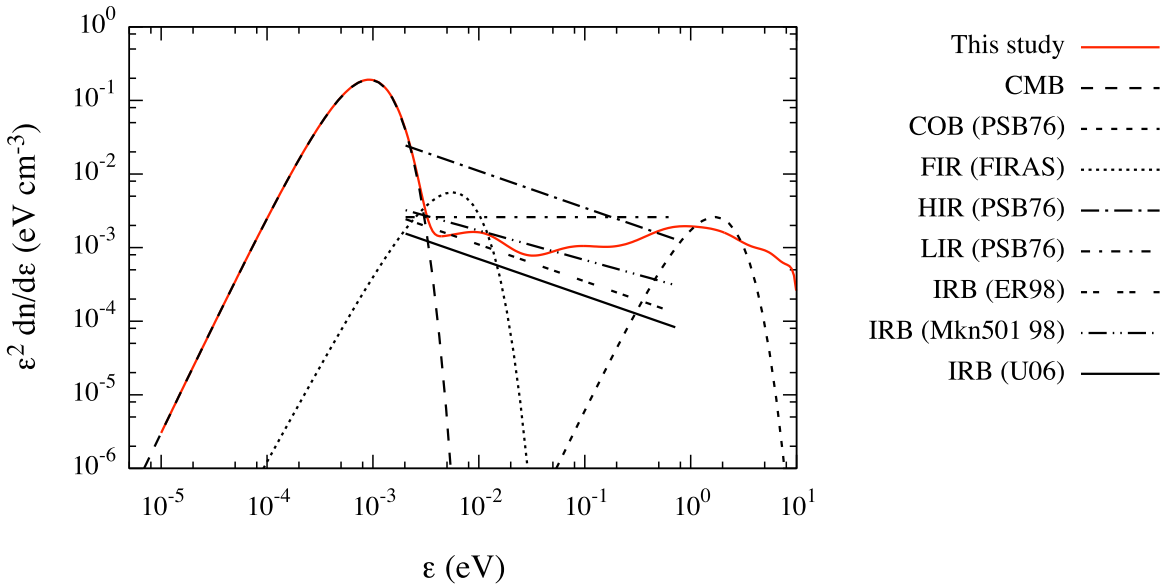


Figure 3.6: Different parameterizations of extragalactic background radiation as a function of relic photon energy: CMB, (Far, Low and High) IRB and COB. The red line indicates the EBL parameterization included in HERMES. The other parameterizations, shown for reference, are taken from PSB76 [155], FIRAS [156], ER98 [157], Mkn501-98 [158], U06 [159]. Photon energy is considered in the laboratory frame.

introduced by Hopkins and Beacom [136] for CIOB, with redshift evolution specified in the corresponding paragraphs of this section.

Some models of extragalactic background radiations are shown in Fig. 3.6, as a function of the photon energy ϵ in the laboratory frame. The red solid line indicates the EBL parameterization included in HERMES, and it should be considered the default, where not specified otherwise. For sake of completeness, we also show the common parameterizations by Puget, Stecker and Bredekamp (PSB76) for COB, lower and higher IRB (LIR and HIR, respectively) [155], and other IRB models, derived from theoretical arguments or experimental observations [156–159].

3.3 Modeling and simulation of energy-loss processes

During their propagation, photons, neutrinos and nuclei (A, Z) (electric charge, mass) with injection energy E_i, generally undergo interactions with background photons. UHECR that reach the Earth are therefore detected with a degraded energy $E_f < E_i$, depending on the type of interactions they were subjected to and on the distance between the source and the Earth. In HERMES, we describe the energy loss in a unit interval of z in terms of equations like

$$\frac{1}{E}\frac{dE}{dz} = -\beta(z, E)\frac{dt}{dz}, \tag{3.26}$$

where

$$-\frac{dt}{dz} = \frac{1}{H_0(1+z)}\left[\Omega_M(1+z)^3 + \Omega_\Lambda + (1-\Omega_M-\Omega_\Lambda)(1+z)^2\right]^{-\frac{1}{2}} \tag{3.27}$$

is the general metric element accounting for the cosmological expansion [140, 160, 161], and the involved cosmological parameters have been introduced in Sec. 3.1. The function $\beta(z, E)$ is related to the cooling

rate of the UHE particle and it depends on the particular energy loss process considered. As we will see further in this section, $\beta(z, E)$ is proportional to the inverse of the mean free path and depends on the density of background photons and their energy, on the energy of the UHECR and on the cross section of the interaction under investigation. In the case of nuclei, it also depends on the nuclear mass and charge. Thus, the total energy loss rate is obtained by

$$\frac{1}{E}\frac{dE}{dz} = -\frac{dt}{dz}\sum_{\text{process}} \beta_{\text{proc}}(z, E), \tag{3.28}$$

where the sum is extended to all interactions acting during the propagation. In HERMES, we include only those interactions which have a significant impact on the propagation of UHECR:

- **Adiabatic loss:** it is due to the expansion of the universe; it is considered for all nuclei with $A \geq 1$, photons and neutrinos;
- **Pair production:** it involves the creation of a positron/electron pair; it is considered for all nuclei with $A \geq 1$ and photons;
- **Photo-pion production:** it involves the creation of one or multiple pions; it is considered for all nuclei with $A \geq 1$;
- **Photodisintegration:** it involves the fragmentation of the original nucleus, with the creation of lighter nuclides (generally referred to as *fragments*); it is considered for all nuclei with $A \geq 2$;
- **Inverse Compton and synchrotron emission:** it is considered for photons and pairs which are part of the electromagnetic cascade generated by nuclei.

In the following we will take into account the interactions of nuclei with cosmic microwave background (CMB) and cosmic infrared/optical background (CIOB) radiations, by adopting the parameterization described in Sec. 3.2.4 (see Fig. 3.6) for the extragalactic background radiation.

Moreover, for the sake of simplicity, we will omit to specify that results shown in the following plots have been obtained from HERMES. Where it will be not the case, we will specify the corresponding source. Finally, in this section we will describe the propagation of UHECR with no regards of magnetic fields: such an approach is generally known as "1D propagation".

3.3.1 Cross section of $p\gamma$ interactions

The probability of UHE protons to interact with background photons rapidly increases with proton energy. If E and ϵ are the energies of the UHE proton and the photon in the observer rest frame, respectively, the interaction is equivalent to a collision with a high energy photon with energy $\epsilon' = \Gamma\epsilon(1 - \cos\theta)$, being θ the collision angle. For instance, when the energy ϵ' equals at least the pion mass $m_\pi c^2 \approx 140$ MeV, the proton undergoes photo-meson production and loses energy. Such a process is known as Greisen-Zatsepin-Kuzmin effect and dominates above $50 - 60$ EeV [29, 30]. The two main channels for the interaction, close to the threshold energy, are

$$p + \gamma_{\text{EBR}} \longrightarrow \Delta(1232 \text{ MeV}) \longrightarrow \begin{array}{l} p + \pi^0 \\ n + \pi^+, \\ \quad n \longrightarrow p + e^- + \bar{\nu}_e. \end{array}$$

involving the resonance $\Delta(1232$ MeV). At higher energies, heavier resonances and multi-pion production channels are likely. Just above the threshold, baryonic resonances dominate and protons are subjected to photo-meson production, mainly through the $\Delta(1232)$-baryon resonance, whereas heavier resonances

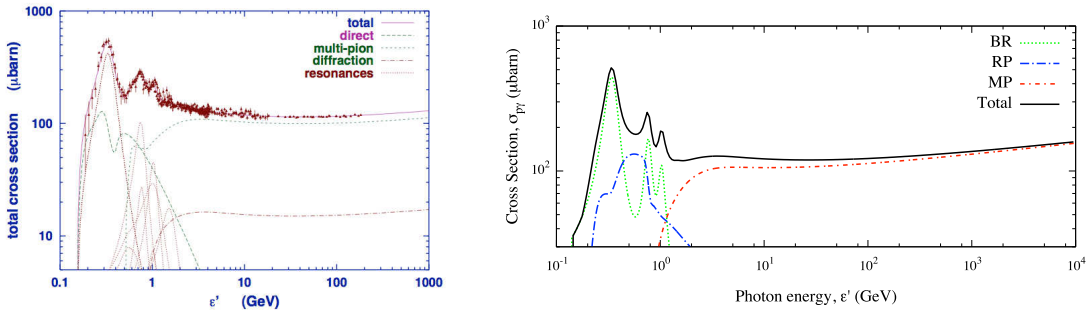


Figure 3.7: Total cross section for $p\gamma_{\rm EBR}$ interactions as a function of the background photon energy ϵ' in the proton rest frame: experimental data (dots) and the different contributions due to baryonic resonances (BR), direct particle (RP) and multi-pion (MP) production are shown as reported in Refs. [162, 163] (left panel) and as obtained from HERMES, following Rachen's parameterizations (right panel).

(up to $\Delta(1950)$-baryon) play a more marginal role. We parameterize the cross-section for baryonic resonances by

$$\sigma_{\rm BR}(\epsilon) \quad = \quad \sum_{i=1}^{4} \sigma_i \sigma_L(\epsilon; \epsilon_i, \Gamma_i)$$

where σ_L is the Lorentzian function, (ϵ_i (GeV), Γ_i (GeV), σ_i (μb)) $= (0.34, 0.17, 351)$, $(0.75, 0.50, 159)$, $(1.00, 0.60, 21)$ and $(1.50, 0.80, 26)$ for $i = 1, 2, 3$ and 4, respectively. For all other processes participating in photo-meson production, including multipions (MP) or direct particle production involving π, η, Δ, ρ, ω and strange-particle channels (RP), we use Rachen's parameterizations [162]. In Fig. 3.7 we show the contributions of each channel to the total cross sections. In the left panel the comparison between the cross section obtained by Rachen and experimental data is shown [162, 163], whereas in the right panel of the same figure we show the cross section obtained from our simulator, putting in evidence the good agreement.

In the following we will use the abbreviation "BR" to refer to baryonic resonances, direct particle and multi-pion production, where not specified otherwise.

3.3.2 Interaction lengths for $p\gamma$ interactions

The general formula for the estimation of the interaction length (or, equivalently, mean free path) is given by

$$\lambda_A^{-1}(z, E) = \mathcal{E}(z) \frac{c}{2\Gamma_A^2} \int_{\epsilon_{\rm thr}/2\Gamma_A}^{\epsilon_{\rm max}} d\epsilon \frac{n(\epsilon)}{\epsilon^2} \int_{\epsilon_{\rm thr}}^{2\Gamma_A \epsilon} d\epsilon' \epsilon' \sigma(\epsilon') \tag{3.29}$$

where now $\Gamma_A = (1+z)\frac{E}{Am_pc^2}$ is the Lorentz factor of the nucleus at redshift z, $\epsilon_{\rm thr}$ is the energy threshold of the considered process in the nucleus rest frame, $n(\epsilon)$ is the density of background photons with energy ϵ in the observer's rest frame, ϵ' is the energy of the photon in the nucleus rest frame and $\mathcal{E}(z)$ is the evolution function of the ambient photon field. It is straightforward to show that $\lambda_A(z, E) = (1+z)^{-3}\lambda_A[z = 0, (1+z)E]$ when the CMB is considered [164, 165], whereas for other background radiations a more complicated evolution should be used. By following Stanev et al [164], we define the average energy loss length by

$$\chi_{\rm loss}(z, E) = \frac{E}{dE/dz} = \frac{\lambda(z, E)}{\kappa(E)}, \tag{3.30}$$

where $\kappa(E) = \langle \Delta E \rangle / E$ is the mean inelasticity, i.e. the average fraction of energy lost by the nucleus because of the interaction. The inelasticity for pair production is $\kappa \approx 2m_e/(Am_p)$ (being m_e and m_p the masses of electron and proton, respectively), i.e. around 10^{-3} in the case of protons, and even smaller for heavier nuclei. Conversely, for photo-pion production by protons, the inelasticity ranges from 0.2 to 0.5, depending on the energy. For the analytical derivation of the inelasticity factor, see Appendix A.

Adiabatic loss

In order to take into account the energy loss due to the expansion of the universe, we use

$$\beta_{\rm rsh}(z) \quad = \quad H_0 \left[\Omega_M (1+z)^3 + \Omega_\Lambda + (1 - \Omega_M - \Omega_\Lambda)(1+z)^2 \right]^{\frac{1}{2}} \tag{3.31}$$

for the adiabatic term, as previously explained in Sec. 3.1.

Pair production

In the rest frame of the nucleus, the pair production process $A + \gamma_{\rm EBR} \longrightarrow A + e^+ + e^-$ occurs at the threshold energy $2m_e c^2 \approx 1$ MeV and it plays an important role only when CMB is considered, the CIOB partecipating marginally [155]. We can treat the process as a continuous energy loss (CEL), because the loss per interaction is very small. The energy loss accounting for the pair production, due to the Bethe-Heitler interaction with ambient photons with density $n(\epsilon)$, is given [166] by

$$\beta_{e^\pm}(E; Z, A) \propto \alpha r_e^2 Z^2 (m_e c^2)^2 \int_2^\infty d\xi n \left(\frac{m_e c^2}{2\gamma} \xi \right) \frac{\varphi(\xi)}{\xi^2}$$

where $\gamma \approx E/(Am_p c^2)$ is the Lorentz factor of the nucleus, m_e is the electron mass, $\alpha = e^2/\hbar c$ is the fine-structure constant and $r_e = e^2/m_e c^2$ is the classical electron radius. In the case of a blackbody at temperature T_0, as for CMB, the energy loss obtained from the Born approximation becomes [166]

$$\beta_{e^\pm}(E; Z, A) = \frac{\alpha r_e^2 Z^2 (m_e c^2 k_B T_0)^2 c}{\pi^2 \hbar^3 c^3 E} f(\nu),$$

with

$$f(\nu) = \nu^2 \int_2^\infty d\xi \varphi(\xi) \left(e^{\nu\xi} - 1 \right)^{-1}, \quad \nu = \frac{m_e c^2}{2\gamma k_B T_0}.$$

Here, k_B is the Boltzmann constant and we multiplied by c in order to obtain a rate. Higher order terms of the Born approximation, proportional to $(Z\alpha v_\pm/c)^m$, where m is the number of interaction with the Coulomb field, should be taken into account in the case of nuclei heavier than protons. In fact, for $Z > 1$ the symmetry between produced electron and positron breaks down. Blumenthal suggested to correct the rate through the Sommerfeld factor [166], although it is only valid in the non-relavistic limit. For nuclei with $Z > 1$, this correction does not agree with experimental data and a better correction, namely

$$S(Z) \simeq \left[1 - 0.29 \times (Z\alpha)^2 + 0.25 \times (Z\alpha)^4 - 0.25 \times (Z\alpha)^6 \right],$$

valid in the case of ultra-relativistic particles, has been proposed by Rachen [162], even if it has been pointed out that Coulomb corrections to the Born approximation have a negligible effect on the pair production loss rate of ultra-relativistic heavy nuclei as ^{56}Fe [167]. By taking into account the evolution with redshift in CMB, after some algebra, in HERMES we use

$$\beta_{e^\pm}(z, E; Z, A) \simeq S(Z) \frac{\alpha^3 Z^2 A^2}{4\pi^2 \hbar} \frac{m_e^2 m_p^2}{E^3} \int_2^\infty d\xi \frac{\varphi(\xi)}{\exp\left[\frac{m_e A m_p}{2E(1+z) k_B T_0} \xi \right] - 1}, \tag{3.32}$$

that is similar to the parameterization adopted in Ref. [168], where the auxiliary function $\varphi(\xi)$ is obtained from Ref. [166, 169] and masses are in units of eV/c^2. However, there is no parameterization in the case of CIOB and, in our code, we estimate the corresponding energy loss rate by using Eq. (3.29).

Photo-pion production

The process has been described in detail in Sec. 3.3.1. In the case of protons propagating in the CMB, the dominant interaction above 50 EeV, we have

$$\begin{aligned}\beta_\pi(E) &= \frac{m_p^2}{2E^2}\int_0^\infty d\epsilon \frac{n(\epsilon)}{\epsilon^2}\int_0^{2\epsilon\frac{E}{m_p}} d\epsilon'\epsilon'\kappa(\epsilon')\sigma(\epsilon') \\ &= -\frac{k_B T_0}{2\pi^2\hbar}\frac{m_p^2}{E^2}\int_0^\infty d\epsilon\kappa(\epsilon)\sigma(\epsilon)\epsilon \times \ln\left[1-\exp\left(-\frac{m_p}{2Ek_BT_0}\epsilon\right)\right], \end{aligned} \tag{3.33}$$

where m_p is the proton mass in units of eV/c^2, $\sigma(\epsilon)$ is the cross-section for pion production in terms of the photon energy ϵ and $\kappa(\epsilon)$ is the inelasticity factor. We parameterize the contribution of this term as in Ref. [168] by

$$\beta_\pi(z,E;1,1) \simeq \begin{cases} A_\pi(1+z)^3\exp\left[\frac{B_\pi}{E(1+z)}\right] & E \le E_{\rm match}(z) \\ C_\pi(1+z)^3 & E > E_{\rm match}(z) \end{cases}, \tag{3.34}$$

in order to avoid further numerical integrations. The function $E_{\rm match}(z) = 6.86e^{-0.807z}\times 10^{20}$ eV ensures the continuity of the function $\beta_\pi(z,E;1,1)$ and $\{A_\pi, B_\pi, C_\pi\} = \{3.66\times 10^{-8}{\rm yr}^{-1}, 2.87\times 10^{20}{\rm eV}, 2.42\times 10^{-8}{\rm yr}^{-1}\}$ are taken from Ref. [170].

In Fig. 3.8 we show the interaction length λ, in the CIOB and the CMB, for each process separately and for all processes together, as well as the energy loss length $\chi_{\rm loss}$, as a function of the energy E of the proton in the observer rest frame. In the CIOB, it is evident that the pair production is negligible with respect to other processes, because occurring on time scales larger than the adiabatic expansion, for all energies above 10^{18} eV. A similar argument applies for the photo-meson production in the CIOB, which, below 10^{20} eV, contributes less than pair production in the CMB, whereas above 10^{20} eV the production of pions in the CMB dominates up to the highest energy. In the energy interval between 2×10^{18} eV and $\sim 5\times 10^{19}$ eV, the main energy loss process is the pair production in the CMB. The obtained results are in perfect agreement with recent literature [77, 165, 171, 172], with small differences related to the different CIOB adopted.

We treat the case of neutron in a similar way, by considering the additional process of the $\beta-$decay. The neutron decay rate is given by $m_N/(\tau_n E)$, with $\tau \simeq 888.6$ s the laboratory lifetime, providing a range of propagation

$$\lambda_\beta = \tau_n\frac{E}{m_N} \simeq 0.9\left(\frac{E}{10^{20}\ {\rm eV}}\right)\ {\rm Mpc}, \tag{3.35}$$

which becomes competitive with photo-pion production only at the highest energy, above 10^{20} eV.

Cascades generated by UHE protons and neutrons

We have discussed the production of electron/positron pairs and of secondary pions. Additionally, electron and pairs interact with the extragalactic background photons, participating to the electromagnetic cascade generated by the primary proton. Conversely, in the case of photo-meson production, close to the threshold the main channels involve the creation of a single neutral pion π^0, with probability $\approx 2/3$, or a charged pion π^+, with probability $\approx 1/3$. We treat the case when the UHE primary is a neutron

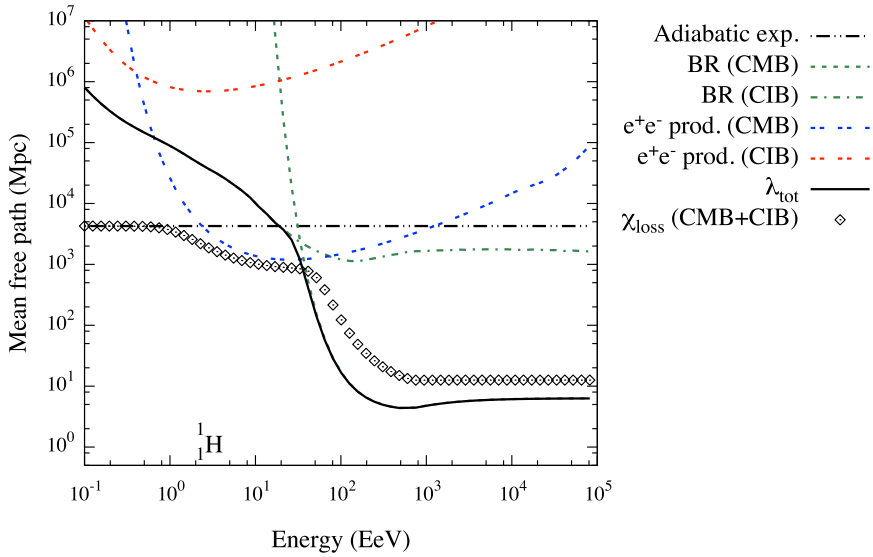


Figure 3.8: Our estimation of the mean free path λ at $z = 0$ as a function of the energy of the proton in the observer rest frame. The contributions due to different processes (adiabatic, pair and photo-pion production) in CMB and CIOB are shown separately, as well as the total interaction length $\lambda_{\rm tot}$ and the total energy loss length $\chi_{\rm loss}$.

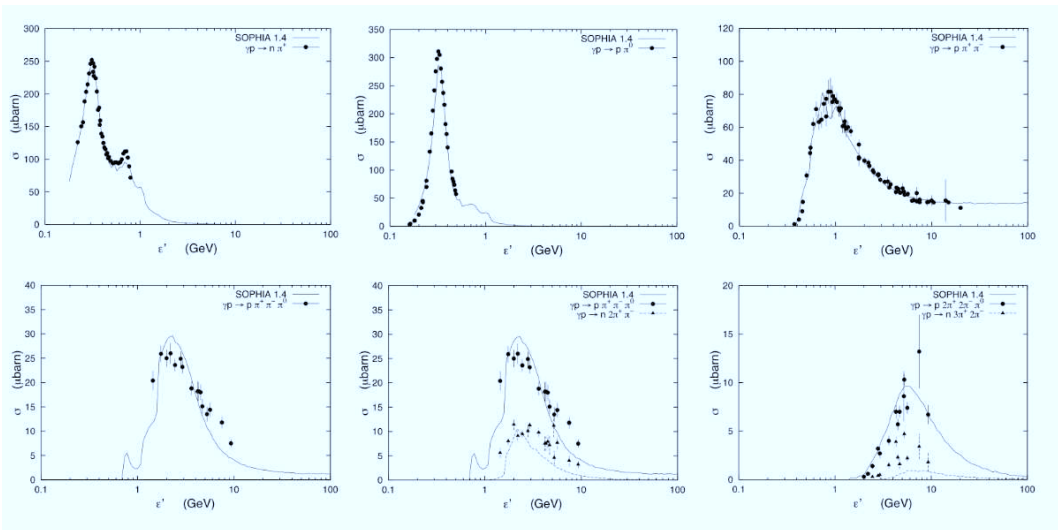


Figure 3.9: Photohadronic interactions of relativistic nucleons with an ambient photon radiation field, obtained from the simulator SOPHIA [173] (lines) and from experiments (dots).

in a similar way, with the production of a π^- instead of π^+. Pions have small lifetime, of the order of 10^{-16} s for π^0 and 10^{-8} s for $\pi^\pm$: thus, we neglect their propagation because they quickly decay to secondary particles, which can decay to other particles (as in the case of secondary muons) generating a cascade of electrons, positrons, photons and neutrinos. In HERMES, the main decay channels involving the production of a single pion and including the $\beta-$decay of neutrons, are reported below:

$$
\begin{aligned}
p+\gamma_{\rm EBR} \longrightarrow \Delta^+ \quad &\longrightarrow \quad \left\{ \begin{array}{l} p \\ \pi^0 \longrightarrow \gamma+\gamma \end{array} \right. \\
p+\gamma_{\rm EBR} \longrightarrow \Delta^+ \quad &\longrightarrow \quad \left\{ \begin{array}{l} n \longrightarrow p+e^-+\bar{\nu}_e \\ \pi^+ \longrightarrow \mu^+ + \nu_\mu \longrightarrow e^+ + \nu_e + \bar{\nu}_\mu + \nu_\mu \end{array} \right. \\
n+\gamma_{\rm EBR} \longrightarrow \Delta^0 \quad &\longrightarrow \quad \left\{ \begin{array}{l} n \longrightarrow p+e^-+\bar{\nu}_e \\ \pi^0 \longrightarrow \gamma+\gamma \end{array} \right. \\
n+\gamma_{\rm EBR} \longrightarrow \Delta^0 \quad &\longrightarrow \quad \left\{ \begin{array}{l} p \\ \pi^- \longrightarrow \mu^- + \bar{\nu}_\mu \longrightarrow e^- + \bar{\nu}_e + \nu_\mu + \bar{\nu}_\mu \end{array} \right. .
\end{aligned}
$$

In Fig. 3.9 the cross sections for nucleon and photon interactions are shown, as obtained from the data and as parameterized in SOPHIA, a Monte Carlo program for photohadronic interactions of relativistic nucleons with an ambient photon radiation field [173]. Channels with single and multi-pion production are present. It is evident that close to the threshold and for $\epsilon' < 1$ GeV, the dominating processes involve single pion production only, whereas at the highest energies channels with two or three pions are available. With good approximation, it can be assumed that the most relevant processes are:

$$
\begin{aligned}
p+\gamma_{\rm EBR} \longrightarrow \Delta^{++}+\pi^- \quad &\longrightarrow \quad p+\pi^++\pi^- \\
p+\gamma_{\rm EBR} \longrightarrow \Delta^{+}+\pi^0 \quad &\longrightarrow \quad \left\{ \begin{array}{l} p+\pi^0+\pi^0 \\ n+\pi^++\pi^0 \end{array} \right. \\
p+\gamma_{\rm EBR} \longrightarrow \Delta^{0}+\pi^+ \quad &\longrightarrow \quad \left\{ \begin{array}{l} n+\pi^0+\pi^+ \\ p+\pi^-+\pi^+ \end{array} \right. \\
n+\gamma_{\rm EBR} \longrightarrow \Delta^{+}+\pi^- \quad &\longrightarrow \quad \left\{ \begin{array}{l} n+\pi^++\pi^- \\ p+\pi^0+\pi^- \end{array} \right. \\
\\
n+\gamma_{\rm EBR} \longrightarrow \Delta^{0}+\pi^0 \quad &\longrightarrow \quad \left\{ \begin{array}{l} n+\pi^0+\pi^0 \\ p+\pi^-+\pi^0 \end{array} \right. \\
n+\gamma_{\rm EBR} \longrightarrow \Delta^{-}+\pi^+ \quad &\longrightarrow \quad n+\pi^-+\pi^+,
\end{aligned}
$$

where the secondary particles produced from pions and $\beta-$decay of neutrons have been omitted for simplicity. The inclusion in HERMES of such channels is currently under development.

3.3.3 Cross section of $A\gamma$ interactions, photo-disintegration and decay chain

As in the case of protons, the probability of UHE nuclei to interact with background photons rapidly increases with nucleus energy. The processes involved in such interactions are the same that we have previously described in the case of protons, namely pair and photo-pion production. However, in the case of heavy nuclei we have to take also into account the photo-disintegration (or photo-erosion) process

$$
{}^{A}_{Z}X + \gamma \longrightarrow {}^{A'}_{Z'}Y + m\alpha + [(Z-Z')-2m]p + [(A-A')-(Z-Z')-2m]n, \tag{3.36}
$$

resulting in the emission of subatomic particles, with the creation of lighter nuclides. Here, m is the multiplicity of α particles, p indicates the proton and n the neutron. In general, in order to describe the changes in abundance of the heavy nuclei as a result of the interaction of the UHECR with the background radiation, a nuclear reaction network including all interactions of interest should be used. Such a network is described by a system of coupled differential equations corresponding to all the

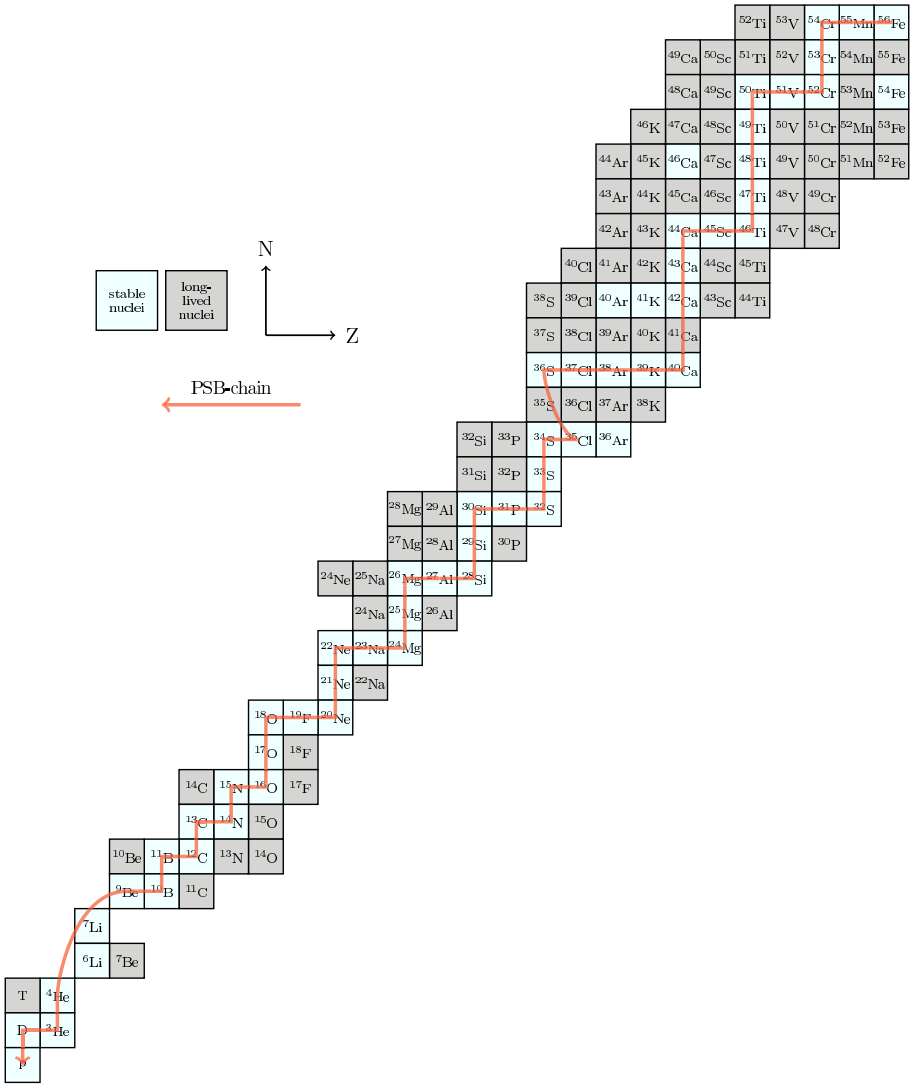


Figure 3.10: Puget-Stecker-Bredekamp (PSB) decay chain for the iron nuclei [155] adopted in HERMES. Picture taken from Ref. [174].

reactions affecting each nucleus, i.e. mainly photo-disintegrations and $\beta-$decays. Such an approach has been recently proposed, and adopted in many successive works, in Ref. [175] for the study of UHE nuclei propagation by using up to date measurements of cross sections [176]. Instead of direct measurements, other recent works related to this topic [174, 177, 178] make use of TALYS [179, 180], a software for the most likely simulation of nuclear reactions.

We adopt the simplest approach to the treatment of the photo-disintegration channels, by following the chain of stable nuclei (stability chain), as suggested for the first time by Puget, Stecker and Bredekamp (PSB) [155]. The PSB chain included in HERMES is shown in Figure 3.10, together with the other nuclei involved in the reaction network. The relative contribution of all decay channels corresponding to nuclei with $A \leq 56$ are taken from Ref. [155] and [167]. However, in order to produce more realistic simulations of the photo-disintegration process, we have obtained from TALYS reactions the branching ratios associated to the most relevant exclusive channels, including one nucleon, two nucleons and multi-nucleons emission on CMB and CIOB, similarly to recent studies [174, 177]. Hence,

in HERMES, we have included different models for the photo-disintegration of nuclei, with cross sections corresponding to: i) the PSB Gaussian approximation; ii) the Rachen's parameterizations; iii) the TALYS reactions. In the following, we will briefly describe the first two models, involving analytical approximations.

In the nucleus rest frame, nuclear photo-disintegration is dominated by the giant dipole resonance (GDR) for a photon energy smaller than 30 MeV. The emission of one or two nucleons represents the main process, although multi-nucleon emission is still possible. Above 30 MeV and below 150 MeV, nuclei are mainly subjected to quasi-deuteron (QD) effect, where the photon interacts with a nucleon pair while they are scattering within the nucleus, leading to the ejection of the pair and possibly additional nucleons [155]. Above 150 MeV, baryonic resonances (BR) generally dominate: nuclei undergo to photo-meson production, mainly through the $\Delta(1232)$-baryon resonance, whereas heavier resonances (up to $\Delta(1950)$-baryon) play a more marginal role. As in the case of protons, Rachen's parameterizations (RP) can be adopted for all processes participating to photo-meson production, including direct particle production involving π, η, Δ, ρ, ω and strange-particle channels [162]. At energies above 1 GeV the photo-fragmentation (PF) occurs and the nucleus breaks into fragments of lower mass and energy. The rate of photo-disintegration with the emission of j nucleons is obtained from Eq. (3.29), and it is given by

$$\mathcal{R}_{\mathrm{dis},j}(z,E;Z,A) = \mathcal{E}(z)\frac{1}{2\Gamma_A^2}\int_{\epsilon_{\mathrm{thr}}/2\Gamma_A}^{\epsilon_{\mathrm{max}}} d\epsilon \frac{n(\epsilon)}{\epsilon^2}\int_{\epsilon_{\mathrm{thr}}}^{2\Gamma_A\epsilon} d\epsilon' \epsilon' \sigma_{A,j}(\epsilon') \tag{3.37}$$

where $\Gamma_A = (1+z)\frac{E}{Am_p}$ is the Lorentz factor of the nucleus at redshift z, ϵ_{thr} is the energy threshold of the considered process characterized by cross section $\sigma_{A,j}$, $n(\epsilon)$ is the density of background photons with energy ϵ in the observer's rest frame and ϵ' is the energy of the photon in the nucleus rest frame. By considering the interaction rate of all reaction channels for a given nucleus, the quantity of interest should be the *effective* nucleon loss rate, defined by

$$\mathcal{R}_{\mathrm{dis,eff}}(z,E;Z,A) = \sum_j j\mathcal{R}_{\mathrm{dis},j}(z,E;Z,A), \tag{3.38}$$

with an energy loss rate is given by

$$\beta_{\mathrm{dis,eff}}(z,E;Z,A) = \mathcal{R}_{\mathrm{dis,eff}}(z,E;Z,A)\kappa(E;A), \tag{3.39}$$

where $\kappa(E;A)$ is the average inelasticity factor accounting for the average fraction of energy loss in one collision. For photo-disintegration processes we can simply consider $\kappa(A) \approx j/A$, whereas for pion production we adopt

$$\kappa(E;A) = \frac{1}{2}\left(1 + \frac{m_\pi^2 - A^2 m_p^2}{s}\right), \tag{3.40}$$

where $\sqrt{s}$ is the total 4-momentum of the nucleus-photon system [181].

A common parameterization adopted for the contribution of the GDR has been originally proposed by PSB [155], and later updated in Ref. [167]:

$$\sigma_{\mathrm{GDR}}(\epsilon;A,j) = \xi_j \Sigma_d W_j^{-1} e^{-2(\epsilon-\epsilon_{p,j})^2/\Delta_j^2} \quad \epsilon_{\mathrm{thr}} \le \epsilon \le \epsilon_1, \tag{3.41}$$

where $\xi_j, \epsilon_{p,j}$ and Δ_j are obtained by fitting nuclear data (see tables in Ref. [155]), $\epsilon_1 = 30$ MeV and updated values of the energy thresholds ϵ_{thr}, for both one and two nucleons disintegration channels, are

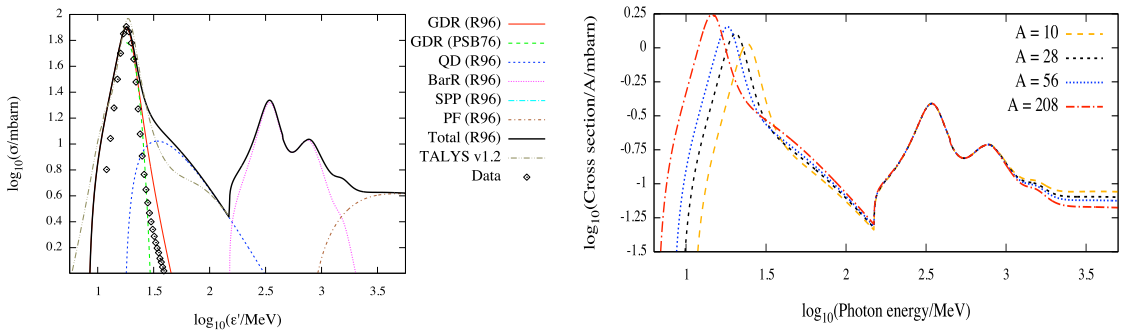


Figure 3.11: Total cross section for $A\gamma$ interactions as a function of the background photon energy ϵ' in the nucleus rest frame, obtained from HERMES, following Rachen's parameterizations. **Left:** Case of an iron nucleus ($A = 56$). The different contributions due to giant dipole resonance (GDR), quasi-deuteron (QD), baryonic resonances+direct particle production (BarR) and multi-pion production (PF) are shown. The parameterization of the GDR proposed by Puget et al [155] is reported, as well as the up to date measurements of cross sections [176] (diamonds). The cross section for GDR and QD obtained with TALYS [179, 180], a software for the simulation of nuclear reactions, is shown. **Right:** Scaling of the cross section with respect to the nuclear mass, as a consequence of Rachen's parameterizations.

given in Ref. [167]. The integrated cross section in this region can be approximately estimated through the Thomas-Reiche-Kuhn sum rule

$$\Sigma_d \equiv \int_0^\infty d\epsilon\sigma(\epsilon) = \frac{2\pi^2 e^2 \hbar}{m_p c}\frac{(A-Z)Z}{A} \approx 60\frac{(A-Z)Z}{A}\mu\text{b GeV},$$

while the function W_j is given by

$$W_j = \Delta_j\sqrt{\frac{\pi}{8}}\left[\text{erf}\left(\frac{\epsilon_{\max}-\epsilon_{p,j}}{\Delta_j/\sqrt{2}}\right) + \text{erf}\left(\frac{\epsilon_{p,j}-\epsilon_1}{\Delta_j/\sqrt{2}}\right)\right]. \tag{3.42}$$

The above parameterization makes the assumption that the total photo-disintegration cross-section can be approximated by a Gaussian function with a cut below the theoretical reaction threshold. An alternative approximation, based on Lorentzian functions, has been proposed and has been shown to better fit the data [175]. However, it has been also shown that, in practice, the energy loss rate exhibits small differences in the two cases for $11 \leq A \leq 56$ and virtually no differences for $A < 11$ [182]. Thus, following Rachen [162], we have also included the parameterization given by

$$\sigma_{\text{GDR}}(\epsilon; A) \quad = \quad 720\mu\text{b } A^{\frac{7}{6}} \frac{\Gamma_{\text{GDR}}^2}{\left(\epsilon A^{\frac{1}{6}} - E_{\text{GDR}}\right)^2 + \Gamma_{\text{GDR}}^2} \tag{3.43}$$

with $E_{\text{GDR}} = 35.3$ MeV $A^{-1/6}$ and $\Gamma_{\text{GDR}} = 15.1$ MeV $A^{-1/6}$. We follow Ref. [162, 175, 183–185] for the parameterization of the cross sections for the remaining process, namely QD and photo-meson production (including BR, RP and PF):

$$\sigma_{\rm QD}(\epsilon;A) = 550\mu{\rm b}\ A^{\frac{5}{4}}\left(\frac{\epsilon}{\epsilon_{\rm QD}}-1\right)^{\frac{3}{2}}\left(\frac{\epsilon}{\epsilon_{\rm QD}}\right)^{-3} \tag{3.44}$$

$$\sigma_{\rm BR}(\epsilon;A) = A\sum_{i=1}^{4}\sigma_i\sigma_L(\epsilon;\epsilon_i,\Gamma_i) \tag{3.45}$$

$$\begin{aligned}\sigma_{\rm RP}(\epsilon;A) &= A\left[53.3{\rm Ef}\left(\epsilon;0.15,0.15\right){\rm Ef}\left(-\epsilon;-0.65,-0.2\right)+{\rm Pl}\left(\epsilon;0.15,0.3,6\right)\right.\\ &+ 62.4{\rm Pl}\left(\epsilon;0.321,0.55,3.5\right)+1.5{\rm Pl}\left(\epsilon;0.7,1.2,4\right)+16.7{\rm Pl}\left(\epsilon;0.321,2.0,0.2\right)\\ &+ \left.7.2{\rm Pl}\left(\epsilon;0.512,1.3,1.1\right)+3.2{\rm Pl}\left(\epsilon;1.0,1.4,4.6\right)\right]\end{aligned} \tag{3.46}$$

$$\sigma_{\rm PF}(\epsilon;A) = 1\mu{\rm b}\ A^{0.91}\left(1-2.48e^{-\epsilon/0.8\ {\rm GeV}}\right)\left(69.8s^{0.081}+64.3s^{-0.453}\right) \tag{3.47}$$

where $\epsilon_{\rm QD}\approx 0.0333$ GeV $A^{-\frac{1}{6}}$, $s=0.88+\epsilon/0.532$ GeV, $(\epsilon_i$ (GeV), Γ_i (GeV), σ_i (μb)$)=(0.34,0.17,351)$, $(0.75,0.50,159)$, $(1.00,0.60,21)$ and $(1.50,0.80,26)$ for $i=1,2,3$ and 4, respectively. Rachen's parameterizations are defined by

$$\begin{aligned}{\rm Pl}\left(x;x_{\rm th},x_{\rm max},\alpha\right) &= \left(\frac{x-x_{\rm th}}{x_{\rm max}-x_{\rm th}}\right)^{A-\alpha}\left(\frac{x}{x_{\rm max}}\right)^{-A}, \quad A=\alpha\left[\frac{x_{\rm max}}{x_{\rm th}}\right] \qquad (3.48)\\ {\rm Ef}\left(x;x_0,w\right) &= \frac{1}{2}\left[1+{\rm erf}\left(\frac{3}{\sqrt{2}}\frac{2x-2x_0-w}{|w|}\right)\right] \qquad (3.49)\end{aligned}$$

with ${\rm Pl}\left(x;x_{\rm th},x_{\rm max},\alpha\right)=0$ for $x<x_{\rm th}$. Here, x is in GeV and the resulting cross-section is given in μb units.

In Fig. 3.11 we show the obtained total cross section for $A\gamma$ interactions as a function of the background photon energy ϵ' in the nucleus rest frame. In the left panel we consider the case of an iron nucleus ($A=56$), where the parameterization of the GDR proposed by PSB is reported, as well as the up to date measurements of in this energy range[1] [176] (diamonds), the cross sections for GDR and QD obtained with TALYS are also shown. In the right panel of the same figure, the scaling of the cross section with respect to the nuclear mass, as a consequence of Rachen's parameterizations, is shown. In Fig. 3.12 we show the comparison between the total cross sections estimated for iron (left panel) and proton (right panel) nuclei, together with the contribution of each single process separately.

For the sake of completeness, it is worth remarking that in HERMES the inclusion of some additional processes, not depending on the background radiation, are currently under development:

$$\begin{array}{ll} {}^A_Z X \longrightarrow {}^A_{Z+1} Y + e^- + \bar{\nu}_e & \beta^- - \text{decay}, \\ {}^A_Z X \longrightarrow {}^A_{Z-1} Y + e^+ + \nu_e & \beta^+ - \text{decay}, \\ {}^A_Z X + e^- \longrightarrow {}^A_{Z-1} Y + \nu_e & \text{electron capture.} \end{array}$$

3.3.4 Interaction lengths for $A\gamma$ interactions

For $A\gamma$ interactions, the differences in the cross section (with respect to the case of protons) are reflected in the interaction length. In both cases shown in Fig. 3.12, the available channels above the threshold for single pion production ($\epsilon'\approx 145$ MeV) involve baryonic resonances and direct particle production, with multi-pion production playing a significant role at the highest energies ($\epsilon'>700$ MeV). In the case of iron, the additional channels due to photo-disintegration process are evident at lower energies ($1<\epsilon'<150$ MeV). The energy loss due to the pair production, Eq. (3.32), and to the adiabatic loss,

[1] We are in debt with E. Khan for having provided the data shown in the figure.

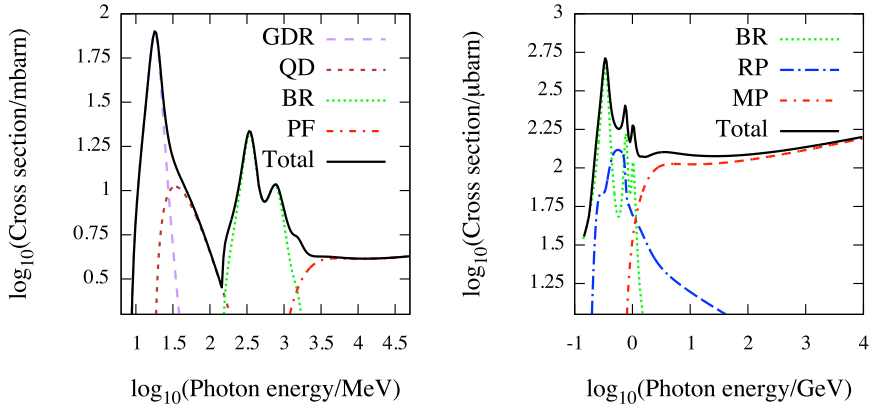


Figure 3.12: Comparison between the total cross section for Fe$\gamma_{\rm EBR}$ (left panel) and $p\gamma_{\rm EBR}$ (right panel) interactions as a function of the background photon energy ϵ' in the nucleus rest frame, obtained from our simulation code, following Rachen's parameterizations.

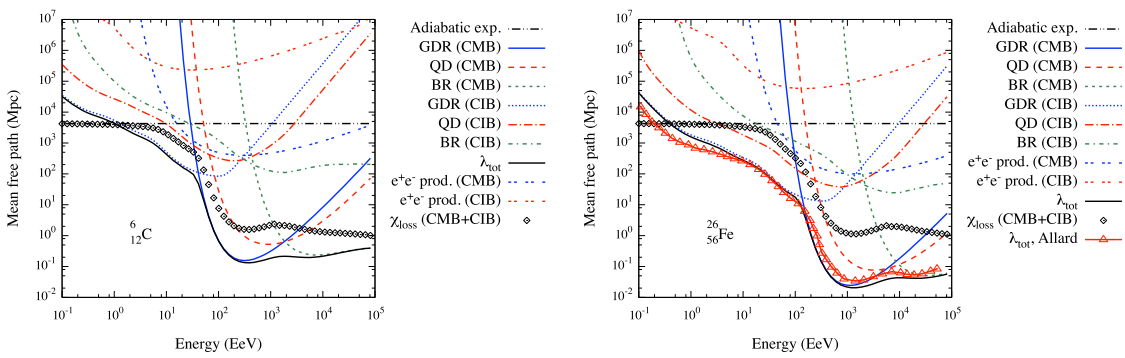


Figure 3.13: Our estimation of the mean free path λ at $z = 0$ as a function of the energy of the nucleus in the observer rest frame. The contributions due to different processes (adiabatic, pair and photo-pion production, as well as giant dipole resonance and quasi-deuteron effect for photo-disintegration) in CMB and CIOB are shown separately. The total interaction length $\lambda_{\rm tot}$ and the total energy loss length $\chi_{\rm loss}$ are shown as well in the case of Carbonium ${}^{12}_{6}C$ (left panel) and Iron ${}^{56}_{26}Fe$ (right panel) nuclei: for the latter, the $\lambda_{\rm tot}$ estimated by Allard et al [185, 186] is reported for reference.

Eq. (3.31), occurs in any case, with significant contributions only in a small range of energies. We treat the photo-pion production similarly to the case of protons by using Eq. (3.29) and the $\Delta-$baryon decay channels. The energy loss equation, defined by Eq. (3.28), still applies but coupled to the nuclear mass loss rate

$$\frac{1}{A}\frac{dA}{dz} = -\frac{dt}{dz}\beta_{\rm dis,eff}(z, E; Z, A), \tag{3.50}$$

leading to

$$\frac{1}{E}\frac{dE}{dz} = \frac{1}{\Gamma}\frac{d\Gamma}{dz} + \frac{1}{A}\frac{dA}{dz}. \tag{3.51}$$

An analytic approach for the estimation of the spectra at Earth, based on the numerical integration of such an equation, has been recently reported in Ref. [187]. In Fig. 3.13 we show our estimation of the

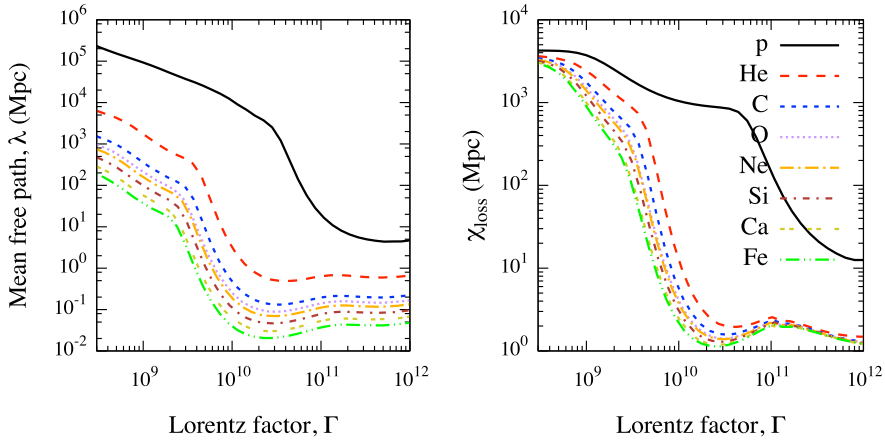


Figure 3.14: **Left:** Total mean free path $\lambda_{\rm tot}$ in CMB and CIOB for several nuclei at $z = 0$, from proton ($A = 1$) to iron ($A = 56$), as a function of the Lorentz factor Γ. **Right:** Same as in the left panel, but for the total energy loss length $\chi_{\rm loss}$.

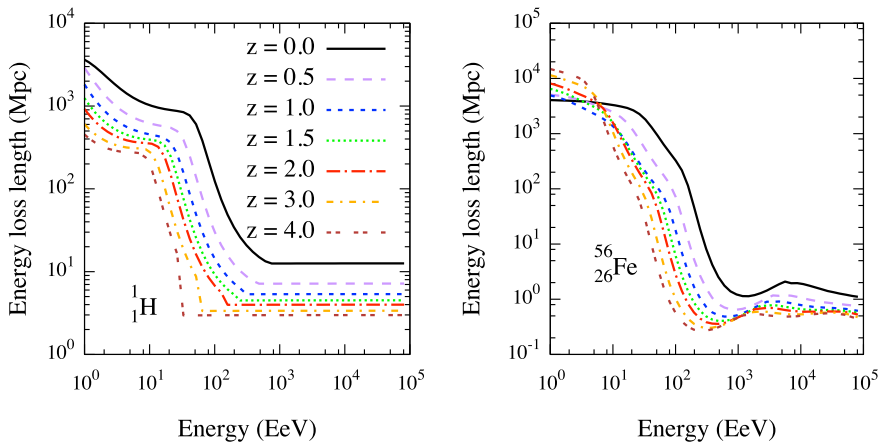


Figure 3.15: Evolution of the total energy loss length $\chi_{\rm loss}$ in CMB and CIOB as a function of the energy of the nucleus in the observer rest frame, for different values of the redshift. **Left:** Case of proton. **Right:** Case of iron.

interaction length λ at $z = 0$, as a function of the energy E of the nucleus in the observer rest frame, in the case of Carbonium ${}^{12}_{6}C$ (left panel) and Iron ${}^{56}_{26}Fe$ (right panel) nuclei. The contributions due to all processes discussed so far, in CMB and CIOB, are also shown separately, as well as total energy loss length $\chi_{\rm loss}$. In the case of iron, the figure shows that the main energy loss below 10^{19} eV is due to the adiabatic expansion of the universe, whereas photo-disintegration process through the giant dipole resonance dominates above 10^{19} eV and photo-meson production becomes dominant above 10^{22} eV.

The estimation of the total interaction (left panel) and energy loss (right panel) lengths at $z = 0$ obtained with HERMES are also shown in Fig. 3.14 as a function of the Lorentz factor Γ, for several nuclei, from proton to iron. Both quantities decrease for increasing nuclear mass and for any value of the energy, although the energy loss length tends to be the same for all nuclei above $\Gamma = 10^{11}$, approximately the value where baryonic resonances occur. In Fig. 3.15 we also show the evolution with redshift of the total energy loss length as a function of the energy, for a proton (left panel) and for an iron (right panel) nucleus, for different values of the redshift, ranging from $z = 0$ to $z = 4$.

3.4 Propagation of secondary neutrinos and photons

In the previous sections we have discussed the production of secondary particles, mainly electron/positron pairs and charged or neutral pions, because of the interaction between UHE nuclei and background photons during the propagation in the extragalactic space. In this section we will briefly describe the propagation of the decay products of pions, causing the flux of neutrinos and the electromagnetic cascade (involving pairs and photons) simulated with HERMES.

Neutrinos

The main mechanisms involved in the creation of neutrinos are the decay of charged pions

$$\begin{aligned} \pi^+ &\longrightarrow \mu^+ + \nu_\mu \longrightarrow e^+ + \nu_e + \bar{\nu}_\mu + \nu_\mu, \\ \pi^- &\longrightarrow \mu^- + \bar{\nu}_\mu \longrightarrow e^- + \bar{\nu}_e + \nu_\mu + \bar{\nu}_\mu, \end{aligned}$$

created from photo-pion production of nuclei, the $\beta-$decay of neutrons

$$n \longrightarrow p + e^- + \bar{\nu}_e,$$

and the instability processes of heavy nuclei previously described ($\beta-$decays and electron capture).

Neutrinos are chargeless particles with negligible mass, undergoing interactions only through the weak nuclear force (and gravity, if they are considered massive particles). Because of such features, neutrinos are likely to traverse the extragalactic space, even for cosmic distances, without interacting with background photons or interstellar medium, and without being deflected by magnetic fields: characteristics that makes neutrinos the ideal candidates for particle astronomy. On the other hand, the flux of cosmogenic neutrinos is relatively small if compared to the flux of charged particles, at the highest energy.

Propagation and energy loss of neutrinos, can be easily described by energy loss equation (3.28), considering only the adiabatic energy loss rate defined by Eq. (3.31).

Photons

High energy photons are produced from the decay of neutral pions ($\pi^0 \longrightarrow \gamma\gamma$). Although photons, as neutrinos, are chargeless and massless particles, they undergo interactions through electroweak force. In the specific case of photons, the cosmic radio background (CRB) plays a fundamental role at highest energy, whereas CMB and CIOB are relevant at smaller energies. Together with electrons and positrons, photons are responsible for the creation and the development of the electromagnetic cascade generated during the propagation of charged nuclei.

Their propagation is dominated by pair production (PP) and double pair production (DPP), despite the latter can be neglected for energies below 1000 EeV. In the case of electrons and positrons the relevant energy-loss processes are the inverse Compton scattering (ICS) and the triplet pair production (TPP). Because of its electromagnetic nature, the development of the cascade strongly depends on the strength of the extragalactic magnetic field: e^-/e^+ loose most of their energy by synchrotron radiation if the intensity of the field is strong enough[2]. The cascade develops no more when the synchrotron cooling rate becomes smaller than ICS one.

[2]For a particle with charge Z and mass m, the energy loss rate in a field of squared r.m.s. strength B^2 is

$$\frac{dE}{dt} = \frac{4}{3}\sigma_T \frac{B^2}{8\pi} \left(\frac{Zm_e}{m}\right)^4 \left(\frac{E}{m_e c^2}\right)^2, \tag{3.52}$$

where σ_T is the Thomson cross section.

At variance with the case of charged and massive nuclei, previously described in this chapter, it is not possible to define the interaction length of photon in terms of the Lorentz factor. In this case, the general expression for the interaction rate reads

$$\mathcal{R}(z,E) = c\int_{\epsilon_{\min}}^{\epsilon_{\max}} d\epsilon \int_{-1}^{1} \frac{d\cos\theta}{2} n(\epsilon,z)\sigma(s)(1-v_p\cos\theta), \tag{3.53}$$

where v_p is the velocity of the incident primary particle. After some algebra, we obtain

$$\mathcal{R}(z,E) = \frac{c}{8\xi E^2}\int_{\epsilon_{\min}}^{\epsilon_{\max}} d\epsilon \frac{n(\epsilon,z)}{\epsilon^2}\int_{s_{\min}}^{s_{\max}} \sigma(s)(s-m^2c^4)\,ds, \tag{3.54}$$

where $\sqrt{s}$ is the energy in the center of mass system and ξ is the velocity of the particle in speed of light units[3]. It is worth noticing that we are not defining s as the standard invariant squared momentum.

Photon pair production occurs when two photons interact to produce an electron-positron pair ($\gamma+\gamma_{\rm EBR} \longrightarrow e^- + e^+$). The total cross section for such a process, in the case of a primary photon with energy E scattering off (with angle θ) a background photon with energy ϵ, is given by

$$\sigma_{PP}(E,\theta) = \frac{3}{16}\sigma_T(1-\xi^2)\left[(3-\xi^4)\log\frac{1+\xi}{1-\xi} - 2\xi(2-\xi^2)\right], \tag{3.55}$$

where σ_T is the Thomson cross section and $\xi \equiv (1-4m_e^2c^4/s)^{\frac{1}{2}}$, with m_e the mass of the electron and $s = 2m_e^2c^4(1-\cos\theta)$. The interaction rate will be given by

$$\mathcal{R}_{PP}(z,E) = \frac{c}{8E^2}\int_{\epsilon_{\min}}^{\epsilon_{\max}} d\epsilon \frac{n(\epsilon,z)}{\epsilon^2}\int_{s_{\min}}^{s_{\max}} \sigma_{PP}(s)\,s\,ds, \tag{3.56}$$

with $s_{\min} = 4m_e^2c^4$ and $s_{\max} = 4E\epsilon$.

In inverse Compton scattering ($e+\gamma_{\rm EBR} \longrightarrow e+\gamma$), an electron (or positron) interacts with a background photon, accelerating it. The total cross section of ICS is given by the Klein-Nishina formula:

$$\sigma_{ICS}(E,\theta) = \frac{3}{8}\sigma_T\frac{m_e^2c^4}{s\xi}\left[\frac{2}{\xi(1+\xi)}(2+2\xi-\xi^2-2\xi^3) - \frac{1}{\xi^2}(2-3\xi^2-\xi^3)\log\frac{1+\xi}{1-\xi}\right], \tag{3.57}$$

where $\xi \equiv (s-m_e^2c^4)/(s+m_e^2c^4)$ is the velocity of the outgoing electron in the center of mass frame, in speed of light units. It is worth noticing that the process has no threshold. Such a formula is valid for $s > 2m_e^2c^4 + 2m_ec^2\epsilon_{\max}$, whereas below this threshold $\sigma_{ICS} = \sigma_T$. The corresponding interaction rate is given by

$$\mathcal{R}_{ICS}(z,E) = \frac{c}{8\xi E^2}\int_{\epsilon_{\min}}^{\epsilon_{\max}} d\epsilon \frac{n(\epsilon,z)}{\epsilon^2}\int_{s_{\min}}^{s_{\max}} \sigma_{ICS}(s)(s-m_e^2c^4)\,ds, \tag{3.58}$$

with $s = 2E\epsilon(1-\xi\cos\theta) + m_e^2c^4$ and $s_{\rm max,min} = 2E\epsilon(1\pm\xi) + m_e^2c^4$.

The total cross section for triplet pair production ($e+\gamma_{\rm EBR} \longrightarrow e+e^-+e^+$) is given by

$$\sigma_{TPP}(E,\theta) = \frac{3\alpha}{8\pi}\sigma_T\left[\frac{28}{9}\log\left(\frac{s}{m^2c^4}\right) - \frac{218}{27}\right], \tag{3.59}$$

if $s >> m_e^2c^4$. The energy threshold of the process is $s_{\min} = 9m_e^2c^4$, whereas the maximum center of mass energy, obtained in head on collisions, is given by $s_{\max} = 2m_e^2c^4 + 2m_ec^2\epsilon_{\max}(1+\xi)$, with an

[3]By noticing that $s = m^2c^4 + 2mc^2\epsilon'$ in rest frame of the particle, and that $ds = 2mc^2d\epsilon'$, the analytical expression previously adopted in the case of nuclei is recovered.

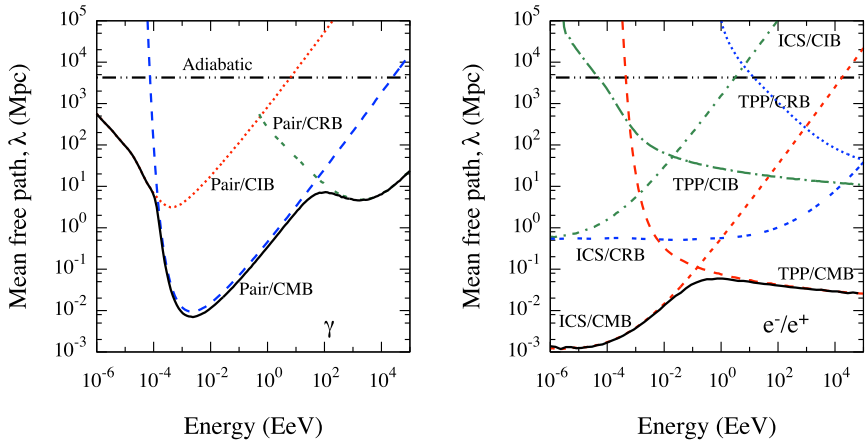


Figure 3.16: Mean free path λ at $z = 0$ as a function of the energy of the cosmic ray in the observer rest frame. The contributions due to different processes in CMB, CIOB and CRB are shown separately. The total interaction length $\lambda_{\rm tot}$ (solid line) is shown in the case of pair production by photon (left panel), and triple pair production (TPP) and inverse Compton scattering (ICS) by e^-/e^+ (right panel).

interaction rate whose analytical expression is the same as in Eq. (3.58). However, the TPP process has inelasticity $\kappa(E) \simeq 1.768(s/m_e^2c^4)^{-3/4}$ for $s >> m_e^2c^4$: although the total cross section for TPP on CMB photons becomes comparable to the ICS cross section already around 10^{17} eV, the energy attenuation is not important up to $\approx 10^{22}$ eV because $\kappa < 10^{-3}$. Moreover, TPP is dominated over by synchrotron cooling, and therefore negligible, if the electrons propagate in a magnetic field of r.m.s. strength larger than 10^{-3} nG [40].

For a detailed description of processes discussed in this section, as well as other important (but less relevant) processes for the development of the electromagnetic cascade, we refer to Ref. [188] and Refs. therein.

In Fig. 3.16 we show the interaction length[4] λ at $z = 0$, as a function of the energy E of the cosmic ray in the observer rest frame, in the case of photon (left panel) and e^-/e^+ (right panel). The contributions due to all processes discussed so far, in CMB, CIOB and CRB, are also shown separately. In the case of photon, the production of pairs is the dominant process: the figure shows that below 10^{14} eV the relevant background radiation is the CIOB, with an interaction length ranging from a few Mpc to almost 1 Gpc at TeV energy. For energies up to tens of EeV the CMB is the only relevant background radiation, with an interaction length smaller than 1 Mpc and, in particular, of the order of tens kpc at PeV energy. At higher energy, above 100 EeV, the CRB is the only background radiation relevant for energy loss, with interaction length of the order of a few Mpc. In practice, HE photons produced by extragalactic sources can not be observed at Earth, in the energy range between hundreds TeV and ≈ 0.1 EeV. Closest extragalactic sources of UHE photon can be still observed, provided that their distance is within a few tens of Mpc and the energy of photons is larger than ≈ 1 EeV.

In the case of electrons and positrons, the interaction length is very small for any value of the energy, ranging from ≈ 1 kpc at TeV energy to at most ≈ 100 kpc at higher energies. The relevant energy-loss process are the inverse Compton scattering on CMB up to ≈ 0.1 EeV and the triple pair production on CMB at higher energy.

[4]We are in debt with M. Settimo for having provided the numerical estimation of mean free paths shown in the figure.

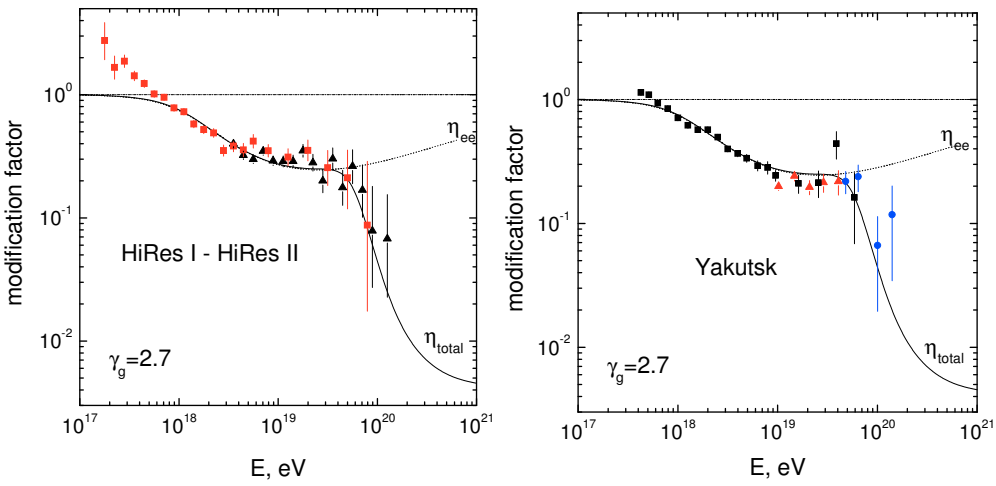


Figure 3.17: Modification factor for the power-law generation spectra with injection index 2.7 in comparison with the some observational data, Hires I-II (left) and Yakutsk (right), for non-evolutionary models. Curve $\eta = 1$ corresponds to adiabatic energy losses, curves η_{ee} to adiabatic and pair production energy losses and curves $\eta_{\rm total}$ to total energy losses [189].

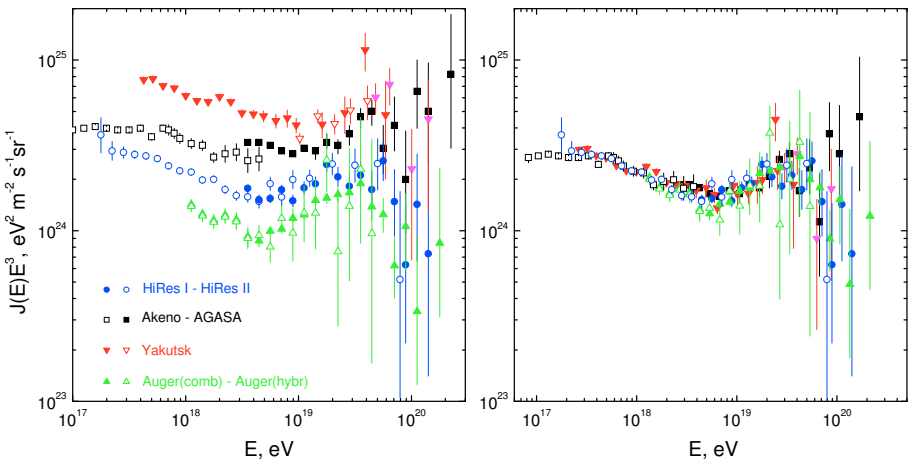


Figure 3.18: The spectra and fluxes measured by Yakutsk, AGASA, HiRes and Auger before (left panel) and after (right panel) energy calibration due to the dip interpretation [190].

3.5 Impact of propagation on the spectrum of UHECR

The energy loss process discussed in this chapter should produce observable features in the spectrum of UHECR: in this section we will treat separately the case of nuclei and the case of neutrinos and γ-rays.

3.5.1 Protons and heavy nuclei

If UHECR are assumed to be protons, an interesting feature called "dip" is expected to be observed if the general assumptions on the transition from galactic to extragalactic cosmic rays are revised [191–193]. The dip should emerge in the diffuse spectrum of UHE protons in the energy range between 10^{18} eV and 4×10^{19} eV, and it is caused by electron-positron pair production on CMB photons.

If we define the *modification factor* η as the ratio between the full spectrum $J(E)$, incorporating energy losses, and the unmodified spectrum $J^*(E)$, where only adiabatic loss is considered. The pre-

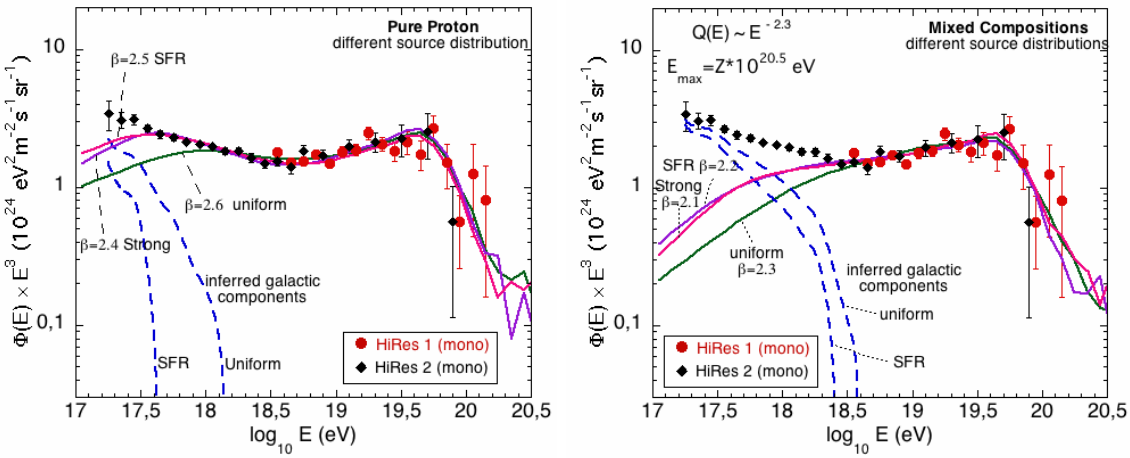


Figure 3.19: Propagated spectra, $E^3\Phi(E)$, for pure proton models (left) and mixed composition models (right) compared with HiRes monocular data. Different source evolution models are indicated by the labels [185].

dicted modification factor for protons is shown in Fig. 3.17 in comparison with modification factors of some observational data, namely Hires I-II (left panel) and Yakutsk (right panel): the agreement is evident. For heavy nuclei the dip is quite different and does not agree with such data, although a mixed composition with a small fraction of heavier nuclei is still allowed.

The predicted shape of the dip is robust, being valid for the rectilinear and diffusive propagation, for different discretenesses in the source distribution, for local source over-density and deficit and for source inhomogeneities on scales smaller than 100 Mpc. By definition $\eta(E) \leq 1$, whereas for $E < 10^{18}$ eV it exceeds 1, a clear indication of an additional component to the flux, of galactic origin. Being thought of as purely galactic feature, the position of the second knee in the dip analysis appears as direct consequence of extragalactic proton energy losses [191]. By adopting the dip prescription, the fluxes of UHECR observed by different experiments can be recalibrated: in Fig. 3.18 such experimental spectra are shown before (left panel) and after (right panel) such a recalibration, and a nice overall agreement emerges [189, 190].

In Fig. 3.19 the propagated spectra are shown for pure proton models (left panel) and mixed composition models (right panel), for different source evolution models and injection indices, in comparison with the HiRes monocular data. The corresponding galactic components are inferred from the overall spectrum by subtracting the extragalactic cosmic ray component, in the case of the uniform and SFR source evolution models [185]. The effect of source evolution is stronger at lower energy in both cases. Moreover, it is worth noticing that the two cases infer very different regions for the galactic component: in the case of protons such a component sharply decreases, becoming negligible around $10^{17.6} - 10^{18.1}$ eV, depending on the evolution model, whereas in the case of mixed composition a smoother decrease is found, and the inferred component becomes negligible around $10^{18.4} - 10^{18.6}$ eV. If the dip model is used to explain such a difference and the composition is light, there should be no transition from the galactic to extragalactic component around $10^{18.5}$ eV.

3.5.2 Neutrinos and $\gamma-$rays

In Fig. 3.20 the spectra of nuclei, neutrinos and photons expected at Earth recently obtained by Decerprit and Allard from detailed simulations [194], are shown.

The cosmic ray, neutrino (summed over all flavors) and photon spectra are shown in the left panel of Fig. 3.20 assuming pure proton, pure helium and pure iron composition at the source, a source spectral

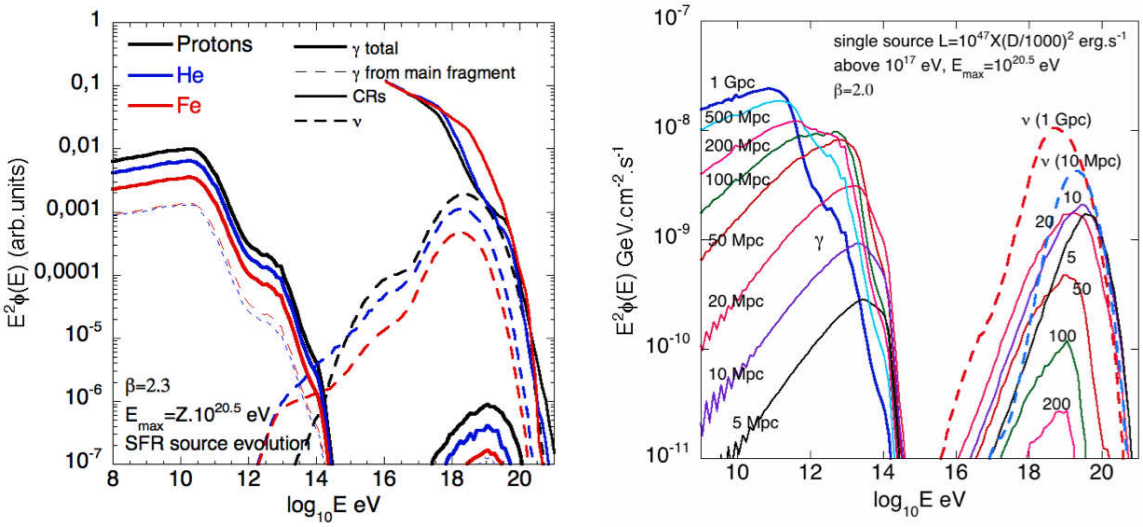


Figure 3.20: Spectra of nuclei, neutrinos and photons expected at Earth recently obtained by Decerprit and Allard from detailed simulations [194]. **Left:** Fluxes obtained by assuming different pure compositions at the source. **Right:** Fluxes for the pure proton (blue) and the pure iron (red) cases, assuming sources located at various distances ($D = 5$, 10, 20, 100, 200, 500, 1000 Mpc). See the text for further details.

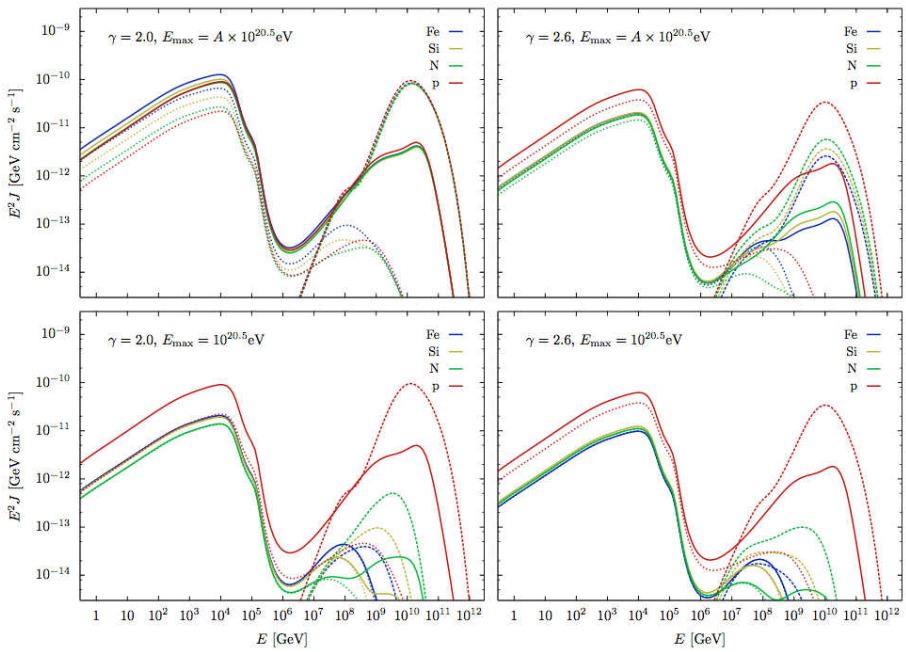


Figure 3.21: Spectra of neutrinos and photons expected at Earth for a point source of primary nuclei, placed at $z = 0.01$ ($\simeq 40$ Mpc), recently obtained by Ahlers and Salvado from simulations [178] based on an analytical approach [174]. The solid line shows the total flux of γ-rays and the dashed line the cosmogenic neutrino flux. The dotted lines show the contribution to γ-rays from Bethe-Heitler loss alone omitting the contribution from contribution from photo-pion interactions. See the text for further details.

index 2.3 and a maximum energy at the source $E_{\max}(Z) = Z \times 10^{20.5}$ eV. The same cosmic ray luminosity between 10^{16} eV and $E_{\max}(Z)$ is assumed. The contribution of the main fragment is shown in thin dashed

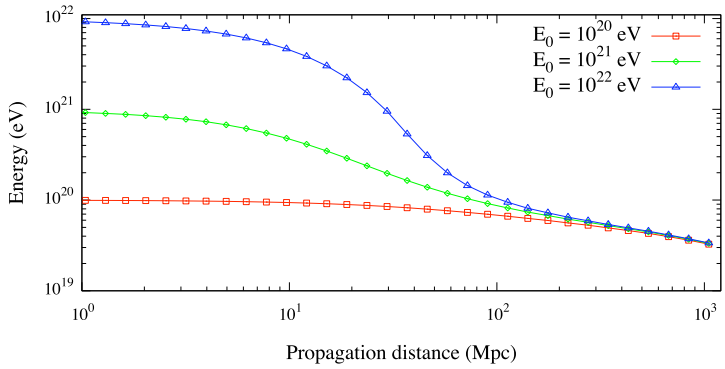


Figure 3.22: Proton energy as a function of propagation distance through the CMB for three different values of injection energy E_i, namely 10^{20}, 10^{21} and 10^{22} eV, obtained with our simulator.

lines. With no regards for the primary composition, the resulting photon spectra appear very similar, although in the case of heavy nuclei, most of the photon flux is caused by secondary nucleons.

The cosmogenic neutrinos (dashed lines) and photons (solid lines) fluxes for the pure proton (blue) and the pure iron (red) cases, assuming sources located at various distances ($D = 5$, 10, 20, 100, 200, 500, 1000 Mpc) and luminosity of $10^{47} \times (D/1000 \text{ Mpc})^2$ erg s^{-1}, between 10^{17} eV and $E_{\max} = Z \times 10^{20.5}$ eV, are shown in the right panel of Fig. 3.20.

The spectra of neutrinos and photons expected at Earth for a point source of primary nuclei, placed at $z = 0.01$ ($\simeq 40$ Mpc), recently obtained by Ahlers and Salvado from simulations [178] based on an analytical approach [174], are shown in Fig. 3.21. The injection spectrum of primary nuclei has been assumed to be of the form $Q_A(E) \propto E^{-\gamma} \exp(-E/E_{\max})$, where the values of γ and $E_{\max}$ are specified in the plots. The solid line shows the total flux of γ-rays and the dashed line the cosmogenic neutrino flux. The dotted lines show the contribution to γ-rays from Bethe-Heitler loss alone omitting the contribution from contribution from photo-pion interactions.

We refer to the original papers [178, 194] for more information about such results, their interpretation and their comparison with experimental observation.

3.6 Investigating the GZK horizon of UHE protons

The solution of Eq. (3.28) is the degraded energy E_f of the proton with initial energy E_i, after the propagation from the source at distance D (or at redshift z) to the Earth ($z = 0$). In Fig. 3.22 we show the proton energy as a function of propagation distance through the CMB for three different values of injection energy E_i, ranging from 10^{20} to 10^{22} eV. Results from HERMES show that a proton propagating for at least 200 Mpc, will be detected with energy $E_f \leq 60 \times 10^{18}$ eV, with no regards if its initial energy is larger, for instance, than 10^{22} eV. Such results are in perfect agreement with literature [195].

Under the assumption of a power-law injection spectrum with spectral index s, we have

$$\omega_{\text{GZK}}(z; E_f) = \frac{s-1}{E_f^{-s+1}} \int_{E_i(z;E_f)}^{\infty} E^{-s} dE, \tag{3.60}$$

while assuming equal intrinsic luminosity and homogenous distribution of sources (i.e. their number at

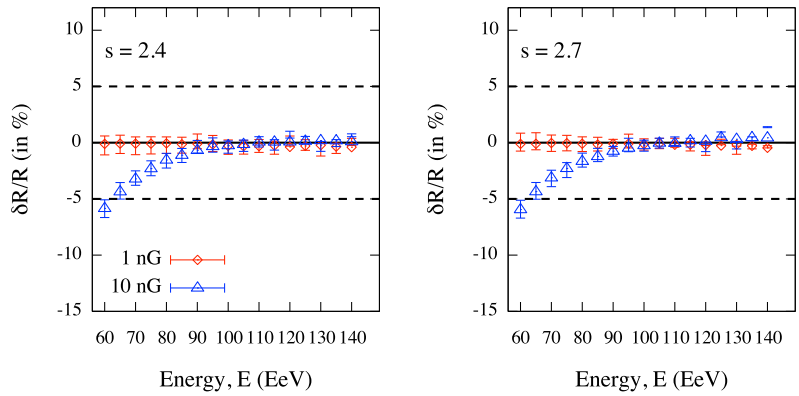


Figure 3.23: Relative difference between GZK horizons estimated with and without EMF. Two values of the r.m.s. strength of the EMF are considered, namely 1 and 10 nG, as well as two different injection index, namely 2.4 (left panel) and 2.7 (right panel). Error bars indicates only the statistical uncertainty on $\delta R/R$. Lines corresponding to $\pm 5\%$ bands and $\delta R/R = 0$ are shown for reference.

redshift z is proportional to z^2), we obtain the function

$$\Omega_{\rm GZK}(z; E_f) = \frac{\int_z^\infty dz' \int_{E_i(z';E_f)}^\infty E^{-s} dE}{\int_0^\infty dz' \int_{E_i(z';E_f)}^\infty E^{-s} dE}, \tag{3.61}$$

for the probability of detecting at $z = 0$ a proton emitted within distance z and with $E \geq E_f$ [196]. The energy $E_i\,(z; E_f)$ of the injected proton is estimated by evolving Eq. (3.28) backward in time [171, 197]. The GZK horizon R is the distance such that $1 - \Omega_{\rm GZK}(R, E_f) = 0.9$.

3.6.1 Influence of extragalactic magnetic field

It is worthwhile discussing in this section the impact of the extragalactic magnetic field (EMF) on the probability $\Omega_{\rm GZK}(z; E_f)$. Unfortunately, a direct or indirect measurement of the EMF is still missing, and only bounds to its r.m.s. strength $B_{\rm rms}$ have been estimated. However, bounds on the EMF strength also depend on the field correlation length ℓ, which is also unknown.

Upper limits of the order of 10^{-9} G (1 nG) on the intensity of the EMF have been measured through i) the Faraday rotation in the polarized radio emission from distant quasars [198] and ii) the characteristic distortions that it induced on the spectrum and the polarization of CMB radiation [199]. More recent analyses, based on the recent WMAP measurements [200], provide the more stringent upper bound of $\approx$2 nG on the present value of the cosmic magnetic field of primordial origin, more than one order of magnitude smaller than previous estimates [201].

Intergalactic magnetic field might also be structured inside and around clusters or groups of galaxies, with filaments extending over few Mpc, as shown, for instance, in recent detailed simulations [107]. The topology of such a structured magnetic field would have a non-negligible impact on the trajectories of UHE protons and, of course, on the average deflections they experience in the case of EMF with regular structures above 200 kpc [202]. Additionally, longitude-averaged X-ray emission observed with ROSAT near 0.65 keV and 0.85 keV towards the center of the Galaxy, are in agreement with a Galactic wind thermally-driven by cosmic rays and hot gas [203, 204]. In our study, we can neglect the effect of such a magnetic wind, because it is expected to have a non-negligible impact for protons below 60 EeV or heavier nuclei. Moreover, we neglect the case of a structured EMF because of the lack of direct or indirect measurements about its structure, and we reserve to consider it in a future work.

Here, we consider only the case of a turbulent EMF characterized by $B_{\rm rms}$ and ℓ. In this case, protons propagate randomly following a brownian trajectory. The average deflection they are subjected to can be parameterized by

$$\theta(E,D) \simeq 0.8^\circ \left(\frac{E}{10^{20}\ \rm eV}\right)^{-1} \left(\frac{D}{10\ \rm Mpc}\right)^{\frac{1}{2}} \left(\frac{\ell}{1\ \rm Mpc}\right)^{\frac{1}{2}} \left(\frac{B_{\rm rms}}{10^{-9}\ \rm G}\right), \tag{3.62}$$

in absence of energy losses [40], with $D \approx zc/H_0$ for distances considered in the present study. Such a deflection implies an average time delay

$$\tau(E,D) \simeq 1.5 \left(\frac{E}{10^{20}\ \rm eV}\right)^{-2} \left(\frac{D}{10\ \rm Mpc}\right)^{2} \left(\frac{\ell}{1\ \rm Mpc}\right) \left(\frac{B_{\rm rms}}{10^{-9}\ \rm G}\right)^{2}\ \rm kyr, \tag{3.63}$$

relative to the rectilinear propagation with the speed of light. Let us consider a propagation distance equal to the GZK radius, i.e. $D = R$: the difference between the trajectory corresponding to the propagation in absence of magnetic field and the brownian trajectory of the proton propagating in the EMF is $\delta R \approx c\tau$, leading to

$$\frac{\delta R}{R} \simeq 0.5 \times 10^{-4} \left(\frac{E}{10^{20}\ \rm eV}\right)^{-2} \left(\frac{R}{10\ \rm Mpc}\right) \left(\frac{\ell}{1\ \rm Mpc}\right) \left(\frac{B_{\rm rms}}{10^{-9}\ \rm G}\right)^{2}, \tag{3.64}$$

where we have omitted to explicitly report the dependence of R and δR on the energy E, for simplicity. We estimate the bound to $B_{\rm rms}\sqrt{\ell}$ in order to have a negligible impact on the GZK horizon, i.e. $\delta R/R < 5\%$, by

$$\sqrt{\frac{\ell}{1\ \rm Mpc}}\frac{B_{\rm rms}}{10^{-9}\ \rm G} < 10^{\frac{3}{2}} \left(\frac{E}{10^{20}\ \rm eV}\right) \left(\frac{R}{10\ \rm Mpc}\right)^{-\frac{1}{2}}. \tag{3.65}$$

For a UHE proton with $E = 60$ EeV the GZK horizon is $R \approx 180$ Mpc: if we consider the propagation in an EMF characterized by $\ell = 1$ Mpc, any r.m.s. strength $B_{\rm rms} < 4.5$ nG will not significantly affect the horizon. For a proton with higher energy as 95 EeV, such a bound reads $B_{\rm rms} < 11.5$ nG, i.e. even more intense EMFs have a negligible impact on the result. For instance, the estimated upper bounds are still conservative even if larger correlation length as 16 Mpc are considered, reducing to ≈1 nG and ≈3 nG, respectively.

We have investigated the validity of such an argument by performing several Monte Carlo realizations of protons propagating in the Universe, with and without EMF. For such a purpose, the well known propagation software CRPropa v1.4 [205] has been adopted: two different values of the r.m.s. strength of the EMF, namely 1 and 10 nG, as well as two different values of the injection index, namely 2.4 and 2.7, has been considered for this study. In any case, the correlation length has been fixed to $\ell = 1$ Mpc. The main advantage in using CRPropa is that relevant energy-loss processes are considered during the propagation in a turbulent magnetic field: in this case we can quantify the average relative deviation $\delta R/R$ in a more realistic scenario with respect to Eq. (3.64), where energy loss has been switched off to estimate $\theta(E,D)$ and $\tau(E,D)$. In Fig. 3.23 we show the result of such a study. It is evident that the the injection index has a negligible impact on the relative deviation $\delta R/R$. For both $B_{\rm rms} \simeq 1$ nG and $B_{\rm rms} \simeq 10$ nG, the resulting relative deviation is within 5% at any energy, in agreement with expectation. It is worthwhile noticing that in any case, the value of $\delta R/R$ is non-positive, i.e. the GZK horizon for the propagation in an EMF is smaller than the horizon for the propagation without magnetic field, as expected. Moreover, it is relevant to remark that for the weak EMF (1 nG) $\delta R/R \approx 0$, whereas for the strong EMF (10 nG) $\delta R/R$ approaches zero above 90 EeV: the EMF has no impact on the GZK horizon of UHE protons.

From such a result, we can also deduce that above 60 EeV the effect of energy loss on the propagation in an EMF is negligible: hence, the parameterizations given by Eq. (3.62) and (3.63), although they have been obtained by considering no energy losses, can be safely used in the scenarios previously described.

Such a result is rather conservative: in fact, the most up-to-date bounds on the correlation length and the r.m.s. strength of the EMF are much lower of the values considered in this study [206].

3.6.2 Influence of cosmology

In Fig. 3.24 we show the weight function $\Omega_{\rm GZK}(z; E_f)$ for some values of the energy threshold $E_f = E_{\rm thr}$, as a function of the distance. By assuming the most recent values of ΛCDM model parameters ($\Omega_b = 0.0456$, $\Omega_c = 0.227$, $\Omega_\Lambda = 0.728$ and $H_0 = 70.4$ km/s^{-1}/Mpc) [200], the value of the injection index is fixed ($s = 2.7$) while the energy threshold $E_{\rm thr}$ is varied (left panel): as expected, the GZK horizon decreases by increasing the energy threshold. In the right panel of Fig. 3.24, we show the weight for two fixed values of the energy threshold, $E_{\rm thr} = 50$ EeV and $E_{\rm thr} = 100$ EeV, while the injection index is varied: although differences among curves corresponding to different values of the injection index are not exaggerated as in the previous case, we find that they increase by decreasing the energy threshold $E_{\rm thr}$.

Cosmological parameters as the density of matter, the density of dark energy and the Hubble parameter at the present time, should influence the function $E_i\,(z; E_f)$, discussed at the end of the previous section, and of consequence the probability $\Omega_{\rm GZK}(z; E_f)$. From Eq. (3.28) it is clear that the factor involving such parameters is only $-dt/dz$.

In a flat universe, we consider the deceleration parameter q_0 constrained by $\Omega_M + \Omega_\Lambda = 1$, whereas in a curved universe, we consider the curvature parameter Ω_κ constrained by $\Omega_\kappa = 1 - \Omega_M - \Omega_\Lambda$. By varying two among the three parameters, we investigate different models of universes. In Tab. 3.2 and Tab. 3.3 the values of the parameters are summarized for flat and closed models, respectively, that will be adopted in the successive analysis.

For small values of z, corresponding to the nearby universe, it is not expected a significant difference among cosmological models. At the highest energies, above 50 EeV, we have found that photomeson production dominates energy losses, thus only the term $-\beta_\pi(z, E) \times dt/dz$ is important. By considering two extreme scenarios, as matter-dominated ($\Omega_M = 1; \Omega_\Lambda = 0$) or energy-dominated ($\Omega_M = 0; \Omega_\Lambda = 1$) universes, the energy-loss term is modified at most by 10%.

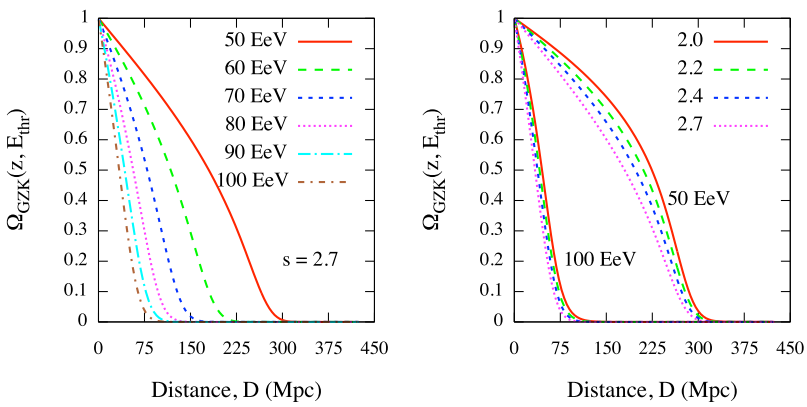


Figure 3.24: Surviving probability, defined by Equation (3.61), due to cosmological effects described in the text, in the case of a proton and assuming ΛCDM model parameters. *Left panel:* the value of the injection index is fixed ($s = 2.7$) while the energy threshold E_f is varied; *Right panel:* the value of the energy threshold E_f is fixed (two cases, $E = 50$ EeV and $E = 100$ EeV) while the injection index is varied.

Name	Ω_M	Ω_Λ	q_0	Description
ΛCDM	0.272	0.728	-0.592	Standard cosmological model
EdS	1	0	0.5	Einstein-de Sitter model (matter-dominated)
AFU1	0	1	-1	Decelerating flat Universe (vacuum energy-dominated)
AFU2	$\frac{1}{3}$	$\frac{2}{3}$	-0.5	Decelerating flat Universe

Table 3.2: Models for flat universes considered in this study and corresponding to different values of the density of matter, the density of dark energy and the deceleration parameter.

Name	Ω_M	Ω_Λ	Ω_κ	Description
FLO1	0.27	0.93	-0.2	Friedmann-Lamaitre open
FLO2		1.23	-0.5	
FLO3		1.43	-0.7	
FLC1	0.27	0.53	0.2	Friedmann-Lamaitre closed
FLC2		0.23	0.5	
FLC3		0.03	0.7	

Table 3.3: Models for curved universes.

In Fig. 3.25 we show the weight function $\Omega_{\rm GZK}(z; E_f)$ as a function of the distance, for both flat and curved models listed in Tab. 3.2 and 3.3, respectively. With no regards for the value of the injection index s, weights corresponding to very different cosmological models do not differ significantly below the GZK horizon, which is not influenced by the choice of a particular cosmology. The difference becomes significant toward higher redshifts where, however, the surviving probability for a proton is very small.

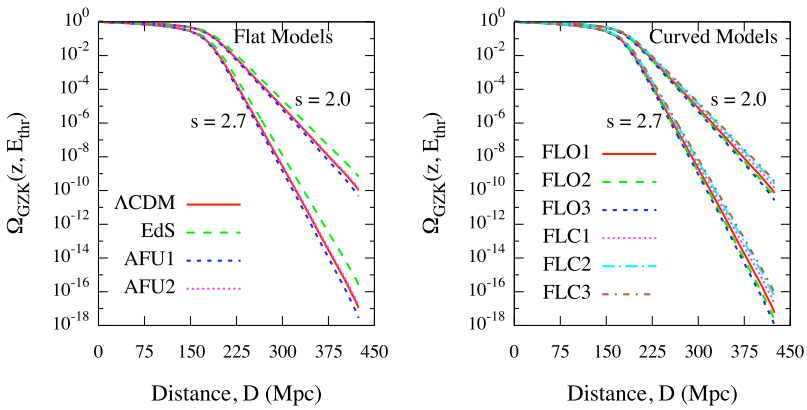


Figure 3.25: Surviving probability, defined by Equation (3.61), due to cosmological effects described in the text, in the case of a proton with $E_{\rm thr} = 60$ EeV and assuming different models of universe, for two values of the injection index ($s = 2.0$ and $s = 2.7$). *Left panel:* flat models described in Tab. 3.2 are considered; *Right panel:* curved models described in Tab. 3.3 are considered.

Indeed, in Fig. 3.26 the same weight function is shown by assuming ΛCDM model values for density parameters [200], protons with energy threshold $E_{\rm thr} = 60$ EeV and different values of the Hubble constant H_0 at the present time. For simplicity, in the following we will not explicitly specify the unit of H_0, that should be considered km s^{-1} Mpc^{-1}. Left panel shows the case for $s = 2.0$, whereas right panel shows the case for $s = 2.7$. It is evident that the value of the Hubble parameter significantly

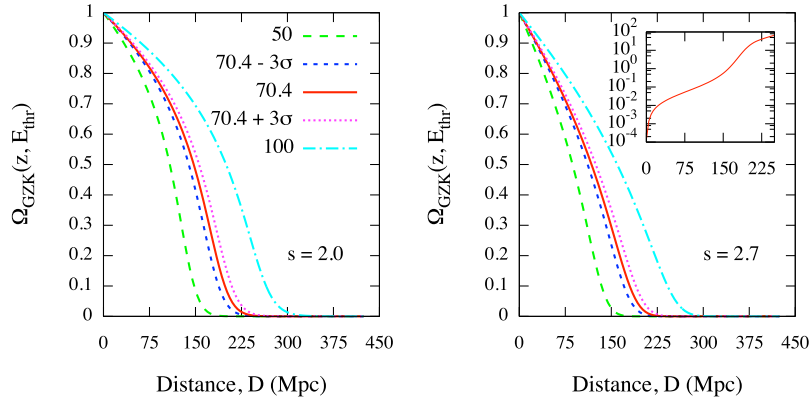


Figure 3.26: Surviving probability, as in Fig. 3.25, assuming ΛCDM model values for density parameters and protons with energy threshold $E_{\rm thr} = 60$ EeV, for two values of the injection index ($s = 2.0$ and $s = 2.7$). Different values of the Hubble parameter H_0 (in km/s^{-1}/Mpc units) at the present time are considered. The inset in the right panel shows the relative difference between curves corresponding to $H_0^+ = 70.4 + 3\sigma$ and $H_0^- = 70.4 - 3\sigma$, as a function of the distance.

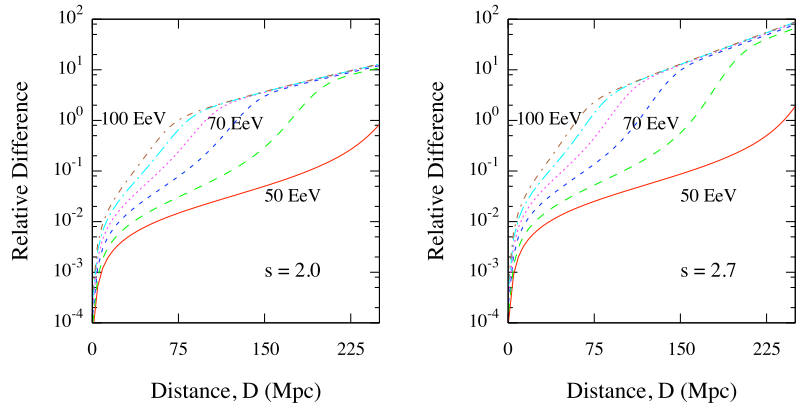


Figure 3.27: Relative difference between curves corresponding to probability functions for H_0^+ and H_0^- in Fig. 3.26, for different values of energy threshold in the case of $s = 2.0$ (left panel) and $s = 2.7$ (right panel).

affects the weight function, as already suggested by previous studies [168, 207]. If the 3σ uncertainty in the value ($H_0 = 70.4$) obtained from the ΛCDM model is taken into account [200], as shown in both panels of Fig. 3.26, we find that differences between curves corresponding to $70.4 + 3\sigma$ and $70.4 - 3\sigma$ are negligible ($\approx 5\%$) only below 75 Mpc, as shown in the inset of the right panel, for $s = 2.7$. From the lower bound curve ($H_0^- = 70.4 - 3\sigma$) the estimated GZK horizon is about 166 Mpc, whereas for the upper bound curve ($H_0^+ = 70.4 + 3\sigma$) it is about 186 Mpc, with a relative difference of about 11%. A relative difference of about 100% between the corresponding weight functions is reached around 160 Mpc, tending to increase with the distance.

In Fig. 3.27 we show the relative differences between curves corresponding to weight functions for H_0^+ and H_0^-, by varying the energy threshold and the injection index, $s = 2.0$ (left panel) and $s = 2.7$ (right panel). By increasing the energy threshold the relative difference increases, although it remains smaller than 15% below the GZK horizon. All significant differences are found to be above the GZK horizon, where only 10% of protons are expected to be produced. In the last chapter we will discuss two simple applications of the arguments discussed so far, in order to show how current and future experiments could probe the value of the Hubble parameter.

3.6.3 Conclusion

The GZK effect plays a fundamental role in the search of sources of UHECRs. Here, we have investigated the influence of cosmology on the GZK horizon of extragalactic UHE protons, with energy ranging from 50 to 100 EeV. By considering very different models of universe, from flat to curved ones, we have shown that significant differences among cosmological models appear to be important above the GZK horizon, where the surviving probability for the protons is very small. Moreover, we have investigated the impact of uncertainty in the Hubble parameter at the present time, according to the ΛCDM model of Universe and to the experimental constraints obtained from recent WMAP observations. Our results suggest the existence of non-negligible differences between the estimated values of the GZK horizon in universes with Hubble parameter $H_0 = 70.4 + 3\sigma$ and $H_0 = 70.4 - 3\sigma$, respectively. However, our numerical results show that such differences should have a small impact on studies involving distances below 250 Mpc, as for instance the recent correlation analyses between observed data and the distribution of nearby active galactic nuclei (AGN) reported by the Pierre Auger and the HiRes collaborations [88, 89, 208, 209].

3.7 Comparison of HERMES with other propagation codes

In this section, we show the comparison between results obtained with HERMES and those from other propagation codes available in the UHECR community. Four representative tests are reported: in all cases magnetic fields have been switched off.

For the first check of consistency, we consider a single source at redshift $z = 0.1$ (≈ 42 Mpc) emitting protons with a hard power-law injection, whose spectral index is 1.4 and the maximum energy achievable is $10^{20.5}$ eV. The resulting flux at Earth is shown in Fig. 3.28, where we compare our result with that one obtained by using CRPropa v1.4 [205], a Monte Carlo simulator designed for the propagation of protons in a magnetized Universe. The result is rather encouraging: the energy spectrum simulated with HERMES, here multiplied by the factor $E^{1.4}$, nicely reproduces that simulated with CRPropa over (almost) two order of magnitude, with the expected *dip* and *bump* features. The differences at the highest energy are just amplified by the multiplicative factor, the unmodified spectra being in excellent agreement within the corresponding statistical uncertainty.

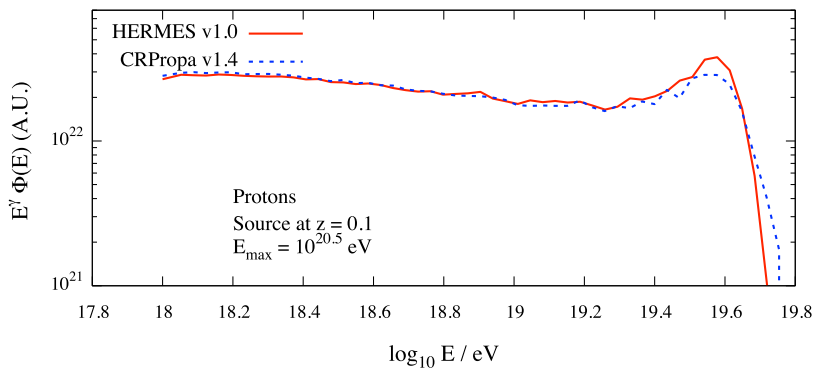


Figure 3.28: Simulated energy spectrum, multiplied by the factor $E^{1.4}$, observed at Earth. The comparison between HERMES and CRPropa v1.4 is shown.

In the second check of consistency, we investigate the surviving probability of protons, i.e. the probability that a proton produced by a sources at distance D could reach the Earth with an energy above a given threshold. We consider an homogenous distribution of equal-intrinsic-luminosity sources in the nearby Universe, up to ≈ 300 Mpc: each source emits protons following a power-law injection

spectrum with spectral index 2.4 and energy cutoff 10^{21} eV. Hence, we estimate the attenuation factor defined by Eq. (3.60) for different energy threshold $E_{\rm thr}$ at Earth, ranging from 60 EeV to 100 EeV. The result is shown in Fig. 3.29, where a comparison between HERMES, CRPropa v1.4 and the simulations kindly provided by D. Allard, adopting his own Monte Carlo propagator, are reported. The resulting curves are in very nice agreement among each other, putting in evidence the goodness of our simulator.

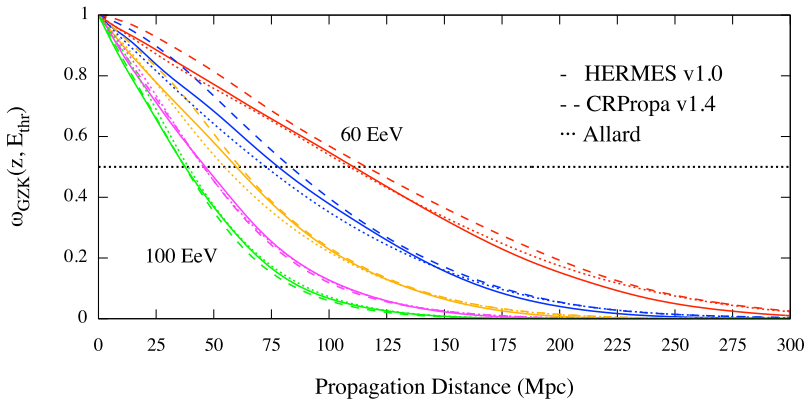


Figure 3.29: Surviving probability of protons (see Eq. 3.60) produced by an homogenous distribution of sources within 300 Mpc, as a function of the propagation distance and for different energy threshold $E_{\rm thr}$ at Earth. A power-law injection with energy cutoff 10^{21} eV and spectral index 2.4 is used. We show the result of our simulations performed with HERMES (solid line), compared to those obtained with CRPropa v1.4 (dashed line) and by D. Allard (dotted line).

In the previous tests we have only considered the comparison against other pure Monte Carlo propagators. In the third test we compare the same surviving probability, against a propagator which adopts the continuous energy loss (CEL) approximation for both pair and photo-pion production. We consider the same simulation setup of the second test, but we focus only on results obtained for an energy threshold of 60 EeV. In the left panel of Fig. 3.30 we show the resulting surviving probability, obtained by HERMES, CRPropa v1.4 and Allard, against the curve kindly provided by S. Mollerach, obtained by adopting the CEL approximation.

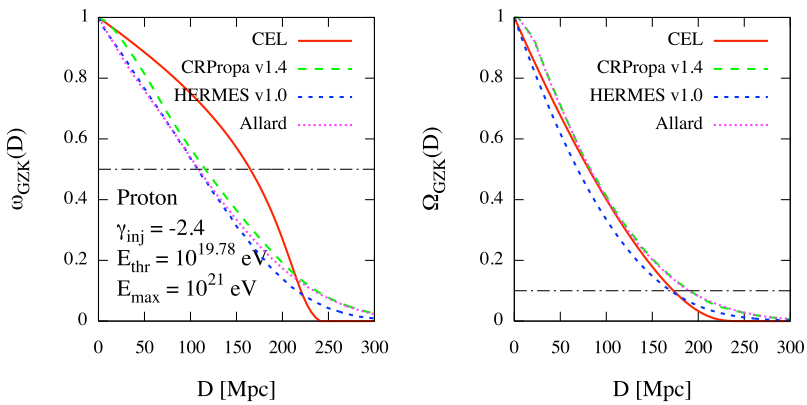


Figure 3.30: Surviving probability of protons, under the same conditions of Fig. 3.29. The particular case of $E_{\rm thr} = 60$ EeV ($10^{19.78}$ eV) is considered. The comparison between our simulator, other Monte Carlo propagators and the CEL approximation is shown for the surviving probability (left panel) and the surviving flux (right panel). See the text for further information.

It is evident that in such a second case, the surviving probability is overestimated up to ≈ 220 Mpc

and underestimated for higher distances. Such a behavior is due to the stochasticity of the photo-pion production, which is reflected in the probability of interaction between protons and photons of the EBR. The CEL approximates such an interaction as a continuous process, constraining the proton to interact at each propagation step. The result of such approximation is that particles produced by distant sources interact more than what is really expected, causing a steep decrease in the surviving probability. In the right panel of Fig. 3.30 we show the corresponding surviving flux, defined by Eq. (3.61): differences between propagators are reduced, and the GZK horizon estimated by each simulator, corresponding to the distance R such that $\Omega_{\rm GZK}(R) = 0.1$, is almost the same in each case.

The last check of consistency reported in this section deals with the estimation of the GZK horizon for both protons and iron nuclei, as a function of the energy threshold at Earth. For such a test, we consider three new propagators, recently published or ready for publication. In particular, we consider[5]:

- **PARSEC:** a fast Monte Carlo code simulating the 3D propagation of protons in a magnetized Universe; it makes use of parameterized probabilities and it is based on CEL approximation [210];
- **SimProp:** a fast Monte Carlo code simulating the 1D propagation of nuclei [211], where protons are propagated by making use of the CEL approximation;
- **CRPropa v2.0β:** a Monte Carlo code simulating the 3D propagation of nuclei in a magnetized Universe [212].

In Fig. 3.31 we show the GZK horizon estimated in the case of protons (left panel) and iron nuclei (right panel), as a function of the energy threshold at Earth. In the case of protons, propagators of the same nature provide similar results: in fact, the horizons obtained by HERMES are in good agreement with those of CRPropa v1.4 over the whole energy range under consideration, whereas PARSEC and SimProp, adopting a forward propagation under the CEL approximation, are in agreement between each other but underestimate the horizons. In the case of irons, the estimated horizons are much more similar, with an overall agreement among all simulators, but some differences at the lowest energy.

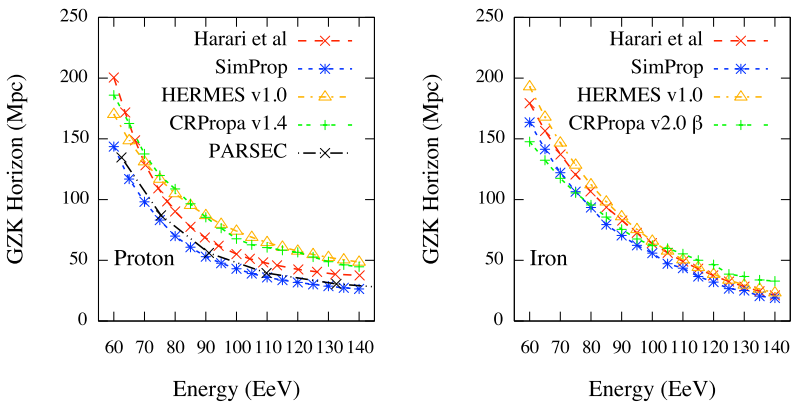


Figure 3.31: GZK horizon estimated in the case of protons (left panel) and iron nuclei (right panel), as a function of the energy threshold at Earth. Results from Harari et al [171] are shown for reference.

[5]We are in debt with: Dr. T. Winchen for having kindly provided the curve corresponding to PARSEC; the developers of SimProp for having kindly provided the simulated particles and the macros required to elaborate the output; Prof. G. Sigl for having recently allowed our access to the source code of CRPropa v2.0β before its official release.

3.8 Conclusions

In this chapter we have treated the energy-loss processes affecting the propagation of UHECRs in the Universe. We have presented the cosmological framework included in HERMES and the chosen parameterizations for the extragalactic background radiation, including cosmic microwave, infrared/optical and radio backgrounds. Successively, we have discussed the models and the parameterizations adopted to simulate the interaction between UHE nuclei and extragalactic background photons, treating with some detail the creation of secondary particles (including electron/positron pairs, photons and neutrinos) and their propagation. All the relevant energy-loss processes included in our simulator, as the adiabatic loss (due to the expansion of the Universe), the pair and photo-pion production, and, in the particular case of heavy nuclei, the photo-disintegration processes, have been discussed.

Hence, we have estimated interaction and energy-loss lengths for several nuclei, playing a fundamental role in simulations performed with HERMES, putting in evidence the agreement with known results. We have also shown the impact of energy-loss processes on the propagation of UHECRs and, in particular, on the GZK horizon of UHE protons. Finally, we have compared our results with those obtained by other simulators used in the community, completing the presentation of our original contribution to the simulation of the propagation of UHECRs. Results put in evidence the goodness of HERMES if compared to other propagators of the same nature, which employ a forward propagation with a full Monte Carlo approach.

What we observe is not nature itself, but nature exposed to our method of questioning.

W. Heisenberg

We are what we think. All that we are arises with our thoughts. With our thoughts, we make our world.

Siddharta Gautama

Chapter 4

Search for UHECR anisotropies: the multiscale method

In this chapter we will discuss the methods currently used for finding an anisotropy signal in the arrival direction distribution of UHECRs, as well as the methods adopted for estimating the correlation signal with a catalog of candidate sources. After a brief overview of hypothesis testing and of common methods designed for this purpose, we will present our original contribution, i.e. the introduction of a new multiscale method, designed to perform both catalog-independent and catalog-dependent analyses, based on information theory and extreme value statistics. In the following, we will show that the multiscale method is a competing tool for the study of both small and large scale anisotropies and correlations, providing a great discrimination power even in presence of a strong background contamination and for quite different astrophysical scenarios.

The part complementary to this chapter, regarding details about probability, statistics and information theory, can be found in Appendix B.

4.1 Hypothesis testing

A *null hypothesis* $\mathcal{H}_0$ is a statistical hypothesis that is tested for possible rejection as a true one, i.e. the observed effect is the result of chance. The hypothesis contrary to the null hypothesis, that observations are the result of a real effect, is known as the *alternative hypothesis* $\mathcal{H}_1$ [213]. *Statistical hypothesis testing* is a statistical procedure to establish the truth of a given hypothesis.

Fisher's approach for a statistical test requires the formulation of a null hypothesis $\mathcal{H}_0$, a test statistics $\mathcal{E}$, the computation of the so called *p-value* and the comparison with the (Fisherian) significance α, to make a decision based on the strength of the evidence.

Neyman-Pearson's approach requires the formulation of a null hypothesis $\mathcal{H}_0$ and of an alternative hypothesis $\mathcal{H}_1$, a test statistics $\mathcal{E}$, the computation of the so called *type I and type II errors* to make a decision between the two competing hypotheses.

Bayes' approach requires the formulation of a null hypothesis $\mathcal{H}_0$ and of an alternative hypothesis $\mathcal{H}_1$ (or more), the computation of the *prior* probability (often unknown, i.e. require assumptions) and of the *posterior* probability to make a decision between the two competing hypotheses.

For a statistical test of a null hypothesis $\mathcal{H}_0$ through a test statistics $\mathcal{E}$, the Fisher's *p-value* of $\mathcal{E}_0$ is defined as the chance probability to observe a value of $\mathcal{E}$ at least equal to $\mathcal{E}_0$, assuming that $\mathcal{H}_0$ is true: $P(\mathcal{E}_0) = p(\mathcal{E} \geq \mathcal{E}_0|\mathcal{H}_0)$. In Fig. 4.1 we show the meaning of the p-value for one-tail (left panel) and two-tail (right panel) tests. If α is the required (Fisherian) significance for the test, the null hypothesis is rejected if $P(\mathcal{E}_0) \leq \alpha$ and accepted otherwise.

In the non-Fisherian approaches, hypothesis testing can not be performed by using a single number, namely the p-value, because it only quantifies the amount of our "surprise" of how unusual are the data when compared to the null models. In principle, p-values should be not able to capture both the long-run outcomes of an experiment and the evidencial meaning of a single result.

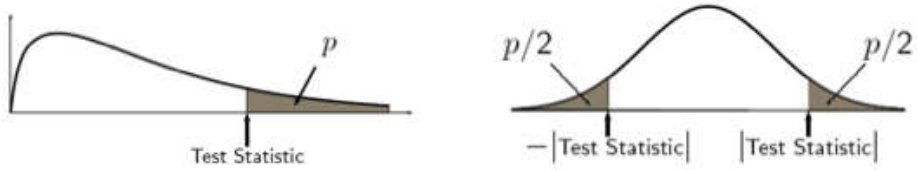


Figure 4.1: P-value for one-tail (left) and two-tail (right) tests.

The best way to estimate p-values is by mean of Montecarlo simulations. Let $\{x_n\}$ be a data sample of size n: in practical applications a large number of random realizations $\{X_n\}$, satisfying the null hypothesis $\mathcal{H}_0$, is used to estimate p-values for data. For instance, if N is the total number of Montecarlo simulations, the p-value of an autocorrelation function[1] (ACF) $w(\Theta)$ at some angular scale Θ_0 is given by

$$P\left[w(\Theta_0)\right] = \frac{\#\left[w^{\mathrm{MC}}(\Theta_0) \geq w(\Theta_0)\right]}{N},$$

i.e. the fraction of simulations with an ACF greater or equal than that of the data at the angular scale Θ_0. However, in practical analyses there is often an additional lack of *a priori* information about the scale of autocorrelation of a data set, i.e. the value of Θ_0 is not known before performing the statistical test (introducing further ignorance). In such cases, we are interested in the angular scale $\Theta^\star$ that minimizes the chance probability. The price to pay for such *a priori* ignorance is an additional scan over the whole parameter space $\mathbb{S}$, the angular scale Θ in this example. Such an approach defines a *penalization procedure* due to the parameter Θ.

If $P^* = \min\{P\left[w(\Theta)\right]\}$ is the minimum p-value of the ACF, obtained at the scale $\Theta^\star$ for the data from the exploratory scan over all values of Θ, and $\mathcal{P}(\Theta) = P\left[w^{(\mathrm{MC})}(\Theta)\right]$ is the p-value to have a null model with ACF $w^{(\mathrm{MC})}(\Theta)$ at the scale $\Theta \in [0, \pi]$, the penalized chance probability will be given by

$$P_{\mathrm{pen}}\left[w(\Theta^\star)\right] = P\left(\min\{\mathcal{P}(\Theta)\} \leq P^* | \mathcal{H}_0\right),$$

i.e. the fraction of Montecarlo simulations with a chance probability smaller (higher ACF) than that obtained from the data after a scanning over all possible angular scales. The generalization to the case of a parameter space with higher dimensionality is straightforward.

For any test statistics $\mathcal{E}$ with continuous density $f(\mathcal{E})$, the distribution of p-values (or penalized p-values) should be flat if the null hypothesis $\mathcal{H}_0$ is true[2]. Intuitively, such a theorem states that if the null hypothesis is true, any realization of the data sample is equally likely and, as a consequence, any value of the test statistics should occur with the same probability.

Although penalized p-values can be small, they do not still represent a significance. In fact, p-values do not require an alternative hypothesis in order to be estimated: they are an intrinsic feature of data

[1]The definition of such a function will be given in the next section. For the moment, it can be assumed as a statistical estimator depending on an angular variable Θ.

[2]Technically, an equivalent definition of this theorem states that the p-value is a random variable itself, and it is distributed as $\mathcal{U}(0, 1)$, i.e. the uniform distribution in the interval $[0, 1]$, if the null hypothesis $\mathcal{H}_0$ is true. We refer to [214] for the proof.

[215] and can not be used in hypothesis testing if an *a priori* (Fisherian) significance has not been fixed. Because of the absence of an alternative hypothesis, p-values are *biased* against the null hypothesis.

For this reason two new quantities, the type I and type II errors, have been introduced in hypothesis testing and decision making processes: they are not features of data but values defined *before* starting the test between a null hypothesis $\mathcal{H}_0$ and an explicit alternative hypothesis $\mathcal{H}_1$ [216]. The *Neyman-Pearsons's type I error* α is the probability to reject $\mathcal{H}_0$ when $\mathcal{H}_0$ is in fact true, while the *Neyman-Pearsons's type II error* β is the probability to accept $\mathcal{H}_0$ when $\mathcal{H}_1$ is in fact true:

$$\begin{aligned} \alpha &= \Pr\left(\text{reject } \mathcal{H}_0 | \mathcal{H}_0\right) \\ \beta &= \Pr\left(\text{reject } \mathcal{H}_1 | \mathcal{H}_1\right) \end{aligned}$$

The *significance* (or *size*) of the statistical test is given by α, which defines the confidence level $\mathrm{CL} = 1 - \alpha$ of the test. The *power* of the test statistics $\mathcal{E}$ used for the statistical test is defined as $1 - \beta$ and it depends, in general, on the sample size. The confidence level is thus the probability to correctly accept the null hypothesis when it is true, while the power is the probability to correctly accept the alternative hypothesis when it is true by mean of the test statistics $\mathcal{E}$. It follows that different estimators can lead to different powers (see Fig. 4.2).

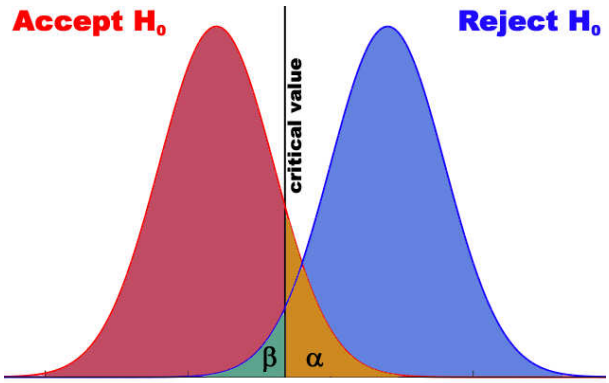


Figure 4.2: Type I and Type II errors.

It is important to point out that a p-values and Fisherian significance are not type I errors: the former estimates how *unlikely* is the value of the test statistics if the null hypothesis is known to be true, while the latter is the frequency of wrong rejection of $\mathcal{H}_0$. Thus, for instance, $P = 0.01$ does not correspond to $\alpha = 1\%$. We refer to Tab. 4.1 [217] and to [218–220] (and ref. therein) for further information.

Note. Although p-values are not equivalent to significances, it is still possible to obtain a significance from a p-value. Sellke *et al.* [221] presented a rigorous estimation of the lower limit on significance (type I error) from a p-value, which does not require assumptions on the distribution of the p-value and an explicit alternative hypothesis (see Appendix B).

For the sake of completeness, an alternative to classical hypothesis testing described so far, is the Bayesian decision making process, based on Bayes factors (see Appendix B for a brief treatment of this argument).

P-value	**α-level**
Fisherian significance level	Neyman-Pearson significance level
Significance test	Hypothesis test
Evidence against $\mathcal{H}_0$	Erroneous rejection of $\mathcal{H}_0$
Inductive: from particular to general	Deductive: from general to particular
Inductive inference: interpreting strength of evidence	Inductive: making decisions
Data-based random variable	Pre-assigned fixed value
Property of data	Property of test
Short run: applies to any single study	Long run: applies only to ongoing repetitions of original study
Hypothetical infinite population	Clearly defined population

Table 4.1: Main differences between p-value and $\alpha-$level approaches [217].

4.2 Review of anisotropy and correlation estimators

The measure of clustering is of fundamental importance in the physics of UHECRs, because it should shed light on the possibility of multi-messenger astronomy. In fact, because of Liouville's theorem, clustering can not occur because of magnetic fields if it is not an intrinsic feature of the data: thus, the presence of a clustering signal should be associated with an anisotropic distribution of either sources or their luminosity, responsible of event excesses in a privileged direction instead of another one. Moreover, when energy losses are taken into account during the propagation, as it should be, the distribution of sources with respect to redshift plays a significant role: the existence of the GZK effect should drastically reduce the number of candidate sources, by restricting their allowed positions to a sphere with radius of few hundreds Mpc. Observations suggest that the distribution of candidate sources in the nearby universe, e.g. AGN or rapidly rotating neutron stars, is strongly anisotropic. Hence, in the absence of clustering, sources are expected to be isotropically distributed and characterized by equally intrinsic luminosity.

Analytical methods based on probability distributions

Let us start by considering the limit of a continuous and uniform distribution of independent equal-luminosity sources in the sky, i.e. the case of an isotropic distribution. It is possible to measure the number of event pairs N_p, out of n, separated by an angle between Θ and $\Theta + d\Theta$, that under our assumptions is defined by

$$\frac{dN_p}{d\Theta} = \frac{2}{n(n-1)} \sin\Theta \int d\Omega_1 \int d\Omega_2 \omega(\delta_1)\omega(\delta_2)\delta\left(\hat{\Omega}_1 \cdot \hat{\Omega}_2 - \cos\Theta\right), \tag{4.1}$$

where $d\Omega_i = \cos\delta_i d\delta_i d\alpha_i$, δ_i and α_i denotes declinations and right ascensions, respectively, and δ_i spans over the declination observable by the experiment. If n is small, a typical situation in UHECR experiments, it is convenient to consider the integrated correlation function

$$N_p(<\Theta) = \int_0^\Theta \frac{dN_p}{d\Theta'} d\Theta', \tag{4.2}$$

to reduce the impact of intervening magnetic fields that may disperse the arrival directions, spreading the clustering signal over larger angular scales. For a surface detector like AGASA or the SD part of the Pierre Auger Observatory, the exposure function [222] is given by

$$\omega(\delta) \propto \cos\phi_0 \cos\delta \sin\alpha_m + \alpha_m \sin\phi_0 \sin\delta, \tag{4.3}$$

where ϕ_0 is the detector latitude (35.8° and −35.2°, for AGASA and Pierre Auger Observatory, respectively) and

$$\alpha_m = \begin{cases} 0 & \xi > 1 \\ \pi & \xi < -1 \\ \cos^{-1}\xi & \text{otherwise} \end{cases} \tag{4.4}$$

with

$$\xi \equiv \frac{\cos\theta_{\max} - \sin\phi_0 \sin\delta}{\cos\phi_0 \cos\delta},$$

being $\theta_{\max}$ the maximum zenith angle of the cosmic ray showers considered. Let us consider that the sky coverage consists of a solid angle Ω and let us divide it into M equal angular bins, each covering a solid angle $w \simeq \pi\Theta^2$ sr, being 2Θ the aperture of a cone covering the same solid angle. The number of bins is $M \simeq \Omega/(\pi\Theta^2)$. If we randomly toss n events into such bins, we have to consider the number m_0 of empty bins, m_1 of single hits, m_2 of double hits and so on. The probability to obtain a certain configuration is given by

$$P(\{m_i\}, n, M) = \frac{1}{M^n} \frac{M!}{m_0! m_1! m_2! ...} \frac{n!}{(0!)^{m_0}(1!)^{m_1}(2!)^{m_2}...}, \tag{4.5}$$

where $M!$ and $n!$ count the permutations of the bins and the events, respectively, $m_j!$ and $j!$ remove the overcounting of those bins containing j events, and the events within those bins, respectively. The normalization factor M^n is the total number of ways to partition n events among M bins [223]. However, it is worth noticing that variables involved in such a calculation are not independent:

$$\sum_{j=1} j m_j = n \quad \text{and} \quad \sum_{j=0} m_j = M. \tag{4.6}$$

By using such constraints, we obtain

$$P(\{m_i\}, n, M) = \frac{M!}{M^n} \frac{n!}{n^n} \prod_{j=0} \frac{(\overline{m}_j)^{m_j}}{m_j!}, \tag{4.7}$$

which is not a Poissonian distribution and where

$$\overline{m}_j = \frac{M}{j!} \left(\frac{n}{M}\right)^j. \tag{4.8}$$

The Poissonian distribution is retrieved in the limit $n << M$, where $\overline{m}_j$ approximates the mean number of $j-$plets. After taking into account the relationship between the number of bins and the angle, it is straightforward to obtain

$$\overline{m}_j(\Theta) = \frac{n}{j!} \left(\frac{\pi n \Theta^2}{\Omega}\right)^{j-1} \quad \text{and} \quad \frac{\overline{m}_j(\Theta)}{\overline{m}_{j-1}(\Theta)} = \frac{n}{jM} \sim \frac{\pi n \Theta^2}{j\Omega}. \tag{4.9}$$

If $p = 1/M$ is the probability of any event to fall into any of the independent bin, the average number of cluster $\overline{C}_m$ with a given multiplicity m, in a set of n events, is given by [224]

$$\overline{C}_m(n) = (1-p)\overline{C}_m(n-1) + p\overline{C}_m(n-1), \tag{4.10}$$

and thus they follow the binomial distribution

$$\overline{C}_m(n) = \frac{n!}{m!(n-m)!} p^{m-1}(1-p)^{n-m}. \tag{4.11}$$

In Fig. 4.3 we show some representative results. In the left panel we report the probability to obtain multiplets, estimated by the exact analytical model and the corresponding Poissonian approximation, as a function of the correlation angle Θ for the Pierre Auger Observatory. In the right panel is shown the average number of multiplets expected from the binomial model and obtained from simulations as a function of the number of events n in the data set. Simulations and approximate models only partially agree on the results. In both cases, any realization of n events adopted for simulations follows an isotropic distribution on the sky.

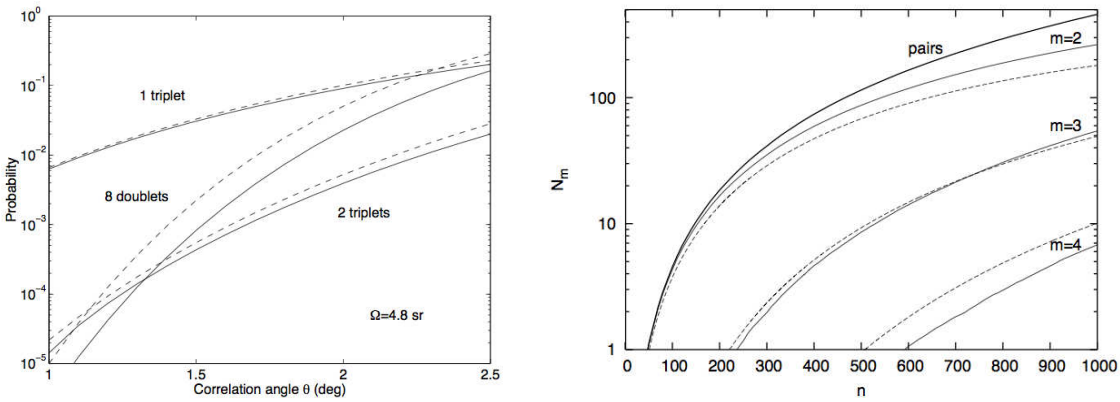


Figure 4.3: Isotropic distribution of events in the sky. **Left:** Inclusive probabilities for various clusters in a 100 event sample at Pierre Auger Observatory; exact analytical method (solid) and Poissonian approximation (dashed) are shown [223]. **Right:** Mean number of pairs (thick solid line), doublets, triplets and quadruplets (thin solid lines) in simulated realizations of n events randomly distributed with AGASA's declination-dependent exposure. The result of the binomial model for doublets ($m = 2$), triplets ($m = 3$) and quadruplets ($m = 4$) is plotted with dashed lines [224].

We can also be only interested in estimating the clustering due to a single source with flux Φ. If it is the case, the probability to observe a cluster of multiplicity m follows a Poissonian distribution

$$P_m(\Phi) = \frac{\overline{m}^m}{m!} e^{-\overline{m}}, \tag{4.12}$$

where $\overline{m} = \Phi\mathcal{E}$ is the mean multiplicity, with $\mathcal{E}$ the experimental exposure in the direction of the source. If dn_s/dF is the number density of sources providing a flux in the interval between Φ and $\Phi + d\Phi$, the average number of multiplets will be

$$\overline{C}_m = \int_0^\infty d\Phi \frac{dn_s}{d\Phi} P_m(\Phi), \tag{4.13}$$

whereas the probability of finding k clusters of multiplicity m will be

$$P_m(k) = \frac{(\overline{C}^m)^k}{k!} e^{-\overline{C}^m}. \tag{4.14}$$

We refer to Refs. [223–226] for further details.

Autocorrelation functions

Many estimators, mainly correlation functions [227–229], have been proposed and widely used to search for clustering of objects and to measure deviation of angular distributions from isotropy. These methods apply to angular coordinates of objects as well to distributions of arrival directions of events: in this

work we will indifferently refer to both as *arrival direction distributions of events*. In this section we will describe correlation functions based on Montecarlo simulations instead of analytical predictions.

If n is the number of experimental coordinates on some region $\mathcal{S}$ of a spherical surface and r is the number of coordinates coming from several Monte Carlo realizations of $\mathcal{S}$, the common anisotropy analysis involves the computation of different estimators: Peebles (or natural), Davis-Peebles, Landy-Szalay and Hamilton [227–229] angular correlation function, respectively defined by

$$\omega_1(\Theta) = \frac{r(r-1)}{n(n-1)}\frac{DD(\Theta)}{RR(\Theta)} - 1 \tag{4.15}$$

$$\omega_2(\Theta) = \frac{2r}{n-1}\frac{DD(\Theta)}{DR(\Theta)} - 1 \tag{4.16}$$

$$\omega_3(\Theta) = \frac{r(r-1)}{n(n-1)}\frac{DD(\Theta)}{RR(\Theta)} - \frac{r-1}{n}\frac{DR(\Theta)}{RR(\Theta)} + 1 \tag{4.17}$$

$$\omega_4(\Theta) = \frac{4nr}{(n-1)(r-1)}\frac{DD(\Theta)\times RR(\Theta)}{DR^2(\Theta)} - 1, \tag{4.18}$$

where DD is the number of pairs lying in the angular range between Θ and $\Theta+\Delta\Theta$ for the experimental distribution, RR is the same number calculated for Monte Carlo realizations and DR is the cross-pair counts between experimental and simulated events. By definition (see Ref. [230]) the function $\omega_i(\Theta)$ ($i = 1, 2, 3, 4$) is strictly related to the excess on the area $d\Omega$ of the pairs number $dN_{\rm pairs}$ over the random background $dN_{\rm MC}$:

$$dN_{\rm pairs} - dN_{\rm MC} = \frac{1}{2}nn_0\omega_i(\Theta)d\Omega,$$

where n_0 is the expected average density of events on $\mathcal{S}$ assuming an isotropic distribution. It has been shown [228] that $\omega_3(\Theta)$ is an unbiased estimator of the autocorrelation function, while $\omega_1(\Theta)$ and $\omega_2(\Theta)$ are not. An integrated variant of the two-point angular correlation function which reduces the impact of fluctuations, widely used when only small data sets of points are available [231–235], is defined by the number of pairs closer than an angle $\Theta \in [0, \pi]$:

$$w_{\rm int}(\Theta) = \sum_{i=1}^{n}\sum_{j<i} H(\Theta - \Theta_{ij}), \tag{4.19}$$

where H is the Heaviside step function and

$$\Theta_{ij} = \arccos\left(\cos\theta_i\cos\theta_j + \sin\theta_i\sin\theta_j\cos(\phi_i - \phi_j)\right)$$

is the angular distance between two directions i and j with coordinates (ϕ, θ) on the sphere. Such estimators are straightforwardly generalized to the estimation of the cross-correlation between two different data sets of objects.

For all of test statistics described so far, the parameter space $\mathbb{S}$ is spanned by the angle Θ, and coincides with the interval $\mathbb{D}_1 = [0, \pi]$. If another variable is taken into account, e.g. the energy E of each event corresponding to a pairwise coordinate, a new space $\mathbb{D}_2$ is introduced, and the parameter space of the ACF $w_{\rm int}(\Theta, E)$ will be defined by the bidimensional set $\mathbb{S} = [0, \pi] \otimes [0, +\infty[$.

Recently, new estimators have been introduced to study the anisotropy signature of sky's arrival direction distributions: the modified two-point Rayleigh (2pt+) [236], and shape-strength method derived from a principle component analysis of triplets of events (3pt) [237]. Such a test statistics have been recently applied to both Pierre Auger data and to synthetic maps of events, the latter generated by sampling the Veron-Cetty & Veron catalog [238] of nearby candidate active galactic nuclei (AGN) within 75 Mpc ($z \leq 0.018$), showing a higher discrimination power than other estimators [239].

2pt+ method

Two-point ACFs described so far are sensitive to the size of cluster but not to the relative orientation of the clustering points. The 2pt+ method [236] takes also into account any possible alignment of event pairs by using two additional variables related to the orientation of each vector connecting event pairs, as well as their angular distance. Let $\cos\beta$ the cosine of the vector's polar angle β, γ the vector's azimuthal angle and let us consider an angular binning of 5°. If $P(n_{j,k}^{\beta\gamma,\mathrm{obs}}, \mu)$ is the Poisson distribution with mean μ and $n_{j,k}^{\beta\gamma,\mathrm{obs}}$ is the observed number of pairs in the $j^{\mathrm{th}}k^{\mathrm{th}}$ bin, corresponding to $(\cos\beta, \gamma)$, the pseduo-log-likelihood

$$\mathcal{L}_{\beta,\gamma}^{\mathrm{data}} = \sum_{j,k} \ln P(n_{j,k}^{\beta\gamma,\mathrm{obs}}, \mu)$$

is built. Analogously, the pseduo-log-likelihood $\mathcal{L}_\alpha$ is built for the distribution of $\cos\alpha$, i.e. the angular distance between event pairs. The p-values p_α and $p_{\beta,\gamma}$ are obtained as the fraction of samples such that $\mathcal{L}_\alpha \leq \mathcal{L}_\alpha^{\mathrm{data}}$ and $\mathcal{L}_{\beta,\gamma} \leq \mathcal{L}_{\beta,\gamma}^{\mathrm{data}}$, respectively. Finally, both p-values are combined by mean of the Fisher's method to obtain

$$p^\dagger = p_\alpha p_{\beta,\gamma}(1 - \ln p_\alpha p_{\beta,\gamma}),$$

although further corrections are needed through Montecarlo simulations, because of the correlation between the two log-likelihood tests. For the definition of the log-likelihood and the Fisher's method, see Appendix B.

3pt method

Instead of pairs, the 3pt method considers triplets of events [237]. For each event, the arrival direction is converted into a Cartesian vector $\vec{r}$ and the orientation matrix is calculated. The corresponding three eigenvalues of the orientation matrix, subjected to the condition $\tau_1 + \tau_2 + \tau_3 = 1$, are estimated and ordered. Two new parameters are defined: 1) the strength parameter $\zeta = \ln(\tau_1/\tau_3)$ and 2) the shape parameter $\gamma = \ln[\ln(\tau_1/\tau_2)/\ln(\tau_2/\tau_3)]$. For each triplet a pair (ζ, γ) of shape-strength parameters is obtained and the result is successively binned.

The same procedure is repeated for a large number of Montecarlo realizations and the departure of the data from isotropy is estimated by building the pseudo-log-likelihood $\mathcal{L}_{\zeta,\gamma}$, where the Poisson distribution is adopted to estimate the probability to observe n_{obs} counts instead of the isotropic expectation n_{iso}, in each bin of (ζ, γ).

4.3 MAF: a new approach to anisotropy studies

Within the present work, we introduce a new fast and simple method for anisotropy analysis, which makes use of a multiscale approach and depends on one parameter only, namely the angular scale of the intrinsic anisotropy. The main advantage of our estimator is the possibility to analytically treat the results: the analytical approach drastically reduces computation time and makes available the possibility of applications to very large data sets of objects. We test the method on several simulated isotropic and anisotropic arrival direction distributions (mock maps) and perform an extensive analysis of its statistical features under both the null and the alternative hypotheses. However, it is worth remarking that the scope of applicability of our method is not limited to UHECR physics, and it is valid for any distribution of angular coordinates of objects.

Let $\mathcal{S}$ be a region of a spherical surface and let $P_i(\phi, \theta)$ $(i = 1, 2, ..., n)$ be a set of points locating n arrival directions on $\mathcal{S}$, defining a *sky*. The sky $\mathcal{S}$ is partitioned within a grid of N equal-area (and

almost-equal shape) disjoint boxes $\mathcal{B}_k$ ($k = 1, 2, ..., N$) as described in Ref. [240]. Let Ω be the solid angle covered by $\mathcal{S}$, whereas each box $\mathcal{B}_k$ covers the solid angle

$$\Omega_k = \frac{1}{N} \int_{\theta_{\min}}^{\theta_{\max}} \int_{\phi_{\min}}^{\phi_{\max}} d\cos\theta d\phi = 2\pi(1 - \cos\Theta)$$

where 2Θ is the apex angle of a cone covering the same solid angle: N, Θ and Ω are deeply related quantities that define a scale.

Let $\psi_k(\Theta)$ be the fraction of points in the data set falling into the box $\mathcal{B}_k$: the function $A(\Theta)$ that quantifies the deviation of data from an isotropic distribution at the scale Θ, is chosen to be the Kullback-Leibler divergence [241, 242]

$$A(\Theta) = \mathcal{D}_{\mathrm{KL}}\left(\psi(\Theta)||\overline{\psi}(\Theta)\right) = \sum_{k=1}^{N} \psi_k(\Theta) \log \frac{\psi_k(\Theta)}{\overline{\psi}_k(\Theta)} \tag{4.20}$$

where $\overline{\psi}_k(\Theta)$, generally a function of the domain meshing, is the expected fraction of points isotropically distributed on $\mathcal{S}$ falling into the box $\mathcal{B}_k$. The Kullback-Leibler divergence is an information theoretic measure widely used in hypothesis testing and model selection criteria [243–245], statistical mechanics [246–248], quantum mechanics [249–252], medical [253] and ecological [254] studies, to cite some of the most known. This measure quantifies the error in selecting the fraction $\overline{\psi}(\Theta)$ to approximate the fraction $\psi(\Theta)$ and it is strictly connected to maximum likelihood estimation (see Appendix B and notes at the end of this section). It is straightforward to show that $A(\Theta)$ is minimum for an isotropic distribution of points, or, in general, when $\psi(\Theta) \sim \overline{\psi}(\Theta)$, i.e. if the model is correct.

If $A_{\mathrm{data}}(\Theta)$ and $A_{\mathrm{iso}}(\Theta)$ refer, respectively, to the data and to an isotropic realization with the same number of events, we define *multiscale autocorrelation function* (MAF) the estimator

$$s(\Theta) = \frac{|A_{\mathrm{data}}(\Theta) - \langle A_{\mathrm{iso}}(\Theta)\rangle|}{\sigma_{A_{\mathrm{iso}}}(\Theta)} \tag{4.21}$$

where $\langle A_{\mathrm{iso}}(\Theta)\rangle$ and $\sigma_{A_{\mathrm{iso}}}(\Theta)$ are the sample mean and the sample standard deviation, respectively, estimated from several isotropic realizations of the data. If $\mathcal{H}_0$ denotes the null hypothesis of an underlying isotropic distribution for the data, the chance probability at the angular scale Θ, properly penalized because of the scan on Θ, is the probability

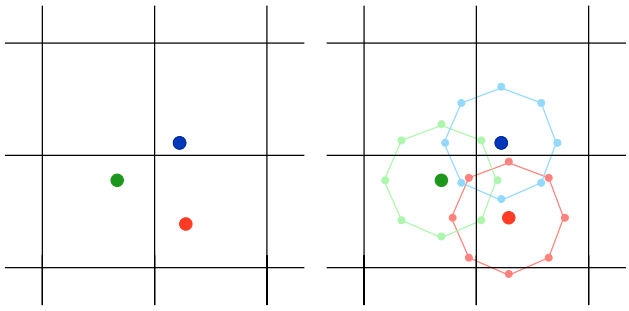
$$p(\Theta) = \Pr\left(s_{\mathrm{iso}}(\Theta') \geq s_{\mathrm{data}}(\Theta)|\mathcal{H}_0, \forall\Theta' \in \mathcal{P}\right) \tag{4.22}$$

obtained from the fraction of null models giving a multiscale autocorrelation, at any angular scale Θ' in the parameter space $\mathcal{P}$, greater or equal than that of data at the scale Θ. The null hypothesis is eventually rejected in favor of the alternative $\mathcal{H}_1 = \neg\mathcal{H}_0$ – being $\neg$ the negation operator – at the angular scale Θ, with probability $1 - p(\Theta)$.

Under the null hypothesis $\mathcal{H}_0$, the estimator $s(\Theta)$ follows a half-Gaussian distribution, independently on the value of the angular scale Θ and on the number of events on $\mathcal{S}$, as it will be successively shown in the text.

4.3.1 Dynamical counting

The simplest definition of the counting algorithm, as shortly described in the previous section, involves the fixed grid introduced in Ref. [240], where each box only embodies the relative number of events falling in it. Unfortunately, such a *static counting* approach could not reveal an existing cluster. For instance, Fig. 4.4a shows a typical scenario where some points of a given triplet fall into different cells.



(a) Static counting (b) Dynamical counting

Figure 4.4: A cluster of three points falling into different boxes.

Indeed, the fixed grid may cut a cluster of points within one or more edges, causing a further loss of information at the angular scale under investigation. To overcome this possible loss of information, we introduced a type of smoothing of the grid by applying it on the data.

The smoothing, adopted in our study, deals with a new counting procedure for the estimation of the density $\psi_k(\Theta)$. However, such a density depends on the observatory's exposure, generally a function on the celestial sphere depending on both the latitude of the experiment and the maximum zenith of detection, quantifying the effective time-integrated detection area for the flux of particles from each observable sky position. The relative exposure ω is the dimensionless function corresponding to the exposure normalized to its maximum value is given by Eq. 4.3.

Given an angular scale Θ, for each point P_i falling inside a box, we consider a set of 8 new points lying on a virtual box centered on P_i. Let $\alpha_{i,0}$ and $\delta_{i,0}$ be, respectively, the right ascension and the declination of the point P_i. We introduce the following notation:

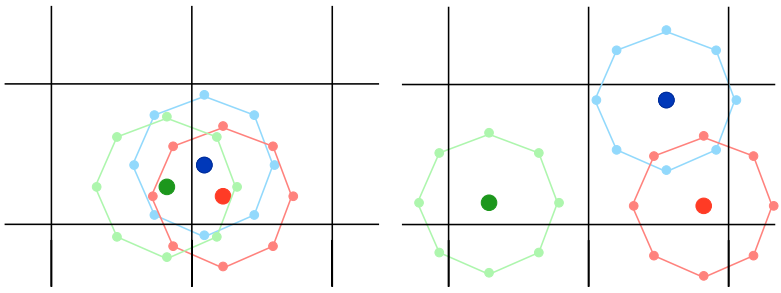
$$\begin{aligned}
\delta_{i,\pm 1} &= \delta_{i,0} \pm \frac{\Theta}{2} \\
\alpha_{i,\pm 1} &= \alpha_{i,0} \pm g\left(\delta_{i,0}\right) \\
\alpha_{i,\pm 2} &= \alpha_{i,0} \pm g\left(\delta_{i,+1}\right) \\
\alpha_{i,\pm 3} &= \alpha_{i,0} \pm g\left(\delta_{i,-1}\right)
\end{aligned}$$

where $g(\cdot)$ is a function which depends on declination, that constrains the angular distance between each of the 8 points and the original one to be $\frac{\Theta}{2}$. Within this framework, for each $P_i(\alpha_{i,0}, \delta_{i,0})$, we have the following 9 *extended* points:

- The original point $P_i(\alpha_{i,0}, \delta_{i,0})$;
- The up $P_i(\alpha_{i,0}, \delta_{i,+1})$ and down $P_i(\alpha_{i,0}, \delta_{i,-1})$ points;
- The left $P_i(\alpha_{i,-1}, \delta_{i,0})$ and right $P_i(\alpha_{i,1}, \delta_{i,0})$ points;
- The up-left $P_i(\alpha_{i,-2}, \delta_{i,+1})$ and up-right $P_i(\alpha_{i,2}, \delta_{i,+1})$ points;
- The down-left $P_i(\alpha_{i,-3}, \delta_{i,-1})$ and down-right $P_i(\alpha_{i,3}, \delta_{i,-1})$ points.

We define the function $h(\delta_{i,j}) = \omega(\delta_{i,j})/\omega(\delta_{i,0})$ $(j = 0, \pm 1)$ and introduce the weights

$$f(\delta_{i,j}) = \frac{h(\delta_{i,j})}{3h(\delta_{i,-1}) + 3 + 3h(\delta_{i,+1})} \tag{4.23}$$



(a) Clustered points (b) Unclustered points

Figure 4.5: a) Three clustered points: the *extended* points are mainly concentrated in two adjacent boxes. b) Three unclustered points: the *extended* points are mainly distributed on the neighbor cells.

In other words, we weight the angular region around a given direction with the local value of the exposure. Finally, we follow the procedure previously described by using the weighted distribution of points instead of the original one, as shown in Fig. 4.4b: thus, the density function $\psi(\Theta)$ is defined as the fraction of *extended* points, opportunely weighted the by function $f(\delta_{i,j})$ defined in Eq. (4.23). Our numerical studies show that such a *dynamical counting* approach recovers the correct information on the amount of clustering in the data.

In fact, the main difference between the static and the dynamical counting lies in the value of the estimator when the procedure is applied to Monte Carlo realizations of the sky. For instance, let us consider the Fig. 4.5, where we show a clustered (Fig. 4.5a) and an unclustered (Fig. 4.5b) set of points. The static counting is not able to recover the differences between the two configurations. Conversely, if the dynamical counting is applied, the *extended* points in Fig. 4.5a are concentrated in two adjacent boxes while in Fig. 4.5b they are distributed on the neighbor cells. This fundamental difference is reflected in the density function, leading to two different $\psi(\Theta)$. Monte Carlo skies producing the same clustered configuration shown in Fig. 4.5a, and as a consequence the same weight distribution, are not frequently expected: in this case, the value of $s(\Theta)$ should be greater than that one estimated from the static method. The direct consequence of a greater value of the estimator $s(\Theta)$ is a lower chance probability and the main advantage of using the dynamical counting, instead of the static one, should be the lowest penalization of $s(\Theta)$ only if an anisotropy signal is really present.

In order to illustrate the importance of dynamical counting in the anisotropy signal detection, we have generated 5000 isotropic and anisotropic skies of 100 events each. In each anisotropic sky, 60% of events are normally distributed, with dispersion ρ, around 10 random sources and 40% of events are isotropically distributed. For each angular scale Θ, we have estimated the average value of $s(\Theta)$ with the static and the dynamical counting, separately. Results are shown in Fig. 4.6 for $\rho = 5°$ (a), $\rho = 10°$ (b), $\rho = 20°$ (c) and for the isotropic map (d). As expected, the two counting methods do not show differences in the estimation of MAF in the case of isotropic skies, resulting in the same flat average value of $s(\Theta)$. Conversely, in the case of the anisotropic skies, the dynamical counting provides a greater estimation of $s(\Theta)$ than the static counting, leading to a smaller estimation of the corresponding chance probability and improving the signal-to-noise ratio. In the next section we will show how the dynamical counting is able to correctly recover the most significant clustering scale. For the sake of completeness, we have generated all mock maps with a full-sky coverage and a uniform exposure.

4.3.2 Interpretation of MAF

Any catalog-independent method provides information about the angular scale $\Theta^\star$ where the significance is minimum. In the case of a simple two-point method, such an angular scale is quite difficult

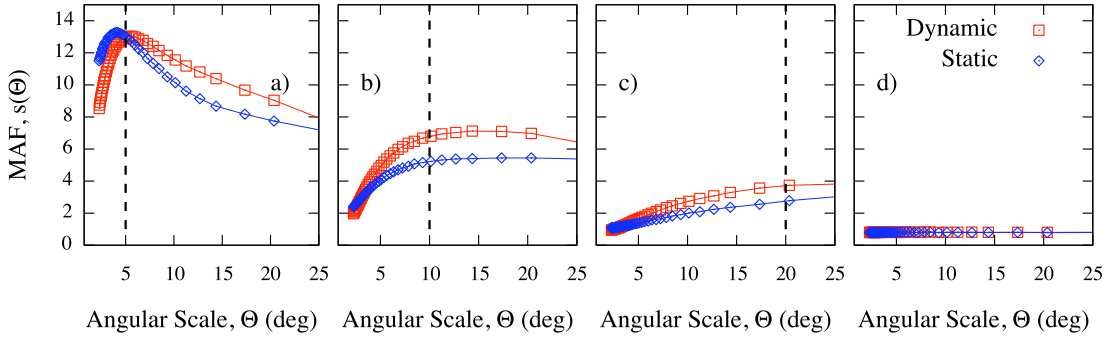


Figure 4.6: MAF: average $s(\Theta)$ (solid line) estimated from 5000 isotropic and anisotropic skies of 100 events each. In each anisotropic sky, 60% of events are normally distributed, with dispersion ρ, around 10 random sources and 40% of events are isotropically distributed. The dashed line indicates the value of the dispersion adopted to generate the corresponding mock map: a) 5, b) 10 and c) 20 degrees; d) isotropic map.

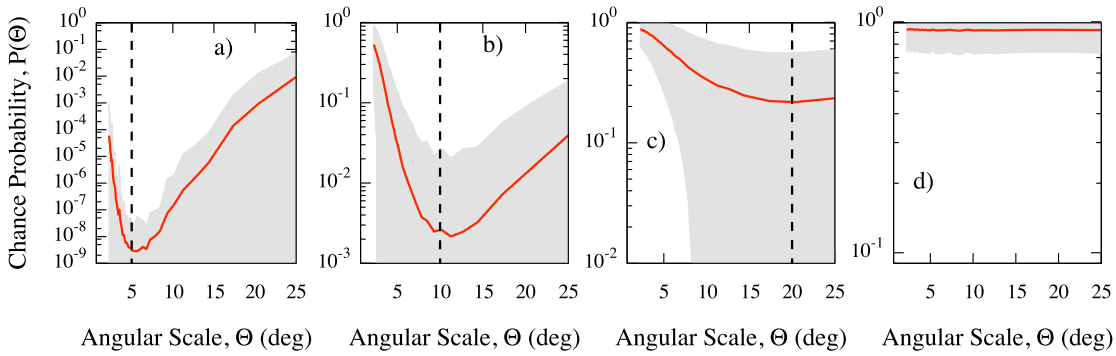


Figure 4.7: MAF: average chance probability (solid line), with 68% region around the mean value, estimated from isotropic and anisotropic skies generated as explained in Fig. 4.6. Dynamical counting is used. The dashed line indicates the value of the dispersion adopted to generate the corresponding mock map: a) 5, b) 10 and c) 20 degrees; d) isotropic map.

to interpret and topologically different configurations of events lead to the same significance. In the case of the modified two-point Rayleigh method, the estimation of the significance includes another set of parameters, independent from the angular distribution, as described in Ref. [236]: parameters are sensitive to the orientation of the pairs and therefore to skies showing preferential directions and filamentary structure of points. It follows that $\Theta^\star$ is the most significant angular size for the mix of these informations, still linked to the pair configuration. In the case of the shape-strength method, the estimation of two parameters, namely the shape and strengh, is performed: both can be interpreted, respectively, in terms of size and elongation of the triangles defined by a triplet of points. It follows that all information is recovered from the configuration of triplets.

In the specific case of MAF, the angular scale $\Theta^\star$, where the significance is minimum, turns to be the significative *clustering scale*: it is the scale at which occurs a greater accumulation of points respect to that one occuring by chance, with no regard for a particular configuration of points, e.g. doublets or triplets. To illustrate better the clustering scale detection feature of MAF, we have generated 5000 isotropic and anisotropic skies of 100 events each, as previously described at the end of Section 4.3.1. Fig. 4.7 shows the average chance probability, with 68% region around the mean value, versus the angular scale for three values of the dispersion, namely $\rho = 5°$ (a), $\rho = 10°$ (b), $\rho = 20°$ (c), and for the isotropic map (d). As expected, chance probability is close to one and nearly flat in the case of the isotropic map, because all clustering scales are equally likely. Conversely, for all anisotropic maps, the average chance probability gets a minimum around the corresponding value of ρ. Thus, our estimator is able to

recover the most significant clustering scale. It should be remarked that when the 20° dispersion is used, the angular scale of the minimum is less obvious because of the large fluctuations due to the isotropic contamination. Finally, it is worth noticing that we have observed that the curve around the value of ρ gets narrower by increasing the number of events.

4.3.3 Statistical analysis of MAF

In this section, we investigate the statistical features of MAF by inspecting its behavior under both the null or the alternative hypothesis. In particular, we estimate the significance α (or Type I error), i.e. the probability to wrongly reject the null hypothesis when it is actually true, and the power $1-\beta$ (where β is known as Type II error), i.e. the probability to accept the alternative hypothesis when it is in fact true. In the following we will adopt the dynamical counting previously discussed.

Null hypothesis. We generate isotropic maps of 10^5 skies, by varying the number of events from 20 to 500: for each sky in each map, we estimate the MAF for several values of the angular scale Θ. Hence, we choose the value of $\Theta = \Theta^\star$ where the chance probability is minimum, as the most significant clustering scale:

$$\tilde{p}(\Theta^\star) = \arg\min_{\Theta} p(\Theta)$$

properly penalized because of the scan on the parameter Θ, according to the definition in Eq. (4.22). Indipendently of the number of events in the mock map, we find an excellent flat distribution of probabilities $\tilde{p}(\Theta^\star)$, shown in Fig. 4.9a for skies of different size, as expected for analyses under the null hypothesis $\mathcal{H}_0$. In other words, MAF is not biased against $\mathcal{H}_0$, as required for good statistical estimators.

Because of the definition in Eq. (4.20) and of the central limit theorem, a Gaussian distribution is expected for the function $A(\Theta)$, and of consequence, the half-normal distribution

$$\mathcal{G}_{1/2}[s(\Theta)] = \frac{2}{\sqrt{2\pi}\sigma(\Theta)} e^{-\frac{s^2(\Theta)}{2\sigma^2(\Theta)}} \tag{4.24}$$

for $\sigma(\Theta) = 1$, is expected for the estimator $s(\Theta)$ defined as in Eq. (4.21), being normalized to zero mean and unitary variance. In Fig. 4.8 the distributions of the MAF estimator are shown for $n = 40$ and $n = 100$ events, for angular scales Θ ranging from 2° to 26°, separately. We find an excellent agreement between the distribution for Monte Carlo realizations and the expected one. It follows that the (unpenalized) probability to obtain by chance a value of the MAF, greater or equal than a given value s_0, is just $1 - \mathrm{erf}\left(\frac{s_0}{\sqrt{2}}\right)$, being erf the standard error function, independently of the angular scale Θ.

Although this is an important feature of the MAF estimator, generally the distribution of $s_{\max} = \max\{s(\Theta)\}$ is more useful for applications, because of the required penalization due to the scan over the parameter Θ. Hence, it is important to identify the distribution of the penalized probability $p(\Theta)$, if any. Intriguingly, our numerical studies show that such a distribution exists and it corresponds to one of the limiting densities in the extreme value theory (see notes at the end of this section). In particular, the cumulative density of maxima is known as the Gumbel distribution [255, 256]:

$$G(x) = \exp\left[-\exp\left(\frac{x-\mu}{\sigma}\right)\right]$$

where μ and σ are the location and shape parameters, respectively, and the corresponding probability density is

$$g(x) = \frac{1}{\sigma}\exp\left[-\frac{x-\mu}{\sigma} - \exp\left(\frac{x-\mu}{\sigma}\right)\right] \tag{4.25}$$

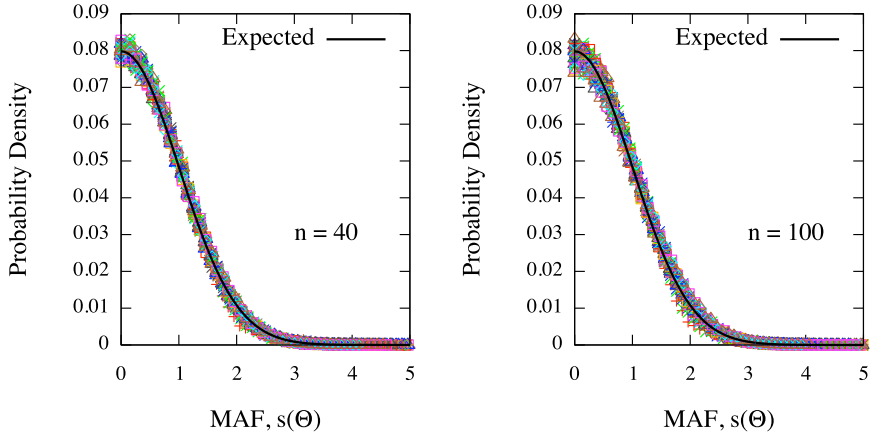


Figure 4.8: Distribution of the MAF estimator for $n = 40$ and $n = 100$ events, for angular scales Θ ranging from 2° to 26°, separately. Solids lines correspond to the expected half-normal distribution.

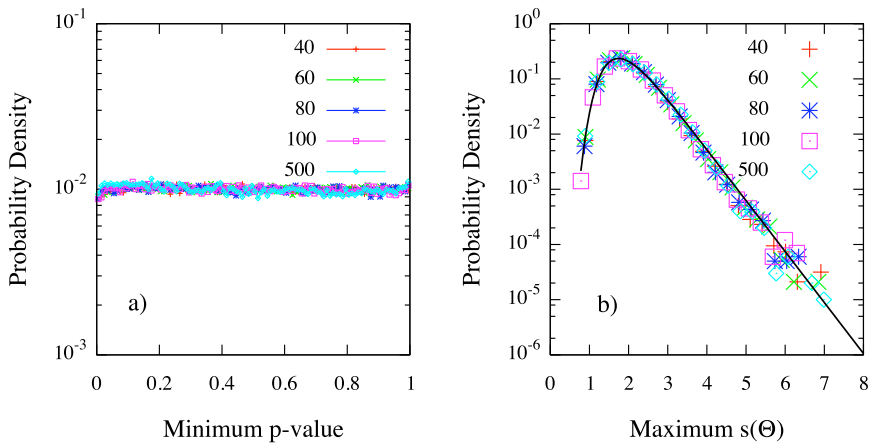


Figure 4.9: MAF. a) Distribution of $\tilde{p}(\Theta^\star)$ for $n = 40, 60, 80, 100$ and 500 events. b) Distribution of $\max\{s(\Theta)\}$ for $n = 40, 60, 80, 100$ and 500 events. Solid line correspond to the least-square fit of the Gumbel density with parameters $\mu = 1.743 \pm 0.002$ and $\sigma = 0.470 \pm 0.002$ ($\chi^2/\text{ndf} = 1.1 \times 10^{-5}$).

In Fig. 4.9b we show the probability densities of $s_{\max}$ for $n = 40, 60, 80, 100$ and 500 events: independently of n, each density is in excellent agreement with the Gumbel distribution of extreme values, for the parameters $\mu = 1.743 \pm 0.002$ and $\sigma = 0.470 \pm 0.002$. Such values correspond to the mean and to the standard deviation of the distribution, $\tilde{\mu} \approx 2.00$ and $\tilde{\sigma} \approx 0.59$, respectively (see notes at the end of this section). It follow that the probability to obtain a maximum value of $s(\Theta)$, at any angular scale Θ, greater or equal than a given value $\max\{s(\Theta)\}$ is

$$p\left(\max\{s(\Theta)\}\right) = 1 - \exp\left[-\exp\left(\frac{\max\{s(\Theta)\} - \mu}{\sigma}\right)\right],$$

providing an analytical expression for the penalized probability defined in Eq. (4.22).

Alternative hypothesis. In order to investigate the behavior of MAF under the alternative hypothesis of an underlying anisotropic distribution of objects, we have generated anisotropic maps of 10^4 skies, by varying the number of events from 20 to 100. In general, the anisotropy of a sky depends on several factors: for instance, in the case of cosmic rays, it depends on the distribution of sources, on magnetic fields and on propagation effects as energy loss or the GZK cutoff [40, 257] (and Ref. therein). Thus,

Experiment	ϕ_0	$\theta_{\max}$	**Exp.** (m^2 s sr)	λ	#**Ev.**
Volcano R.	35.15°N	70°	0.2×10^{16}	1.000	6
Yakutsk	61.60°N	60°	1.8×10^{16}	0.625	20
H. Park	53.97°N	74°	–	1.000	7
AGASA	35.78°N	45°	4.0×10^{16}	0.750	29
SUGAR	30.43°S	70°	5.3×10^{16}	0.500	13
Pierre Auger	35.20°S	60°	28.4×10^{16}	1.200	27

Table 4.2: Surface detectors: positions, maximum zenith angles $\theta_{\max}$, exposures, energy shift factors and number of detected events with rescaled energy $E' \geq 4.0 \times 10^{19}$ EeV.

a more complicated approach is required for the Monte Carlo realization of the maps. In order to estimate the power of MAF, we build reasonable anisotropic maps reflecting in part the real-world scenario, keeping in mind that our purpose is to build an anisotropic set of events for statistical analysis and not to generate events mimicking real data sets with the best available approximation. We proceed as follows:

1. *Catalog of candidate sources.* Although several models for production mechanisms of UHECR are available [40, 257] (and Ref. therein), [41, 258–264], it is generally accepted that the candidate sources are extragalactic and trace the distribution of luminous matter on large scales [225]. In particular, it has been shown that correlation with possible high redshift sources is unlikely [265], whereas compact sources are favored [266, 267]: the recent result reported by the Pierre Auger Collaboration experimentally supports the latter claim, showing a high correlation between the observed data and the distribution of nearby active galactic nuclei (AGN) [88, 208]. For these reasons, we use the Palermo Swift-BAT hard X-ray catalogue of AGN with known redshift within 200 Mpc ($z \leq 0.047$) [268], as the reference catalog of candidate sources providing the most complete and uniform all-sky hard X-ray survey up to date.

2. *Source effects.* Events, from each source in the reference catalog, are generated by weighting for the source flux and for the expected geometrical flux attenuation. Hence, the number of events coming from a source is proportional to its flux and to the factor z^{-2}: because of the small scales and the high energy of cosmic rays involved in anisotropy studies ($E \geq 4.0 \times 10^{19}$ EeV), we assume a flat universe with zero cosmological constant ($\Omega = 1$, $\Lambda = 0$) and nonevolving source. Indeed, we naively take into account the possible deflections of the particles, due to the random component of the magnetic field, by producing arrival directions gaussianly-distributed with dispersion ρ around the source. It is worth remarking that such a dispersion is strictly related to both the injection energy and the mass of the particle, as well as other physical quantities [257].

3. *Background.* We take into account the possibility for a contaminating isotropic background of the anisotropy signal, by generating a number of events isotropically distributed, corresponding to a fraction f_{iso} of the whole data set.

4. *Detection effects.* As previously explained, the number of events detected by a single fully efficient and full-time operating surface detector, depends on its own relative exposure. In order to take into account such a detection effect, we generate the events according to the relative exposure of the single detector. Moreover, for each detector we generate the corresponding number of events reported in Table 4.2, in order to produce skies mimicking as much as possible real data actually available.

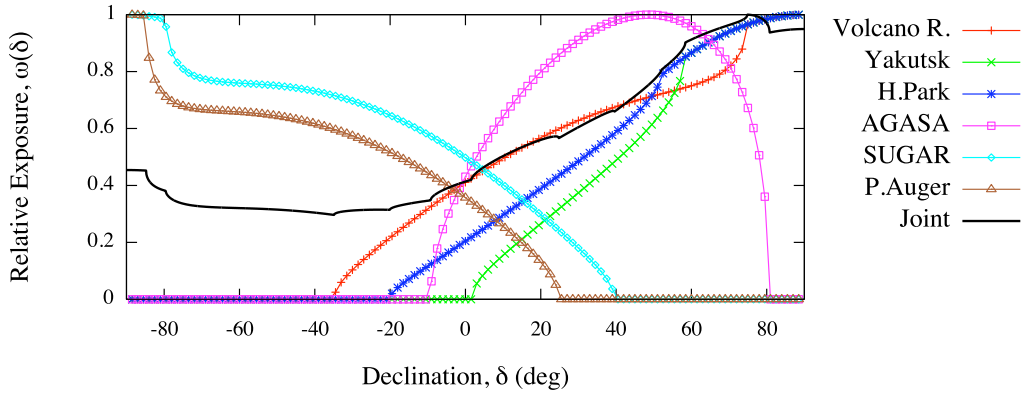


Figure 4.10: Relative geometrical exposure of each single detector listed in Table 4.2 (lines and points), and the joint exposure of all experiments (solid line).

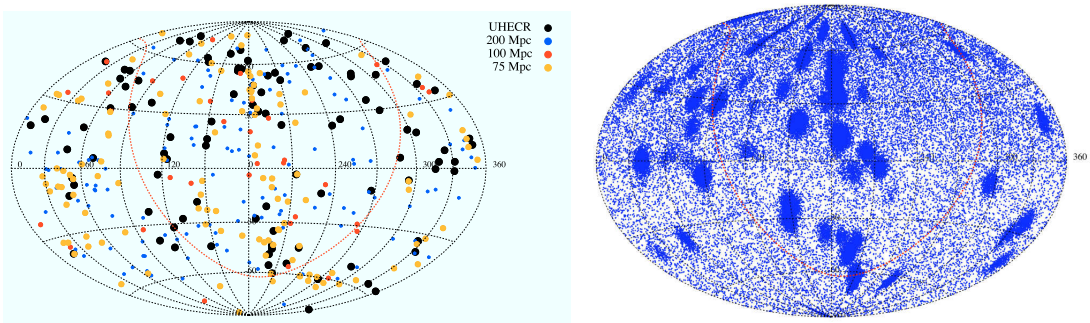
However, for a more realistic distribution of events, several more constraints, in general based on further assumptions or debated models, are required: the mass of the particle, the injection spectrum of the source, the intervening magnetic field, to cite some of the most important. In our study of the MAF discrimination power, we fix $\rho = 3°$, as the mean angular deviation of UHECR in the galactic and extra-galactic magnetic field, and a background fraction $f_{\rm iso} = 0.3$.

In order to produce a likely map of UHECR, we choose to generate events distributed in the whole sky, according to the number of events collected by surface detectors in the last decades. In particular, we consider events with energy $E \geq 4.0 \times 10^{19}$ EeV and error on the arrival direction smaller than 5°, as detected at the Sidney University Giant Airshower Recorder (SUGAR) [269], Akeno Giant Air Shower Array (AGASA) [270], Haverah Park [271], Volcano Ranch (one event from [257] and six events from [272]), Yakutsk [273], Pierre Auger Observatory [208] up to 2008. However, the fluxes of particles as measured by those experiments do not agree each other in the absolute fluxes, and a rescaling is needed [189]. By assuming that the spectrum reported by the HiRes Collaboration [274] corresponds to the correct energy scale, the rescaling, based on some specific characteristics of the UHECR spectrum, fixes the energy shift factors λ for the other experiments [189, 232]. Positions, maximum zenith angles $\theta_{\rm max}$, exposures and energy shift factors are reported in Table 4.2, for each experiment, as well as the number of detected events with rescaled energy $E' \geq 4.0 \times 10^{19}$ EeV ($E' = \lambda E$). In Fig. 4.10 the relative geometrical exposure of each single detector listed in Table 4.2 is shown, as well as the joint exposure of all experiments. For reference, in Fig. 4.11a the all-sky data set of 102 detected events with rescaled energy E' is shown, superimposed on the distribution of AGN within 200 Mpc from the reference catalog, whereas in Fig. 4.11b we show the mock map of simulated events according to physical constraints previously described.

In Fig. 4.12 we show the power $1 - \beta$ vs. the number of events, generated as described above. A sky is labelled as *anisotropic* if, for a fixed value of the significance α, the penalized chance probability as defined in Eq. (4.22) is lesser or equal than α, i.e. if the condition

$$\tilde{p}(\Theta^\star) = \arg\min_{\Theta} p(\Theta) \leq \alpha$$

holds for some angular scale $\Theta^\star$. In Fig. 4.12 we show the power for two values of the significance threshold, namely $\alpha = 0.1\%$ and $\alpha = 1\%$, estimated through the analytical approach. For applications, a power of 90% is generally required: under this threshold the method could fail to detect an existing anisotropy signal. In the case of the MAF, and for the considered anisotropic mock map, the power increases with the number of events n and it is able to detect the anisotropic signal for $n \geq 60$, with significance $\alpha = 1\%$. However, by decreasing the significance for the statistical test, the power requires



(a) UHECR events and candidate sources. (b) Mock map.

Figure 4.11: a) All-sky data set of 102 detected events with rescaled energy $E' \geq 40$ EeV (see the text for further information) superimposed on the distribution of AGN with known redshift ($z < 0.047$) from the Palermo SWIFT-BAT hard X-ray catalogue; b) corresponding mock map generated for the statistical analysis (see details in the text). Equatorial coordinates are shown.

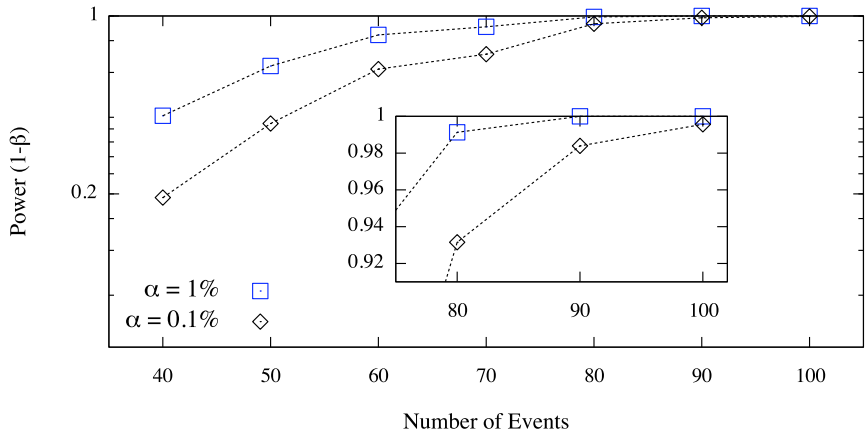


Figure 4.12: MAF power vs. the number of events sampled from anisotropic mock maps generated as described in the text, for values of the significance corresponding to $\alpha = 0.1\%$ and $\alpha = 1\%$.

a greater number of events to reach the 90% threshold, as expected: our test clearly shows that the MAF provides an excellent discrimination power for $n \geq 80$. Indeed, we verified the agreement between the analytical and the Monte Carlo estimations of the discrimination power.

4.3.4 Discussion and conclusion

We introduced a new statistical test, based on a multiscale approach, for detecting an anisotropy signal in the arrival direction distribution of UHECR, that makes use of an information theoretical measure of similarity, namely the Kullback-Leibler divergence, and of the extreme value theory. Within the present work we showed that our procedure is suitable for the analysis of both small and large data sets of events, by applying it on several Monte Carlo realizations of isotropic and anisotropic synthetic data sets, corresponding to plausible scenarios in the physics of highest energy cosmic rays. In fact, for small data sets as well as for larger ones, the method is able to recover the information about the most significant angular scale of clustering in the data, even in presence of strong isotropic contamination.

The advantages of our approach over other methods are multiple. First, the method allows an analytical description of quantities involved in the estimation of the amount of anisotropy signal in the data, avoiding thousands of Monte Carlo realizations needed for the penalizing procedure of results

and drastically reducing the computation time. Second, the method allows the detection of a physical observable, namely the clustering scale, in the case of a point source. In the case of multiple sources, the information is about the most significant clustering scale(s), according to source distribution. Third, the method is unbiased against the null hypothesis and it provides a high discrimination power even in presence of strong contaminating isotropic background, for both small and large data sets. Although in this work we referred to UHECR physics for our applications, it is worth remarking that the method is suitable for the detection of the anisotropy signal in each data set involving a distribution of angular coordinates on the sphere, and it can be adapted to non-spherical spaces by properly redefining the dynamical counting algorithm.

4.3.5 Additional information

The Kullback-Leibler divergence.

Let P and Q be two probability distributions, with densities $p(x)$ and $q(x)$, respectively. The Kullback-Leibler (KL) divergence is a measure quantifying the error in approximating the density $p(x)$ by means of $q(x)$, and it is defined [241, 242] as

$$\mathcal{D}_{KL}(p||q) = \int p(x) \log \frac{p(x)}{q(x)} dx \tag{4.26}$$

The KL divergence is non-negative, i.e. $\mathcal{D}_{KL}(p||q) \geq 0$ with equality if and only if $P = Q$, and asymmetric, i.e. $\mathcal{D}_{KL}(p||q) \neq \mathcal{D}_{KL}(q||p)$. The statistical interpretation of KL divergence is as follows.

Let $\tilde{P}$ the empirical distribution of random outcomes x_i $(i = 1, 2, ..., n)$ of the true distribution P, putting the probability $\frac{1}{n}$ on each outcome as

$$\tilde{p}(x) = \frac{1}{n} \sum_{i=1}^{n} \delta(x - x_i) \tag{4.27}$$

and let Q_Θ be the statistical model for the data, depending on the unknown parameter Θ. It follows

$$\mathcal{D}_{KL}(\tilde{p}||q_\Theta) = -\mathcal{H}(\tilde{p}) - \int \tilde{p}(x) \log q(x|\Theta) dx \tag{4.28}$$

where $\mathcal{H}(\tilde{p})$ is the information entropy of $\tilde{p}$, not depending on Θ, whereas $\tilde{p}$ and $q_\Theta = q(x|\Theta)$ are the corresponding densities of $\tilde{P}$ and Q_Θ, respectively. Putting Eq. (4.27) in the right-hand side of Eq. (4.28):

$$\begin{aligned} \mathcal{D}_{KL}(\tilde{p}||q_\Theta) &= -\mathcal{H}(\tilde{p}) - \frac{1}{n} \sum_{i=1}^{n} \log q(x_i|\Theta) \\ &= -\mathcal{H}(\tilde{p}) - \frac{1}{n} \mathcal{L}_q(\Theta|x) \end{aligned} \tag{4.29}$$

where $\mathcal{L}_q(\Theta|x)$ is the log-likelihood of the statistical model. It directly follows that

$$\arg\min_\Theta \mathcal{D}_{KL}(\tilde{p}||q_\Theta) = \frac{1}{n} \arg\max_\Theta \mathcal{L}_q(\Theta|x) \tag{4.30}$$

where the function $\arg\min(\arg\max) f(\Theta)$ retrieves the minimum (maximum) of the function $f(\Theta)$. Hence, another way to obtain the maximum likelihood estimation it to minimize the KL divergence

[275]; indeed, it can be shown that the KL divergence corresponds to the expected log-likelihood ratio [276].

Extreme value theory.

It is the research area dealing with the statistical analysis of the extremal values of a stochastic variable. Let x_i $(i = 1, 2, ..., n)$ be i.i.d. random outcomes of a distribution F. If $M_n = \max\{x_1, x_2, ..., x_n\}$, the probability to obtain an outcome greater or equal than M_n is:

$$\Pr(M_n \leq x) = \Pr(x_1 \leq x, x_2 \leq x, ..., x_n \leq x) = F^n(x)$$

It can be shown that the limiting distribution $F^n(x)$ is degenerate and should be normalized [277]. However, if there exists sequences of real constants $a_n > 0$ and b_n such that

$$\Pr\left(\frac{M_n - b_n}{a_n} \leq x\right) = F^n(a_n x + b_n)$$

then

$$\lim_{n \longrightarrow \infty} F^n(a_n x + b_n) = G(x) \tag{4.31}$$

The function $G(x)$ is the generalized extreme value (GEV) or Fisher-Tippett distribution

$$G(z) = \begin{cases} \exp\left(-e^{-z}\right) & \xi = 0 \\ \exp\left[-\left(1 - \xi z\right)^{\frac{1}{\xi}}\right] & \xi \neq 0 \end{cases}, \quad z = \frac{x - \mu}{\sigma} \tag{4.32}$$

defined for $1 - \xi z > 0$ if $\xi \neq 0$ and for $z \in \mathbb{R}$ if $\xi = 0$, where μ, σ and ξ are the location, scale and shape parameters, respectively. The Gumbel distribution is related to the distribution of maxima [255, 256] and it is retrieved for $\xi = 0$ [277]. The corresponding probability density $g(x)$ is easily obtained from G as

$$g(x) = \frac{1}{\sigma} \exp\left[-\frac{x - \mu}{\sigma} - \exp\left(\frac{x - \mu}{\sigma}\right)\right] \tag{4.33}$$

Finally, the two parameters μ and σ can be related to the mean $\tilde{\mu}$ and to the standard deviation $\tilde{\sigma}$ of the distribution, by means of the following relations:

$$\tilde{\mu} = \mu + \gamma\sigma \tag{4.34}$$

$$\tilde{\sigma}^2 = \frac{\pi^2}{6}\sigma^2 \tag{4.35}$$

where $\gamma = 0.577215...$ is the Euler constant.

4.4 Discrimination power of MAF in different astrophysical scenarios

A good test statistic should not be specialized on a particular type of signal, showing sensitivity to an as broad as possible set of anisotropies, even if specialized methods could reveal a weak signal that more general methods generally are not able to capture. Moreover, a good method should depend on a small number of external parameters and should be stable when changing internal parameters, if any.

In general, before applying a statistical method to a real data set, it should be tested on synthetic data sets mimicking the features that are supposed to happen in real data. In the following, we will name mock maps such synthetic data sets.

Within our study, we use the Multiscale Autocorrelation Function (MAF) [278] to investigate the clustering features of the arrival directions of simulated UHECR events and we present the results obtained from the statistical analysis of several mock maps, corresponding to different astrophysical models from which anisotropic distributions of events have been generated. In particular, we report the statistical power of MAF against the null hypothesis of an underlying isotropic distribution, for two values of the Type I error, $\alpha = 0.1\%$ and 1%, and by varying the number of events in the sky from 20 to 60. In order to be sure that a signal is still detected in case of an existing background, isotropic contamination has been included to these low statistics data sets.

A discussion on the significant clustering scales is presented, together with a brief comparison with known existing methods as the angular autocorrelation function (ACF), the 2-point correlation function and its enhanced version (2pt+) [236], the multipolar analysis [279] and the shape-strength method (3pt) [237].

4.4.1 Existing methods

Many methods have been proposed to search for an anisotropy signal in the arrival direction distribution of UHECRs, if any. Such methods can be grouped according to the "metric" and the statistical approach they involve.

There are two main categories: binned and unbinned methods. The fundamental difference is that binned methods depend on the bin size variable, whereas unbinned methods do not. The main advantage of unbinned tests is the gain in robustness of the final significance. Unbinned methods can be more sensitive to angular scales smaller than the uncertainty on the arrival directions of events ($\approx 0.9°$ above 50 EeV in the case of Pierre Auger Observatory), even if such scales are generally out of interest, because beyond the experimental capabilities. It follows that such methods lose statistical power at larger angular scales. Hence, unbinned tests are generally expected to be very specialized methods and they should not be suitable for anisotropy analysis. For this reason, in the following we will not consider unbinned methods.

Methods can be also grouped by the configuration of events they use to find the anisotropy signal. For instance, they could involve doublets or triplets, as the ACF or the 3pt methods, respectively. In this study, we consider the following methods:

- **ACF**: a one-parameter 2-point autocorrelation function;
- **UNM2pt**: a one-parameter binned 2-point correlation function;
- **MULTIPOL**: a one-parameter binned method involving the harmonic analysis of the arrival direction distribution;
- **2ptRayBin**: a one-parameter binned 2-point correlation function involving the relative orientation of the events on the sky (Rayleigh method);
- **UNMSs**: a two-parameter binned 3-point function involving principal component analysis;

where we have specified the number of parameters adopted by each method and their main features.

4.4.2 Astrophysical models

Within this study, we consider mock maps of events generated within the Pierre Auger Collaboration [280]. Such maps have been chosen according to the peculiar type of anisotropy they exhibit. In particular, we consider maps with large scale anisotropies (group A) as the following:

- **DIPOLEx**: dipole pointing in $+x$ direction;

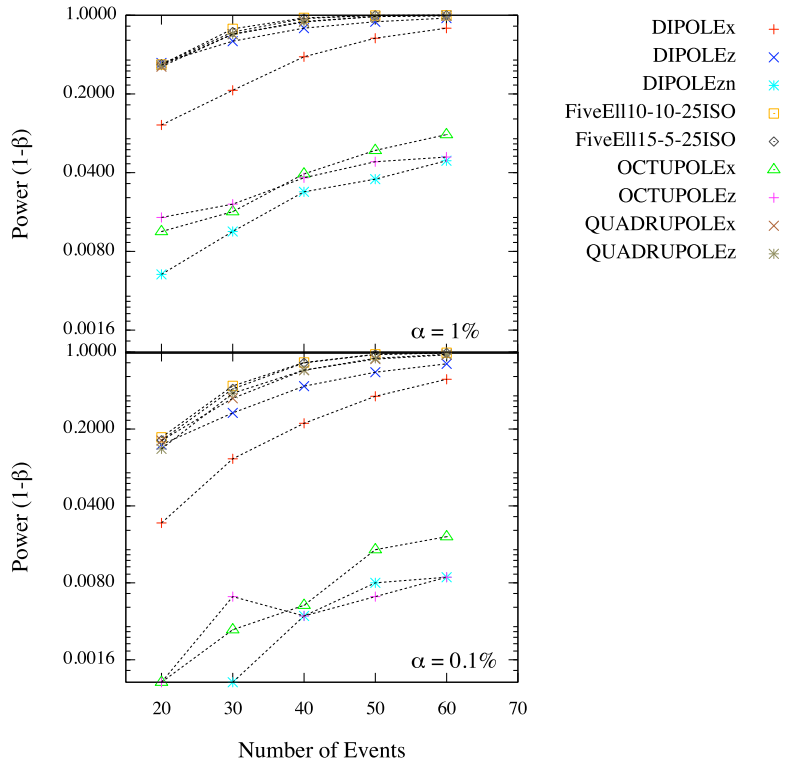


Figure 4.13: Power for two different value of α (0.1% and 1%) as a function of the number of the events, for the astrophysical models investigated in our study.

- **QUADRUPOLEx**: quadrupole pointing in $+x$ direction;
- **OCTUPOLEx, OCTUPOLEz**: octupole pointing in $+x$ and $+z$ direction, respectively;

where the label x (or z) indicates that the axis of symmetry x (or z) is along the equatorial plan (perpendicular to this plan). Then, we consider two mock maps where containing 35 cosmic rays within 5 ellipses (7 CR per ellipse) and 25 isotropically distributed events (group B):

- **FiveEll10-10-25ISO**: ellipses are randomly oriented circles of 10°;
- **FiveEll15-5-25ISO**: randomly oriented ellipses 15° long and 5° wide.

Then, we consider two different cases of the Infra Red Astronomical Satellite (IRAS) catalog (group C), where large scale anisotropy is expected to be dominant:

- **IRAS0020-10ISO**: 50 events are isotropically distributed and 10 events are sampled from astrophysical objects with $z \leq 0.02$;
- **IRAS0020-ISO**: 30 events are isotropically distributed and 30 events are sampled from astrophysical objects with $z \leq 0.02$;

In both cases, the distribution of events is weighted by $1/z^2$ and selection function, the IRAS mask is not taken into account and the direction of events is obtained by a Gaussian smearing around sources, with spread $\rho = 3°$.

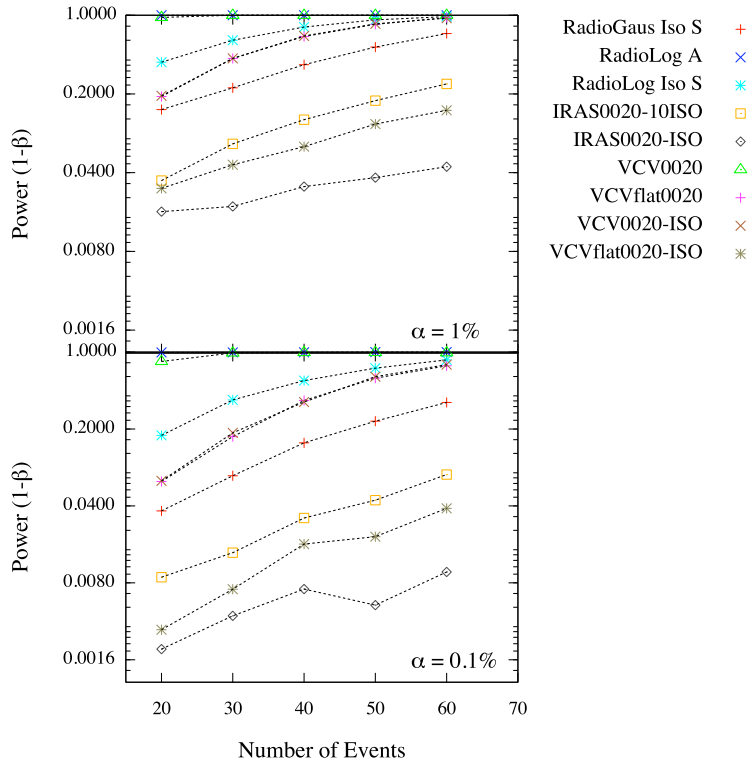


Figure 4.14: Power for two different value of α (0.1% and 1%) as a function of the number of the events, for the other astrophysical models investigated in our study.

This is followed by the Veron-Cetty & Veron (VCV) catalog (group D), where the density of sources is smaller than IRAS case, resulting in an increased number of repeaters and increasing the anisotropy signal at small angular scales:

- **VCVflat0020**: 60 events are sampled from astrophysical objects with $z \leq 0.02$. The distribution of events is not weighted by $1/z^2$;
- **VCVflat0020-ISO**: 30 events are isotropically distributed and 30 events are sampled from astrophysical objects with $z \leq 0.02$. The distribution of events is not weighted by $1/z^2$;
- **VCV0020**: 60 events are sampled from astrophysical objects with $z \leq 0.02$. The distribution of events is weighted by $1/z^2$;
- **VCV0020-ISO**: 30 events are isotropically distributed and 30 events are sampled from astrophysical objects with $z \leq 0.02$. The distribution of events is weighted by $1/z^2$.

In all cases, the direction of events is obtained by a Gaussian smearing around sources, with spread $\rho = 3°$. Finally, we consider mock maps with very few bright sources in the sky (group E), where an even stronger anisotropy signal than previous cases is expected at small angular scales:

- **RadioGaus_Iso_S**: 48 events are isotropically distributed and 12 are sampled from three sources, namely CenA (4.2 Mpc, $\rho = 8.2°$), VirA (16.1 Mpc, $\rho = 16.06°$) and ForA (19.6 Mpc, $\rho = 17.2°$);

Method/Mock	DIPOLEx	DIPOLEz	DIPOLEzn	QUADRUPOLEx	OCTUPOLEx	OCTUPOLEz	IRAS0020-10ISO	IRAS0020-ISO	VCVflat0020-ISO	VCVflat0020	VCV0020-ISO	VCV0020	RadioGaus-Iso-S	RadioLog-Iso-S	RadioLog-A	FiveEll10-10-25ISO	FiveEll15-5-25ISO
ACF	0.10	0.00	0.03	0.33	0.01	0.01	0.09	0.01	0.02	0.58	0.89	1.00	0.92	1.00	1.00	1.00	1.00
UNM2pt	0.20	0.92	0.10	0.98	0.00	0.00	0.01	0.00	0.02	0.47	0.20	1.00	0.15	0.83	1.00	0.92	0.90
MULTIPOL	0.02	0.02	0.01	0.74	0.20	0.03	0.04	0.01	0.08	0.48	0.13	0.94	0.05	0.09	0.87	0.18	0.19
2ptRayBin	0.53	0.83	0.01	0.97	0.14	0.02	0.07	0.01	0.04	0.82	0.85	1.00	0.60	0.98	1.00	1.00	1.00
UNMSss	0.11	0.81	0.02	1.00	0.01	0.01	0.07	0.01	0.08	0.87	0.77	1.00	0.73	0.98	1.00	1.00	1.00
MAF	0.57	0.78	0.01	0.94	0.02	0.01	0.09	0.01	0.04	0.75	0.77	0.99	0.35	0.85	1.00	0.99	0.98

Table 4.3: Fraction of realizations with 60 events providing a significance smaller than 0.1% for some existing methods and the MAF method. Blue text corresponds to the best method, red text to the worst method and green text to methods providing a fraction higher than 90%.

- **RadioLog_A**: 60 events are sampled from the sources and distributed as $dN/d\alpha = \alpha^{-1}$, where α is the separation angle in space;
- **RadioLog_Iso_S**: 18 events are sampled from the sources and distributed as $dN/d\alpha = \alpha^{-1}$, whereas 42 events are isotropically distributed.

It is worth remarking that events in each map are distributed by additionally taking into account the exposure of the Pierre Auger Observatory.

4.4.3 Statistical analysis

In order to estimate the minimum number of events on an anisotropic sky where the MAF method is able to perform a significant discrimination from isotropy, we investigate the statistical power of MAF against the null hypothesis of an underlying isotropic distribution. In this specific study, the power $1-\beta$ (where β is known as Type II error) is defined as the probability of correctly rejecting the null hypothesis when it is known to be false. For such a purpose, we use the MAF as test statistic and, for a given map, we count the fraction of skies providing a clustering signal with penalized p-value $p_{\min} \leq \alpha$, with $p_{\min} = p(s_{\max})$ and α is the significance or, equivalently, the Type I error, defined as the probability of wrongly rejecting the null hypothesis when it is known to be true. In any case, the power is expected to increase with the number of events.

The described procedure is applied to 10^4 realizations of mock maps described in Sec. 4.4.2 by varying the number of events in each sky from 20 to 60, and for two different values of the significance, namely $\alpha = 0.1\%$ and 1%. Results are shown all together in Figs. 4.13 and 4.14, to give a comprehensive idea of how the statistical power behaves as a function of the astrophysical model, the significance and the number of events in the sky.

Additionally, we show from Fig. 4.16 to Fig. 4.32 the power for each mock map separately (left panels) together with the distribution of the angular scales $\Theta^\star$ where the clustering signal is maximum, i.e. such that $s_{\max} = s(\Theta^\star)$, and the chance probability is smaller than 5%, i.e. where $p_{\min} \leq 0.05$ (right panels). A Gaussian kernel estimation has been used for histograms. In particular, we show results for models in:

- **group A**: Fig. 4.16, 4.17, 4.18, 4.19, 4.20 and 4.21;
- **group B**: Fig. 4.22 and 4.23;
- **group C**: Fig. 4.24 and 4.25;

- **group D**: Fig. 4.26, 4.27, 4.28 and 4.29;
- **group E**: Fig. 4.30, 4.31 and 4.32.

As expected, the power increases with the number of events in any case. A more detailed discussion about the power and the angular scales of clustering is provided in the next section.

4.4.4 Comparison with other methods

We report here the comparison between results obtained with the MAF method and methods briefly described in Sec. 4.4.1. In particular, we consider only the cases with skies of 60 events providing a significance smaller than 0.1%. The fraction of skies rejecting the null hypothesis of isotropy and satisfying such requirements are reported in Tab. 4.3 [281].

For astrophysical models in group A, the ACF generally behaves equal or worse than other methods, including MAF, whereas it is the best method for almost all astrophysical models where the anisotropy signal is stronger at the smallest angular scales. UNM2pt, MULTIPOL, 2ptRayBin (2pt+) and UNMSss (3pt) are more sensitive to some types of large scale anisotropy, while maintaining competitive performance even on small scale anisotropy. However, in general the MULTIPOL method is the least sensitive among them.

The MAF method shows a high sensitivity to both small and large scale anisotropies, together with a competitive statistical power comparable to that one of the other methods. Hence, it is suitable for the application on real data sets of events together with ACF, 2pt+ and 3pt methods.

4.4.5 Discussion and conclusion

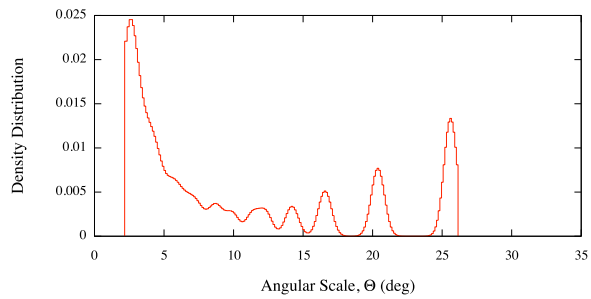


Figure 4.15: *Isotropic* distribution of 60 events. Probability density of angular scales such that the clustering signal gets a maximum corresponding to $p_{min} \leq 0.05$.

In the case of an isotropic distribution of events, the chance probability to obtain a given $\max\{s(\Theta)\}$, or, equivalently, a minimum $p-$value $p_{\min}$, is uniformly distributed in the interval $[0, 1]$. Morover, within such an interval, the distribution of the corresponding angular scales where $s(\Theta)$ gets a maximum, is flat because all clustering scales are equally likely, as shown in Ref. [278]. Conversely, if we focus only on the angular scales such that the clustering signal gets a maximum corresponding to $p_{min} \leq 0.05$, the corresponding probability density depends on the angular scale. In Fig. 4.15 such a distribution is shown in the case of skies of 60 events. It is evident that the number of clusters occurring by chance within $5°$ is larger than the number of clusters occurring by chance at larger angular scales. Such a behavior should be taken into account for the correct interpretation of the probability density of angular scales corresponding to the different kind of anisotropy.

In fact, in those astrophysical scenarios where the MAF provides a lower statistical power, the probability density of angular scales is clearly contaminated by angular scales that appear more frequently

in isotropic data sets with $p_{\min} \leq 0.05$. Conversely, in those models where the power is greater than 80%, the probability density is centered around the typical clustering scale of the anisotropy introduced in the mock skies.

In conclusion, we have shown that the MAF is a competing catalog independent method for capturing the anisotropy signal at both small scales and large scales. Moreover, from a comparison with the methods presented in Sec. 4.4.1, it provides the best power in the case of the DIPOLEx anisotropy, and the same power in the case of small scales anisotropies. Finally, it is worth remarking that the MAF method is able to recover the clustering scale in the data in those astrophysical scenarios where it provides a high statistical power.

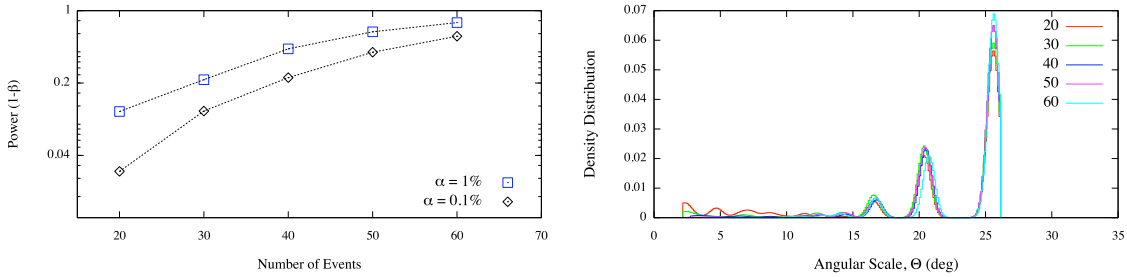


Figure 4.16: DIPOLEx. *Left panel:* Power for two different value of α (0.1% and 1%) as a function of the number of the events. *Right panel:* Probability density of angular scales such that the clustering signal gets a maximum corresponding to $p_{min} \leq 0.05$.

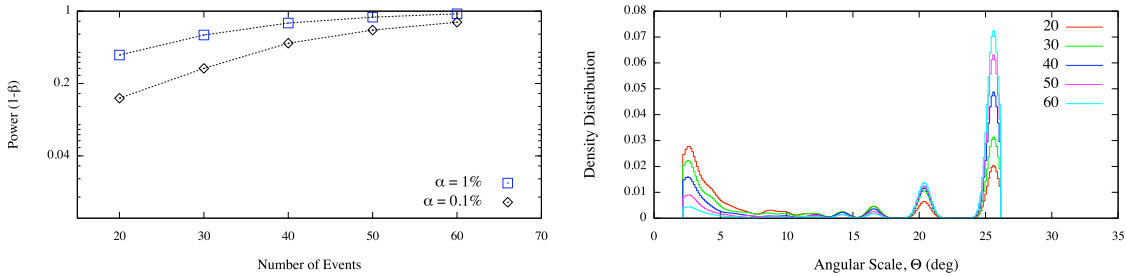


Figure 4.17: DIPOLEz. *Left panel:* Power for two different value of α (0.1% and 1%) as a function of the number of the events. *Right panel:* Probability density of angular scales such that the clustering signal gets a maximum corresponding to $p_{min} \leq 0.05$.

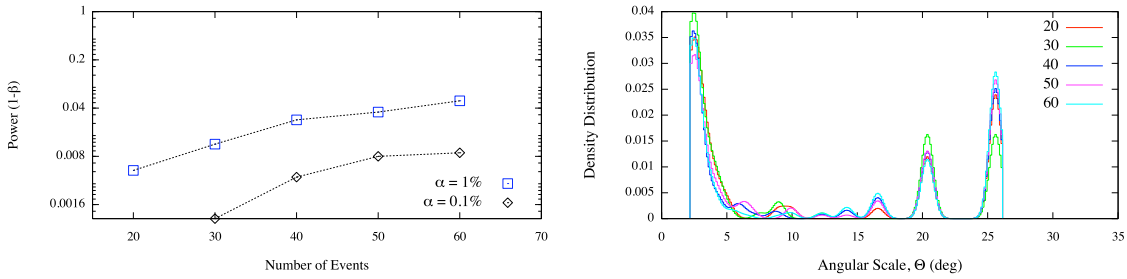


Figure 4.18: DIPOLEzn. *Left panel:* Power for two different value of α (0.1% and 1%) as a function of the number of the events. *Right panel:* Probability density of angular scales such that the clustering signal gets a maximum corresponding to $p_{min} \leq 0.05$.

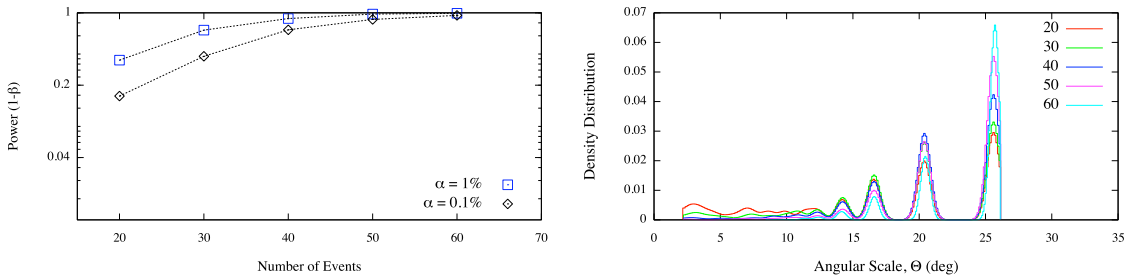


Figure 4.19: QUADRUPOLEx. *Left panel:* Power for two different value of α (0.1% and 1%) as a function of the number of the events. *Right panel:* Probability density of angular scales such that the clustering signal gets a maximum corresponding to $p_{min} \leq 0.05$.

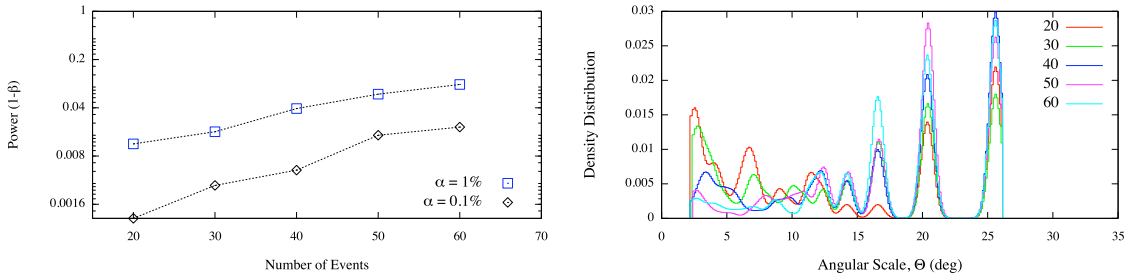


Figure 4.20: OCTUPOLEx. *Left panel:* Power for two different value of α (0.1% and 1%) as a function of the number of the events. *Right panel:* Probability density of angular scales such that the clustering signal gets a maximum corresponding to $p_{min} \leq 0.05$.

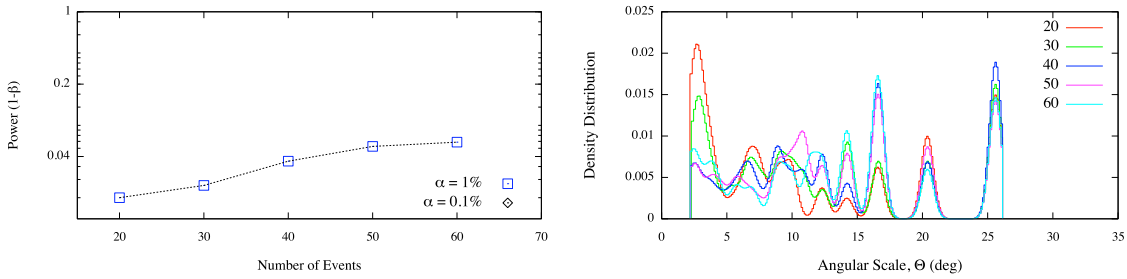


Figure 4.21: OCTUPOLEz. *Left panel:* Power for two different value of α (0.1% and 1%) as a function of the number of the events. *Right panel:* Probability density of angular scales such that the clustering signal gets a maximum corresponding to $p_{min} \leq 0.05$.

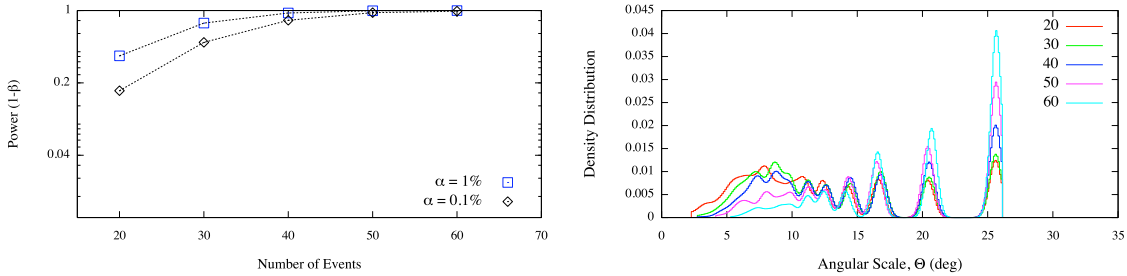


Figure 4.22: FiveEll10-10-25ISO. *Left panel:* Power for two different value of α (0.1% and 1%) as a function of the number of the events. *Right panel:* Probability density of angular scales such that the clustering signal gets a maximum corresponding to $p_{min} \leq 0.05$.

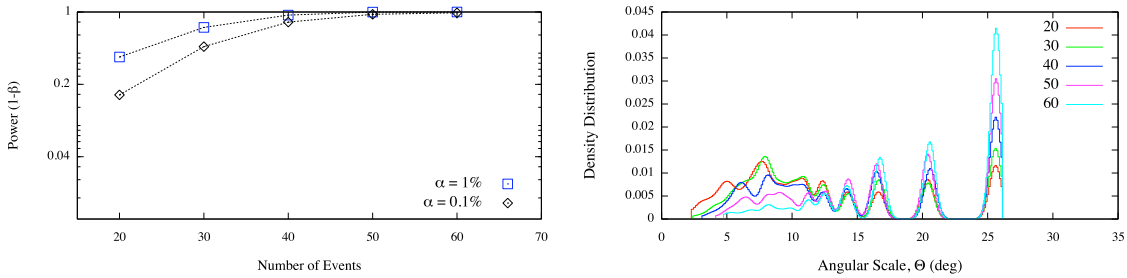


Figure 4.23: FiveEll15-5-25ISO. *Left panel:* Power for two different value of α (0.1% and 1%) as a function of the number of the events. *Right panel:* Probability density of angular scales such that the clustering signal gets a maximum corresponding to $p_{min} \leq 0.05$.

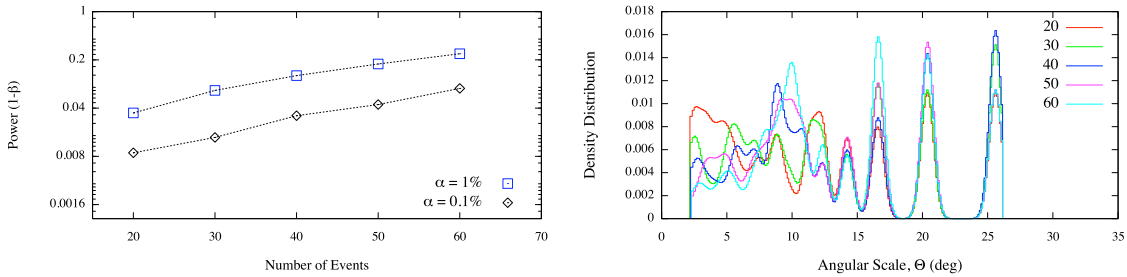


Figure 4.24: IRAS0020-10ISO. *Left panel:* Power for two different value of α (0.1% and 1%) as a function of the number of the events. *Right panel:* Probability density of angular scales such that the clustering signal gets a maximum corresponding to $p_{min} \leq 0.05$.

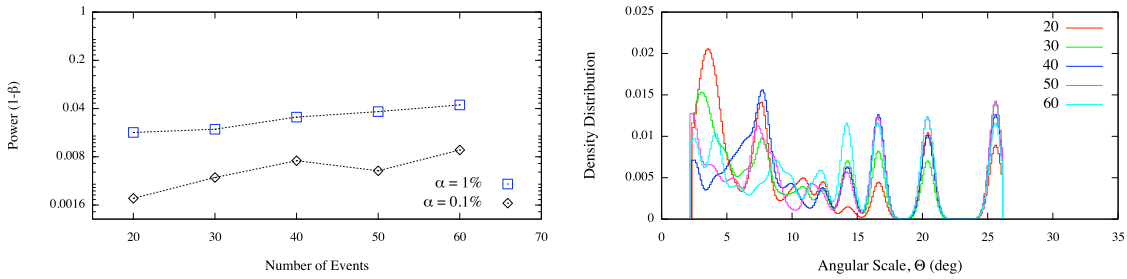


Figure 4.25: IRAS0020-ISO. *Left panel:* Power for two different value of α (0.1% and 1%) as a function of the number of the events. *Right panel:* Probability density of angular scales such that the clustering signal gets a maximum corresponding to $p_{min} \leq 0.05$.

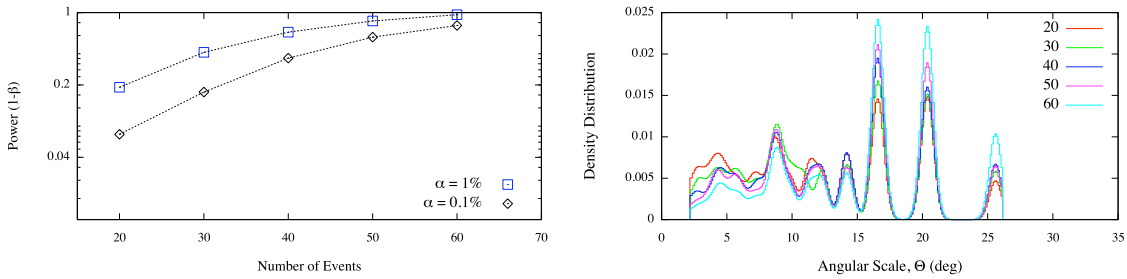


Figure 4.26: VCVflat0020. *Left panel:* Power for two different value of α (0.1% and 1%) as a function of the number of the events. *Right panel:* Probability density of angular scales such that the clustering signal gets a maximum corresponding to $p_{min} \leq 0.05$.

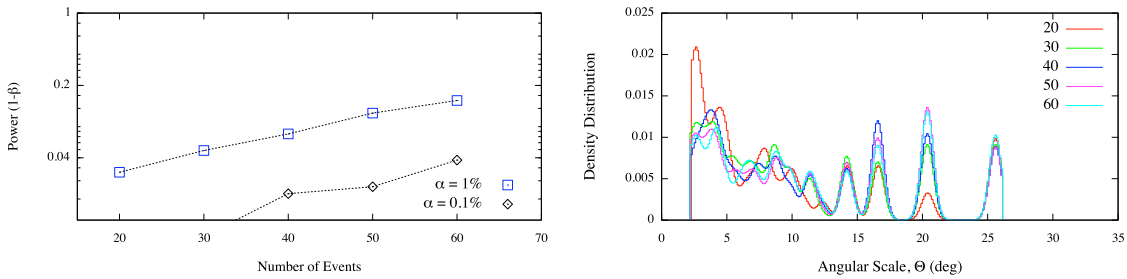


Figure 4.27: VCVflat0020-ISO. *Left panel:* Power for two different value of α (0.1% and 1%) as a function of the number of the events. *Right panel:* Probability density of angular scales such that the clustering signal gets a maximum corresponding to $p_{min} \leq 0.05$.

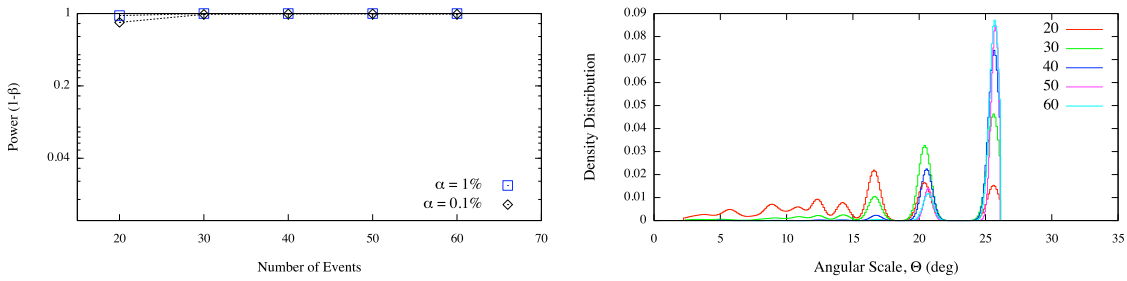


Figure 4.28: VCV0020. *Left panel:* Power for two different value of α (0.1% and 1%) as a function of the number of the events. *Right panel:* Probability density of angular scales such that the clustering signal gets a maximum corresponding to $p_{min} \leq 0.05$.

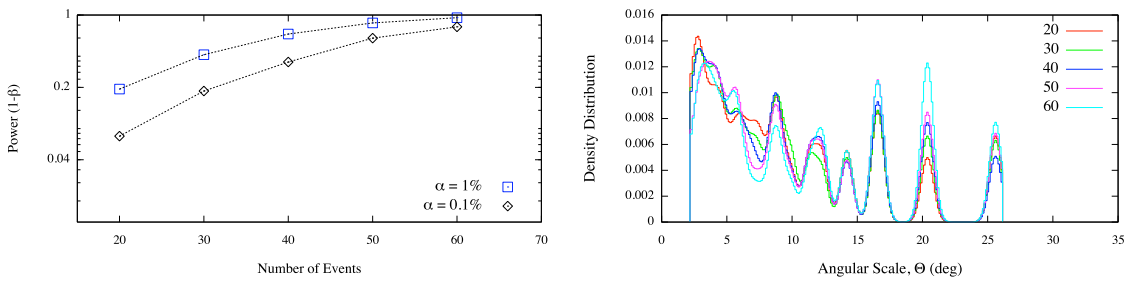


Figure 4.29: VCV0020-ISO. *Left panel:* Power for two different value of α (0.1% and 1%) as a function of the number of the events. *Right panel:* Probability density of angular scales such that the clustering signal gets a maximum corresponding to $p_{min} \leq 0.05$.

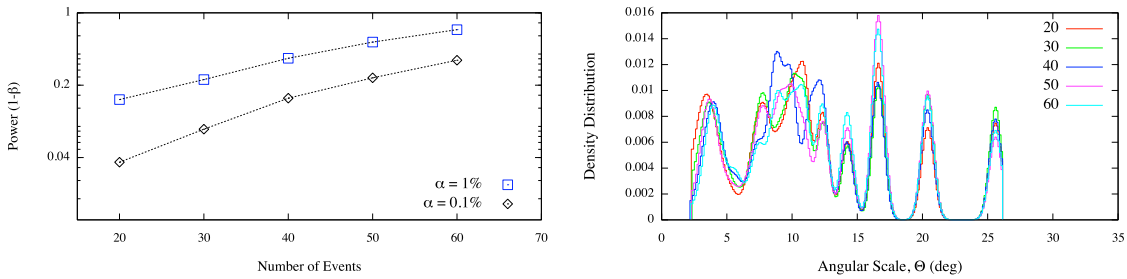


Figure 4.30: RadioGaus-Iso-S. *Left panel:* Power for two different value of α (0.1% and 1%) as a function of the number of the events. *Right panel:* Probability density of angular scales such that the clustering signal gets a maximum corresponding to $p_{min} \leq 0.05$.

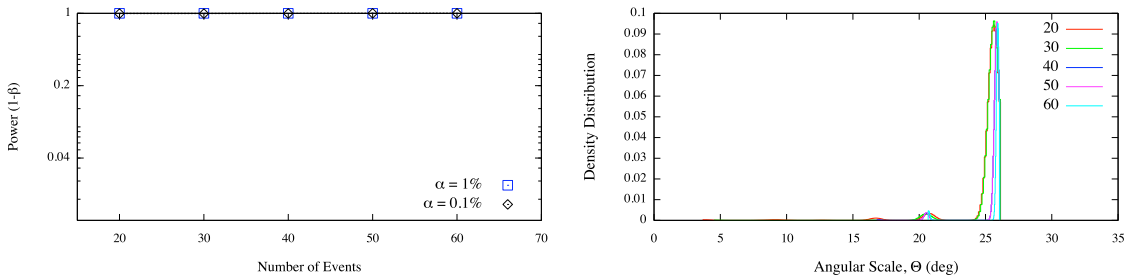


Figure 4.31: RadioLog-A. *Left panel:* Power for two different value of α (0.1% and 1%) as a function of the number of the events. *Right panel:* Probability density of angular scales such that the clustering signal gets a maximum corresponding to $p_{min} \leq 0.05$.

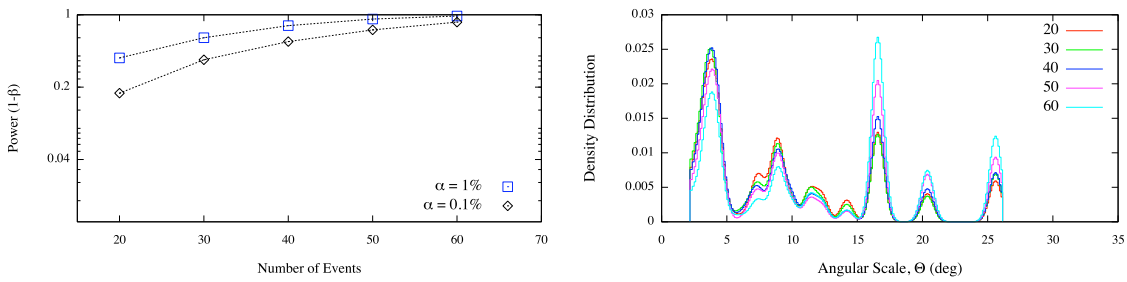


Figure 4.32: RadioLog-Iso-S. *Left panel:* Power for two different value of α (0.1% and 1%) as a function of the number of the events. *Right panel:* Probability density of angular scales such that the clustering signal gets a maximum corresponding to $p_{min} \leq 0.05$.

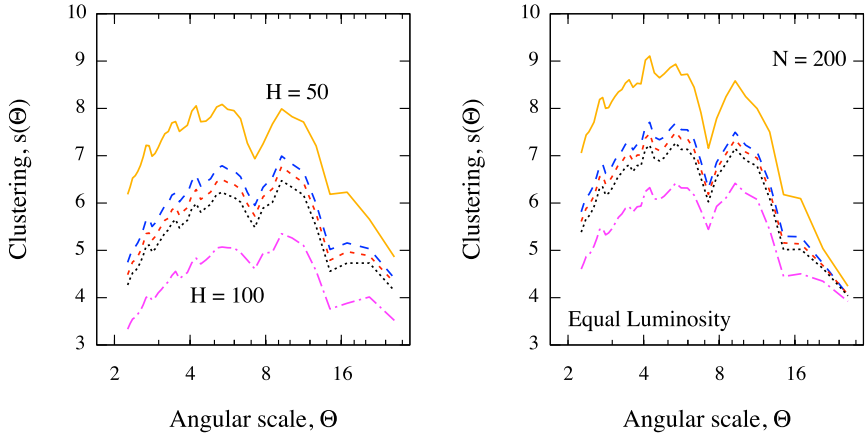


Figure 4.33: Expected clustering signal, as a function of the angular scale, from a sky of $N = 200$ protons with $E \geq 100$ EeV (in the field of view of Pierre Auger Observatory) and for values of the Hubble parameter considered in Fig. 3.26. Sources of 44% of events are AGN within 200 Mpc in the SWIFT-BAT 58-months catalog, whereas the remaining 56% of events are isotropically distributed. Scenarios with intrinsic luminosity taken into account (left panel) and not taken into account (right panel) are considered. The signal at each angular scale is obtained by averaging over 10^4 Monte Carlo realizations.

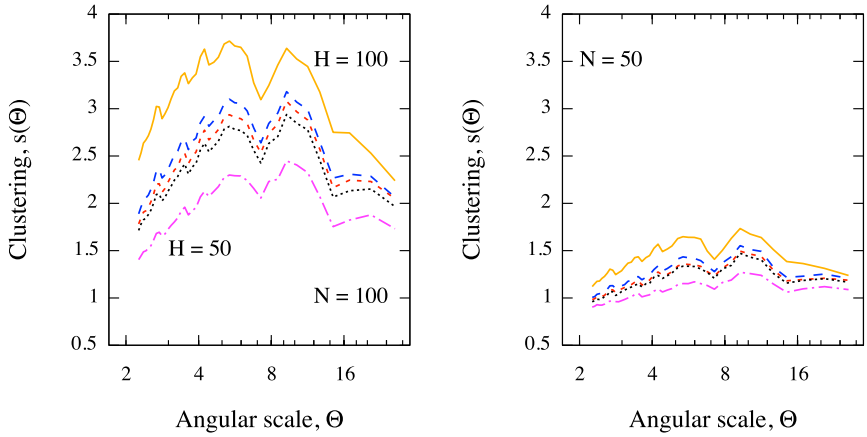


Figure 4.34: Same as in Fig. 4.33, for scenarios where intrinsic luminosity of AGN is taken into account. The clustering for $N = 100$ (left panel) and $N = 50$ (right panel) is considered.

4.5 Probing cosmological parameters with MAF

Motivated by the correlation between the arrival directions of UHECRs detected with the Pierre Auger Observatory and AGN [89], we consider the distribution of sources corresponding to the positions of AGN in the nearby universe (up to ≈200 Mpc), reported in the SWIFT-BAT 58-months catalog [282]. In particular, we consider two scenarios: i) all AGN are treated with equal intrinsic luminosity and ii) the intrinsic luminosity of each AGN is taken into account. If $\mathcal{L}$ indicates the luminosity of an AGN, and by assuming no source evolution, the probability to get an event from such a source is proportional to $\mathcal{L}\ z^{-2}\omega_{\rm GZK}(z, E_{\rm thr})$, being $E_{\rm thr}$ the energy of particles at Earth. Thus, protons are then propagated in a ΛCDM Universe until they reach the Earth. We consider only UHECRs with energy above 100 EeV and with arrival direction lying in the field of view of the Pierre Auger Observatory, whose non-uniform exposure is taken into account, as well as its angular uncertainty of 0.8°. The effect of EMF is also taken into account, smearing the direction around the source by sampling a Fisher-von Mises distribution, i.e. the Gaussian counterpart on the sphere. The spreading angle is given by Eq. (3.62) in the case of r.m.s.

strength $B_{\rm rms} = 2$ nG and correlation length $\ell = 1$ Mpc, according to the most recent upper bounds [201]. Additionally, according to the result reported by Pierre Auger Collaboration in the case of the SWIFT-BAT 58-months catalog, the 56% of events in the simulated sky are isotropically distributed [89].

We investigate the clustering signal averaged over several Monte Carlo realizations (10^4 for each astrophysical scenario), by mean of the multiscale autocorrelation function (MAF) [278], versus the angular scale.

In this study, different values of the parameter H_0 are considered, as well as an increasing number of events in the sky. The results are shown in Fig. 4.33 and 4.34, for different astrophysical scenarios and angular scales. In Fig. 4.33 we consider the cases where the intrinsic luminosity of AGN is taken into account (left panel) and not taken into account (right panel), for skies of $N = 200$ protons. In Fig. 4.34 we focus on scenarios where the intrinsic luminosity is accounted for, and vary the number of protons, to put in evidence the impact of the statistics on the clustering signal. It is evident that, for a fixed number of events, the clustering signal increases for increasing values of H_0, whereas it decreases for decreasing number of events, as expected[3].

Such a study suggests that the observation of 100 protons with energy above 100 EeV should be sufficient to probe the Hubble parameter. The large statistics required for this investigation can be attained in few years of activity by future detectors as Auger North or JEM-EUSO. However, it is worth remarking that we have considered only the most conservative case, with events above 100 EeV. In Fig. 3.23 we show that an energy threshold of 60 EeV is sufficient enough to observe significant differences in the flux, and, indirectly, on the clustering of protons. At such energy threshold, we already expect that current experiments as the Pierre Auger Observatory will collect more than 200 UHECRs within a few years, providing the statistics required for this probe.

4.6 Multiscale cross-correlation function

In this section, we will briefly describe the generalization of the MAF method to the cross-correlation analysis with a catalog of candidate sources. By assuming the same notation of the previous section, we make use of another information measure, called *mutual information*, derived from the more general Kullback-Leibler divergence. If $\psi_{\rm data}(\Theta)$ and $\psi_{\rm cat}(\Theta)$ indicate the distribution on the sky, at the angular scale Θ, of the data and catalog, respectively, and if $\psi_{\rm data,cat}(\Theta)$ denotes their joint distribution, the mutual information is defined by

$$\mathcal{I}\left(\psi_{\rm data}(\Theta), \psi_{\rm cat}(\Theta)\right) = \mathcal{D}_{\rm KL}\left(\psi_{\rm data,cat}(\Theta) || \psi_{\rm data}(\Theta)\psi_{\rm cat}(\Theta)\right), \tag{4.36}$$

i.e. the Kullback-Leibler divergence of the joint distribution from the product of the single distributions. If the catalog and the data are totally independent, or, equivalently, if they are *uncorrelated*, then $\psi_{\rm data,cat}(\Theta) = \psi_{\rm data}(\Theta)\psi_{\rm cat}(\Theta)$ and $\mathcal{I}\left(\psi_{\rm data}(\Theta), \psi_{\rm cat}(\Theta)\right) = 0$ follows directly from the definition of $\mathcal{D}_{\rm KL}$ in Eq. (4.20). See Appendix B for further detail.

The generalization to the Multiscale Cross-correlation Function (MCF) is straightforward. Instead of the function $A(\Theta)$, quantifying the deviation of data from an isotropic distribution at the scale Θ, we will make use of the function

$$C(\Theta) = \mathcal{I}\left(\psi_{\rm data}(\Theta), \psi_{\rm cat}(\Theta)\right) = \sum_{k=1}^{N} \psi_k^{\rm data,cat}(\Theta) \log \frac{\psi_k^{\rm data,cat}(\Theta)}{\psi_k^{\rm data}(\Theta)\psi_k^{\rm cat}(\Theta)}, \tag{4.37}$$

[3]The statistical power of the method increases with the number of events.

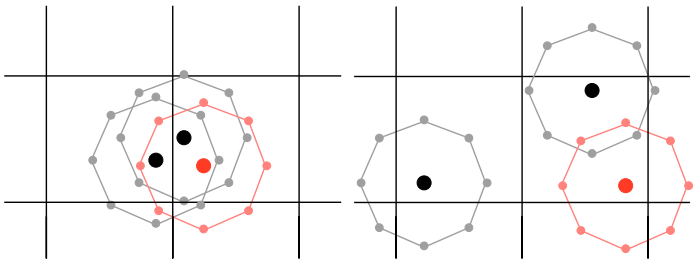


Figure 4.35: Two points (black dots) correlating with one source (red dot): the weighted points are concentrated in two adjacent boxes (left). Two points (black dots) not correlating with one source (red dot): the weighted points are distributed on the neighbor cells (right).

quantifying the amount of correlation between the data and the catalog. As in the case of MAF, the MCF is defined by

$$s(\Theta) = \frac{|C_{\rm data}(\Theta) - \langle C_{\rm iso}(\Theta)\rangle|}{\sigma_{\rm iso}(\Theta)}, \tag{4.38}$$

where the subscript "iso" indicates that the corresponding quantity is estimated for isotropic realizations of the data, thus for uncorrelated data sets. The procedure for the estimation of the distribution of points is the same of the MAF case, involving the dynamical counting method. In Fig. 4.35 we show, respectively, an application of our procedure to a correlating (left panel) and to an uncorrelating (right panel) set of points: in the former case the weighted points are concentrated in two adjacent boxes while in the latter case they are distributed on the neighbor cells.

In the case of MCF, the angular scale $\Theta^\star$, where the significance is minimum, turns to be the most significant *correlation scale*: it is the scale at which a greater number of correlating points occurs with respect to the one occurring by chance, with no regard for a particular number of points, e.g. doublets or triplets, as for standard cross-correlation estimators. In the next chapter, we will show the application of both MAF and MCF to UHECR events detected with the Pierre Auger Observatory.

4.7 Conclusions

In this chapter we have introduced a new multiscale method for the investigation of the anisotropy in the arrival direction distribution of UHECR events. It has been designed to perform both catalog-independent and catalog-dependent analyses, and it is based on information theory and extreme value statistics. We have shown that the multiscale method is a competing tool for the study of both small and large scale anisotropies and correlations, providing a great discrimination power even in presence of a strong background contamination and for quite different astrophysical scenarios.

As a possible application of our method, we have shown how to probe the Hubble parameter with clustering analysis, by assuming that AGN in the nearby Universe (up to ≈ 200 Mpc), reported in the SWIFT-BAT 58-months catalog, are the sources of UHECRs.

In any case, whenever technical progress opened a new window into the surrounding world, I felt the urge to look through this window, hoping to see something unexpected.

B. Rossi

When you know a thing, to hold that you know it, and when you do not know a thing, to allow that you do not know it: this is knowledge.

Kong Fu Tzu

Chapter 5

Search for UHECR anisotropies in Auger and experimental uncertainty effects

In this chapter and in the next ones, we will search for clustering and correlation signals in the arrival direction distribution of UHECRs detected with the Pierre Auger Observatory. Such studies, dealing with the search for the sources of UHECRs, represent two among the hottest topics in astroparticle physics.

The first step is the study of the impact of experimental uncertainties on such signals. This will be the focus of the present chapter. Indeed, the effect of finite angular and energy resolutions is to reduce the clustering signal of the arrival directions and the correlation between UHECRs and catalogs of candidate sources. The angular uncertainty smoothes out any clustered pattern, whereas the energy uncertainty allows low energy events to be ranked as high energy ones, diluting the existing signal with unclustered CRs and further reducing the correlation between their arrival directions and those ones of candidate sources. The multiscale methods, previously introduced, is extensively adopted for such studies. Note that our results have been presented to the Pierre Auger Collaboration as internal reports [283, 284].

In the first section, we will present the method we have developed to smear the arrival directions of simulated UHECRs. Our approach is rather general and it deserves attention because it is one of the main ingredients required by the studies presented in this chapter and in the next one, where the investigation of astrophysical scenarios will be presented. It will be adopted to simulate the finite angular resolution of the Pierre Auger Observatory and to simulate the deflections of UHECRs propagating in the irregular component of the extragalactic magnetic field. In the second section, we will define the Auger data set adopted for our studies, whereas in the remaining sections the impact of experimental uncertainties on clustering and correlation signals, respectively, will be presented.

5.1 Smearing the arrival directions of UHECRs

The angular smearing of the arrival directions of simulated UHECRs represents a fundamental tool for the simulation of realistic sky maps of events. With such a technique, we are able to take into account, in the simulations, the uncertainty on the reconstruction of the "true" arrival direction of UHECRs due to the finite resolution of the Auger Observatory. However, the smearing technique is a quite general method that can also be employed to perform fast simulations, without requiring a full Monte Carlo propagation of particles from the source to the Earth. For instance, it can be adopted to simulate the deflections due to the turbulent component of the extragalactic magnetic field in absence of energy losses. In some cases, as those presented in this chapter, such an approach works very well, whereas in

other cases, a full simulation is required. The extension of this technique to the case of the energy is straightforward and it will be discussed in the next sections.

In general, the smearing of the arrival direction of a simulated UHECR is performed by sampling either a Gaussian or a Rayleigh distribution around its original direction. However, such naive approaches only work in a limited number of circumstances, as we will see in the following, requiring strong assumptions to be fulfilled.

In fact, we show that the distribution of the spherical coordinates significantly deviates from a simple Gaussian distribution, even in the case of small smearing angles, when the point source is placed near a pole. Moreover, we show that the distribution of the angular distances between the source and the simulated events deviates from a Rayleigh distribution for large smearing angles. A wrong smearing of arrival directions may introduce spurious clustering and correlation signals, altering the final results in an unexpected way.

Hence, we have developed a simple method to obtain the correct distribution of events around a source, with no regards for the size of the smearing angle and for the position of the source on the sphere. Such a method, recently presented to the Pierre Auger Collaboration in an internal report [285], will be adopted for any study presented in this chapter and in the next one.

5.1.1 Directional statistics

In one dimension, a Gaussian smearing of spread σ_x around a point $S \equiv (x_0)$ represents the set of points $T_i \equiv (x_i)$ obtained by sampling the normal distribution with mean x_0 and variance σ_x^2. Hence, points are distributed on a straight line as $\mathcal{N}(x_0, \sigma_x)$.

If the line is not straight, as in a circle, the Gaussian smearing requires the circular analogue of the normal distribution, known as the von Mises distribution [286], whose probability density function is given by

$$f(x; \mu, \kappa) = \frac{e^{\kappa \cos(x-\mu)}}{2\pi I_0(\kappa)}, \qquad x \in [a, a+2\pi], a \in \mathbb{R} \tag{5.1}$$

where $\mu \in \mathbb{R}$ is called *location parameter*, $\kappa > 0$ is called *concentration parameter* and $I_0(\kappa)$ is the modified Bessel function of the first kind and of order 0. The points are clustered around $x = \mu$ with dispersion $1/\kappa$, playing the same role of x_0 and σ_x^2, respectively, in the Gaussian smearing on the straight line.

In two dimensions, if x and y are not correlated variables and $\sigma_x^2 = \sigma_y^2 = \sigma^2$, the Gaussian smearing of spread σ around a point $S \equiv (x_0, y_0)$ represents the set of points $T_i \equiv (x_i, y_i)$ obtained by sampling the 2-dimensional normal distribution centered in S and variance σ^2. Hence, points are distributed on a flat plane as $\mathcal{N}(x_0, \sigma) \times \mathcal{N}(y_0, \sigma)$. However, if x and y are correlated and with different dispersions, the most general form of the probability density function is

$$f\left(x, y; x_0, y_0, \sigma_x, \sigma_y\right) = \frac{e^{-\frac{1}{2}\frac{t}{1-\rho^2}}}{2\pi\sigma_x\sigma_y\sqrt{1-\rho^2}} \tag{5.2}$$

with

$$t = \frac{(x-x_0)^2}{\sigma_x^2} + \frac{(y-y_0)^2}{\sigma_y^2} - 2\rho\frac{(x-x_0)(y-y_0)}{\sigma_x\sigma_y}$$

The analogue of a 2-dimensional Gaussian distribution on a spherical surface is known as the von Mises-Fisher distribution [286], that is a particular case of the more general Fisher distribution, whose

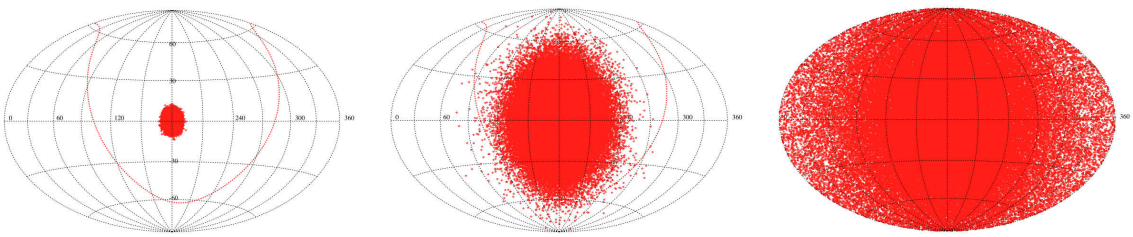


Figure 5.1: Aitoff projection of 10^5 events smeared around $S = (180, 0)$ with smearing $\sigma = 3°$ (left panel), $\sigma = 20°$ (middle panel) and $\sigma = 50°$ (right panel). Equatorial coordinates are shown and the dashed line indicates the galactic plane.

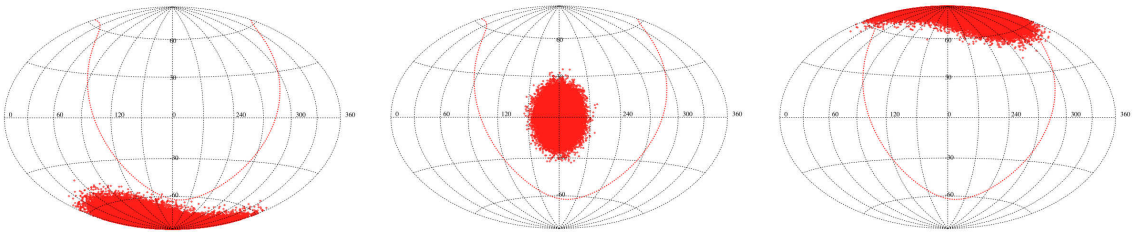


Figure 5.2: Aitoff projection of 10^5 events smeared around a source placed in $S = (60, -80)$ (left panel), $S = (180, 0)$ (middle panel) and $S = (300, 80)$ (right panel). In each case the smearing angle is $\sigma = 8°$. Equatorial coordinates are shown and the dashed line indicates the galactic plane.

probability density function, in the case of a $(p-1)-$dimensional unit sphere, is given by

$$f_{p-1}(\mathbf{x}; \boldsymbol{\mu}, \kappa) = \left(\frac{\kappa}{2}\right)^{\frac{p}{2}-1} \frac{e^{\kappa \boldsymbol{\mu}^T \mathbf{x}}}{\Gamma(p/2)\, I_{p/2-1}(\kappa)} \tag{5.3}$$

where $\kappa \geq 0$, $||\boldsymbol{\mu}|| = 1$ and $I_\nu(\kappa)$ is the modified Bessel function of the first kind and order ν. The parameters $\boldsymbol{\mu}$ and κ are called *mean direction* and *concentration parameter*. In the case of $p = 3$, studied in detail by Fisher [287], points are distributed on a spherical surface, whereas the extension to $p > 3$ is due to Watson and Williams [288]. If $\mathbf{x}$ and $\boldsymbol{\mu}$ are written in spherical polar coordinates, the probability density function becomes

$$f(\phi, \theta; \alpha, \beta, \kappa) = \frac{\kappa \sin\theta}{4\pi \sinh\kappa} e^{\kappa[\cos\theta\cos\alpha + \sin\theta\sin\alpha\cos(\phi-\beta)]} \tag{5.4}$$

where

$$\begin{aligned} \mathbf{x} &= (\cos\theta, \sin\theta\cos\phi, \sin\theta\sin\phi)^T \\ \boldsymbol{\mu} &= (\cos\alpha, \sin\alpha\cos\beta, \sin\alpha\sin\beta)^T \end{aligned}$$

The function in Eq. (5.4) provides the probability density of normally distributed points with random directions (ϕ, θ) around the mean direction $\boldsymbol{\mu}$, and it reduces to the uniform distribution on the sphere for $\kappa = 0$.

Such a probability density function can be further simplified. In fact, if $\boldsymbol{\mu}$ is the pole of a unit sphere then ϕ and θ are independent and we obtain

$$f(\theta; \kappa) = \frac{\kappa \sin\theta}{2\sinh\kappa} e^{\kappa\cos\theta} \tag{5.5}$$

for $\theta \in [0, \pi]$ and ϕ uniformly distributed in the interval $[0, 2\pi]$. In the following we will refer to such a distribution as the *special Fisher distribution*. However, in order to obtain points distributed around $\boldsymbol{\mu}$, the smeared points around the pole should be opportunely rotated.

Sampling the Fisher distribution. Let U be a stochastic variable uniformly distributed in the interval $[a, b]$, with $a, b \in \mathbb{R}$, with probability density function $g(U; a, b) = 1/(b - a)$. In order to obtain a general rule to sample the Fisher distribution by using random uniformly distributed numbers, the differential distribution functions of U and θ are used:

$$g(U; a, b)dU = f(\theta; \kappa)d\theta$$

By integrating the above equation, it follows

$$\frac{U}{b-a} = -\frac{e^{\kappa \cos\theta}}{2\sinh\kappa}$$

The extreme values of θ are 0 and π. By using the relationship $2\sinh\kappa = e^k - e^{-k}$, the corresponding extreme values of the variable U are $(b-a)/(1-e^{2\kappa})$ and $(b-a)/(e^{-2\kappa}-1)$. Without loss of generality, we can assume $a = 0$ and $b = 1$, bounding the variable U to the interval $[(1 - e^{2\kappa})^{-1}, (e^{-2\kappa} - 1)^{-1}]$. With such a prescription, it is straightforward to obtain the variable

$$\theta = \arccos\left[\frac{1}{\kappa}\log\left(-2U\sinh\kappa\right)\right]$$

whose probability density function is given by Eq. (5.5).

Rotating smeared directions. Let (α, δ) the equatorial coordinates of a point source on the unit sphere, and let (ϕ', θ') the angular coordinates of a smeared point around the pole obtained by sampling the uniform distribution and the Fisher distribution defined in Eq. (5.5). The 3-dimensional vector $\mathbf{r}'$ in the Cartesian space, corresponding to the coordinates (ϕ', θ'), is given by

$$\mathbf{r}' = \left(\cos\phi'\cos\theta', \sin\phi'\cos\theta', \sin\theta'\right)^T$$

If, for instance, the pole is assumed to be at $\delta' = -\pi/2$, the rotation matrix R needed to move a smeared point from its sampled direction to the correct one (around the point source), is provided by the product of the two rotation matrices with Euler angles α and $\delta + \pi/2$:

$$R(\alpha, \delta) = \begin{pmatrix} \cos\alpha\cos\delta & -\sin\alpha & \cos\alpha\sin\delta \\ \sin\alpha\cos\delta & \cos\alpha & \sin\alpha\sin\delta \\ -\sin\delta & 0 & \cos\delta \end{pmatrix}$$

By transforming the rotated vector $\mathbf{r} = R(\alpha, \delta)\mathbf{r}'$ to equatorial coordinates, the correct direction (ϕ, θ) of the smeared point around the source is obtained.

Distribution of angular distances. It can be shown that the special Fisher distribution defined by Eq. (5.5) describes the distribution of angular distances between the smeared points and their source. It is worth remarking that the domain of the Rayleigh distribution is the positive real line $[0, +\infty[$, whereas the angular distance is defined in the bounded interval $[0, \pi]$, as the special Fisher distribution. In fact, in the limit of small angles $\sin\theta \approx \theta$ and $\cos\theta \approx 1 - \theta^2/2$, hence:

$$f(\theta; \kappa) \approx \frac{\kappa e^{\kappa}}{2\sinh\kappa}\theta e^{-\frac{1}{2}\kappa\theta^2} \approx \kappa\theta e^{-\frac{1}{2}\kappa\theta^2}$$

that, for $\kappa = \sigma^{-2}$, is equal to the Rayleigh distribution with parameter σ.

5.1.2 Applications

In this section we show some application of our method to some special cases. In Fig. 5.1 we show, in equatorial coordinates, the Aitoff projection of 10^5 events whose directions have been smeared around a source $S = (180, 0)$, with different smearing angles: $\sigma = 3°$ (left panel), $\sigma = 20°$ (middle panel) and $\sigma = 50°$ (right panel).

In Fig. 5.3, for each smearing angle, we show the probability density of the corresponding values of the right ascension (left panel) and of the declination (middle panel), superimposed to the expected Gaussian probability densities. In both cases, the Gaussian density describes the simulations with a good approximation, for small values of the smearing angle. For smearing angles corresponding to medium angular scales, the distribution of simulated events slightly deviate from the Gaussian expectation. At larger smearing angles, the deviation from the Gaussian distribution is significant. In the case of declination, it is worth noticing that the probability of finding an event on the poles is zero, as expected.

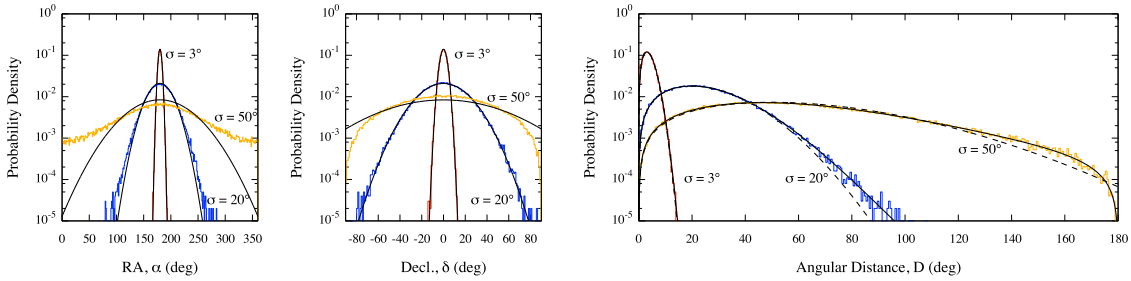


Figure 5.3: **Left:** Probability density of right ascension (left panel) and declination (middle panel) corresponding to sky maps in Fig. 5.1. Solid lines are the expected Gaussian probability densities. **Right:** Probability density of the angular distance of the smeared events from the source, corresponding to sky maps in Fig. 5.1. Solid lines are the expected Fisher probability densities, whereas dashed lines indicate the Gaussian expectations.

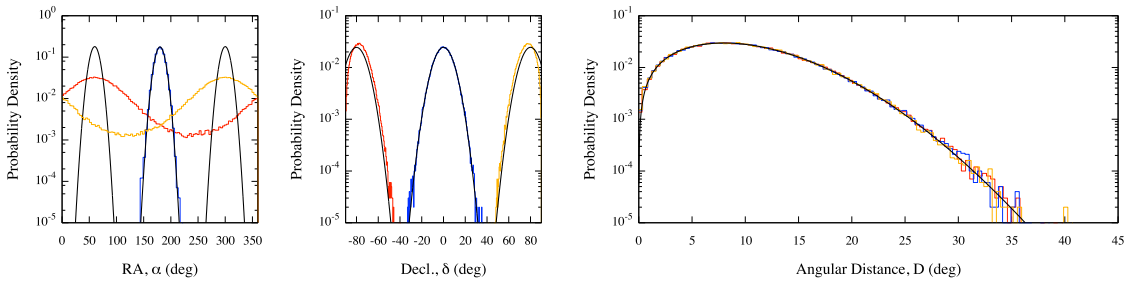


Figure 5.4: **Left:** Probability density of right ascension (left panel) and declination (middle panel) corresponding to sky maps in Fig. 5.2. Solid lines are the expected Gaussian probability densities. **Right:** Probability density of the angular distance of the smeared events from the source, corresponding to sky maps in Fig. 5.2. Solid line indicates the expected Fisher probability density.

In the right panel of Fig. 5.3, for each smearing angle, we show the probability density of the angular distance of the smeared events from the source. For $\sigma = 3°$, the density is well approximated by both a Rayleigh and a Fisher distribution. However, the deviation from a Rayleigh becomes evident by increasing the smearing angle. It is worth remarking that for $\sigma = 50°$, the Fisher distribution shows a cutoff for $D = 180°$, i.e. the maximum value of the angular distance, whereas the Rayleigh distribution, wrongly, do not show such a cutoff, extending to infinity.

In Fig. 5.2 we show, in equatorial coordinates, the Aitoff projection of 10^5 events whose directions have been smeared around three sources, separately. In particular, the sources are placed in $S =$

$(60, -80)$ (left panel), $S = (180, 0)$ (middle panel) and $S = (300, 80)$ (right panel), and in each case the smearing angle is $\sigma = 8°$.

In Fig. 5.4, for each case, we show the probability density of the corresponding values of the right ascension (left panel) and of the declination (middle panel), superimposed to the expected Gaussian probability densities. In both cases, the Gaussian density describes the simulations with a good approximation, if the source is not close to the poles. Conversely, when the source is placed near the poles, the Gaussian density fails to reproduce the simulations.

In the right panel of Fig. 5.4, for each case, we show the probability density of the angular distance of the smeared events from the source. As expected, the density is a special Fisher distribution with concentration parameter $\kappa = 1/8^2$.

Such a method has been independently adopted in PARSEC, a software simulating the propagation of protons in the Universe [210].

5.2 The Auger dataset

The Auger dataset used in this chapter is built by applying to the triggered events a fiducial cut to ensure adequate containment inside the array. The cut requires that at least five active stations surround the station with the highest signal, and that the reconstructed shower core be inside a triangle of active detectors (see Sec. 1.6.2 for details). In the following, we will refer to such an event selection as the standard quality cuts for anisotropy and correlation studies.

For the clustering analysis (Sec. 5.3), we use the data set of UHECR events recorded by the Auger Observatory from 1 January 2004 to 5 June 2010, with reconstructed energies above 40 EeV and zenith angles smaller than 60°. The angular resolution, defined as the angular radius that would contain 68% of the reconstructed events, is better than 0.9°, while the energy resolution is 15%, with a systematic uncertainty of 22%. The total number of CRs passing these criteria is 193. The number of events corresponding to other energy thresholds $E_{\rm thr}$, relevant for the present studies, is reported in Tab. 5.1.

$E_{\rm thr}$ (EeV)	# Events
45	131
50	105
55	73
60	56
65	45
70	29
75	20

Table 5.1: Total number of Auger events above the energy threshold $E_{\rm thr}$.

In the case of analysis in correlation with astronomical catalogues (Sec. 5.4), we focus our attention on the 69 events with energy $\geq$ 55 EeV detected from 1 January 2004 to 31 December 2009. The choice of such a smaller data set is motivated by our interest to investigate the impact of experimental uncertainties on the correlation within 3.1° between the arrival directions of UHECRs with energy above 55 EeV and the positions of active galactic nuclei closer than 75 Mpc, reported by the Pierre Auger Collaboration [89]. For such a purpose, we will also consider three smaller sets of events corresponding to different epochs of data taking. We will name such sub-sets a Period I (1 January 2004 - 26 May 2006), Period II (27 May 2006 - 31 August 2007) and Period III (1 September 2007 - 31 December 2009). We refer to [89] for further information about such periods and for details about the events.

5.3 Search for an anisotropy signal with the MAF and effects of uncertainties

A general approach to investigate the effect of uncertainties is as follows:

1. Smear the physical observables of interest and create ensemble(s) of skies (i.e. *smeared data sets*);
2. For each sky in the ensemble(s) estimate the chance probability of an underlying isotropic distribution of events, by using a method introduced for this purpose;
3. Build the distribution of significances obtained in the previous step and estimate the impact of the smearing on the decision making process about the null hypothesis.

This procedure has been applied in Ref. [289] for the autocorrelation function (ACF) and in Ref. [290, 291] for the enhanced two point (2pt+) [236] and the shape-strength (3pt) [237] methods, to investigate the effect of the angular and the energy resolution on the acceptance/rejection of the null hypothesis in the case of Auger data.

By using the multiscale autocorrelation function (MAF), we have investigated how the chance probability (that Auger data is drawn from an isotropic distribution) changes by varying both energy and angular scales of the analysis. Finally, we have studied, separately, the impact of the angular and energy resolution on the previous result by means of the smearing procedure. In particular, we have created ensembles of skies corresponding to:

- angular smearing of $\sigma = \{0.2°, 0.8°, 2.0°\}$, corresponding to a smaller, equal and greater angular resolution than the Pierre Auger Observatory [292], respectively;
- energy smearing of $\Delta E/E = \{0.05, 0.10, 0.15\}$, corresponding to a smaller, equal and greater energy resolution than the Pierre Auger Observatory [293], respectively;

5.3.1 Search for anisotropy with the MAF method

In order to investigate the dependence on the energy, we have performed a scan over the energy threshold E and the angular scale Θ, according to similar studies for the ACF [294, 295]. In particular, the scan over the energy is performed by fixing the value of $E = E_{\rm thr}$ (ranging from 45 to 70 EeV, in step of 1 EeV) and by selecting all events in the data set with energy equal or greater than $E_{\rm thr}$.

The chance probability $P_{\rm chance}(E, \Theta)$ of an underlying isotropic distribution is estimated for each pair of parameters: Fig. 5.5 (upper panel) shows the result of such an analysis when the penalization due to scan is not taken into account. The global minimum $P_{\rm min} \sim 1.3 \times 10^{-4}$ is found for $E = 52$ EeV and $\Theta = 16.9°$.

If $\tilde{P}_\Theta(E, \Theta)$ is the chance probability penalized for the scan over Θ, we have estimated the dependence on E of the new chance probability, by defining [1]

$$P(E) = \arg\min_{\Theta} \tilde{P}_\Theta(E, \Theta)$$

and obtaining the result shown in Fig. 5.5 (right panel), with a global minimum $P_{\rm min} \sim 0.01$ for $E = 52$ EeV, in great agreement with results from 2pt+ and 3pt methods [296]. Finally, by fixing $E = 52$ EeV, in Fig. 5.6 (left panel) we show the corresponding Auger sky and in Fig. 5.6 (right panel) the penalized chance probability versus the angular scale Θ, remarking the presence of the global minimum at $\Theta = 16.9°$.

[1] The function $\arg\min_x f(x, y)$ retrieves the minimum of the function $f(x, y)$ with respect to the variable x.

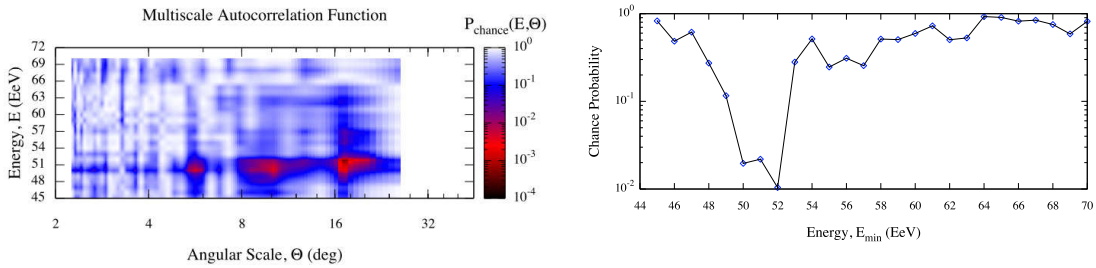


Figure 5.5: Unpenalized (left panel) and penalized (right panel) chance probability for the scan over the energy threshold and the angular scale for Auger data. Each point in the plot shown in the lower panel is obtained from the minimum chance probability, penalized because of the scan over Θ.

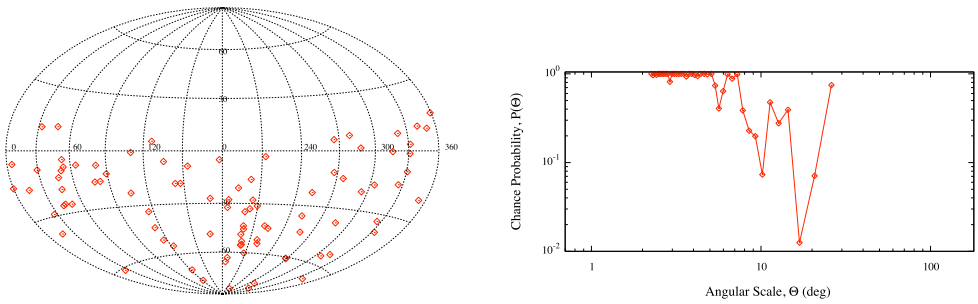


Figure 5.6: Auger events with energy $E \geq 52$ EeV up to 5 June 2010 (left panel) and chance probability versus the angular scale for this sky (right panel).

5.3.2 Effects on the anisotropy signal due to angular resolution

In the angular smearing procedure, the direction of each event in the data set is randomized according to a Fisher-Von Mises distribution, i.e. the Gaussian distribution thrown on the sphere, with dispersion σ. A smeared sky is made up of all randomized events: a smeared mock map is the ensemble of all smeared skies corresponding to a smearing angle σ.

An example ($\sigma = 0.8°$) of the distribution of the angular deviations between the original and 10^3 smeared directions is shown in Fig. 5.7: the solid line corresponds to a Rayleigh distribution with dispersion $\sim 0.8°$, as expected from a Gaussian smearing in angle in the small-angles approximation.

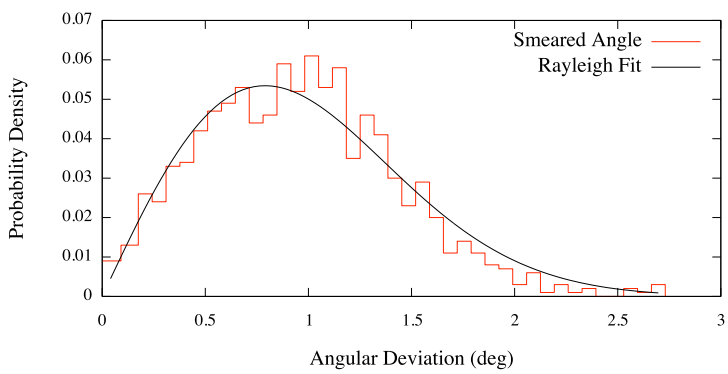


Figure 5.7: Distribution of the angular deviations between the original and 10^3 smeared directions ($\sigma = 0.8°$): the solid line corresponds to a Rayleigh distribution with dispersion $\sim 0.8°$.

Mock map analysis. We have performed the MAF analysis on a smeared map of 10^3 skies, built from a sample of 60 events of the VCV flat mock map. From a previous analysis by means of MAF, it is known that the probability to correctly reject the null hypothesis for the unsmeared skies within such a mock map, is close to 100% for sets of 60 events. In Fig. 5.8 we show the distribution of minimum chance probability as estimated for each smeared sky in the map, for three different values of the dispersion: namely $\sigma = 0.2°$, $\sigma = 0.8°$ and $\sigma = 2.0°$. The value (≈ 0.001) of the minimum chance probability corresponding to the original unsmeared sky is indicated by the arrow. In Tab. 5.2 we report, for each dispersion, the probability to get a chance probability smaller than $1\%, 2\%, 5\%$ and 10%, after the smearing procedure: results show that at least 95% of smeared skies reject the null hypothesis $\mathcal{H}_0$ with a significance of 2%. It is worth remarking that 100% of skies reject $\mathcal{H}_0$ with a significance of 10%, two orders of magnitude higher than the original significance.

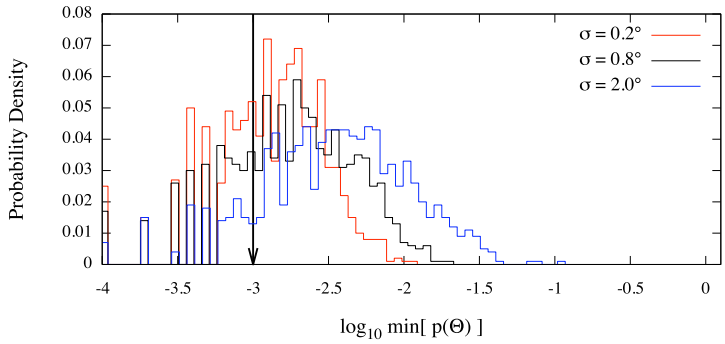


Figure 5.8: Distribution of minimum chance probability estimated for smeared realizations of a VCV flat sample sky of 60 events, for three different values of the dispersion: $\sigma = 0.2°$, $\sigma = 0.8°$ and $\sigma = 2.0°$. The value (≈ 0.001) of the minimum chance probability corresponding to the original unsmeared sky is indicated by the arrow.

σ (deg)	$P \leq 1\%$	$P \leq 2\%$	$P \leq 5\%$	$P \leq 10\%$
0.2°	99.8%	100%	100%	100%
0.8°	97.7%	100%	100%	100%
2.0°	82.5%	94.6%	99.7%	99.9%

Table 5.2: Cumulative chance probability for different significance thresholds by varying the resolution of the angular smearing for a sky sampled from the VCV flat mock map. Chance probability for the unsmeared sky is $\approx 0.1\%$.

Auger data analysis. The same analysis has been performed on Auger data with $E \geq 52$ EeV, i.e. the energy threshold where the chance probability is minimum. In Fig. 5.9 we show the smeared map of Auger data for three values of the dispersion: $\sigma = 0.2°$, $\sigma = 0.8°$ and $\sigma = 2.0°$. In Fig. 5.10 (upper panel) we show the distribution of minimum chance probability as estimated for each smeared sky in the map, for the same values of the dispersion of the previous analysis. The value (≈ 0.01) of the minimum chance probability corresponding to the original unsmeared sky is indicated by the arrow. In Fig. 5.10 (middle panel) we show the distribution, for the three values of the angular dispersion, of the angular scales where the chance probability is minimum: the dominant clustering scale is $\Theta_{min} = 16.9°$. In Fig. 5.10 (lower panel) we show the chance probability versus the angular scale Θ for the unsmeared sky and the average chance probability from the smeared map ($\sigma = 0.8°$) with 68% and 90% regions around the mean value. In Tab. 5.3 we report, for each dispersion, the probability to get a chance probability smaller than $1\%, 2\%, 5\%$ and 10%, after the smearing procedure on Auger data: results show that at least 98.5% of smeared skies reject the null hypothesis $\mathcal{H}_0$ with a significance of 10% when the angular

resolution of the Pierre Auger experiment is taken into account.

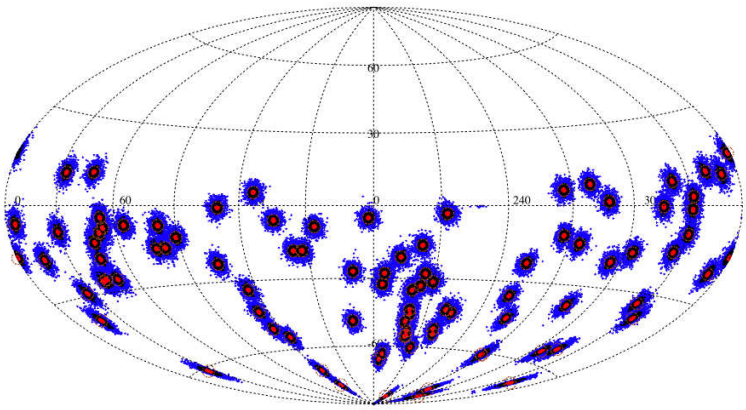


Figure 5.9: Smeared realizations of Auger data with $E \geq 52$ EeV. Map of smeared skies for three different values of the dispersion: $\sigma = 0.2°$ (red dots), $\sigma = 0.8°$ (black dots) and $\sigma = 2.0°$ (blue dots).

σ (deg)	$P \leq 1\%$	$P \leq 2\%$	$P \leq 5\%$	$P \leq 10\%$
0.2°	24.3%	88.6%	99.9%	100%
0.8°	7.3%	35.6%	84.1%	98.5%
2.0°	5.6%	20.5%	58.4%	85.3%

Table 5.3: Cumulative chance probability for different significance thresholds by varying the resolution of the angular smearing for Auger data with $E \geq 52$ EeV. Chance probability for the unsmeared sky is $\approx 1\%$.

5.3.3 Effects on the anisotropy signal due to energy resolution

In the energy smearing procedure, the energy of the $i-$th event in the data set is randomized according to a Gaussian distribution with dispersion σ_{E_i}. A smeared sky is made up of all randomized events: a smeared mock map is the ensemble of all smeared skies corresponding to a smearing energy σ_E.

An example ($\Delta E/E = 0.10$, $E = 52$ EeV) of the distribution of the deviations between the original and 10^3 smeared energies is shown in Fig. 5.11: the solid line corresponds to a Gaussian distribution with a dispersion ~ 5 EeV.

For the angular smearing analysis, we have selected the Auger events with $E \geq 52$ EeV: this data set is composed by the 91 highest energy events. Because of the energy smearing, we use a larger data set, extended down to a lower energy threshold corresponding to 40 EeV: in fact, 27% of events with energy $E = 40$ EeV are expected to be smeared up to an energy of 52 EeV when, for instance, $\Delta E/E = 0.15$.

$\Delta E/E$	$P \leq 1\%$	$P \leq 2\%$	$P \leq 5\%$	$P \leq 10\%$
0.05	5.1%	13.3%	31.3%	53.5%
0.10	2.9%	7.9%	17.7%	34.1%
0.15	6.0%	10.2%	17.7%	28.2%

Table 5.4: Cumulative chance probability for different significance thresholds by varying the resolution of the energy smearing for Auger data with $E \geq 40$ EeV. Chance probability for the unsmeared sky, corresponding to an energy threshold of 52 EeV, is $\approx 1\%$.

After each smearing, the first 91 highest energy events are used for the successive analysis. In order to give a representation of the contaminating events with an unsmeared energy smaller than 52 EeV,

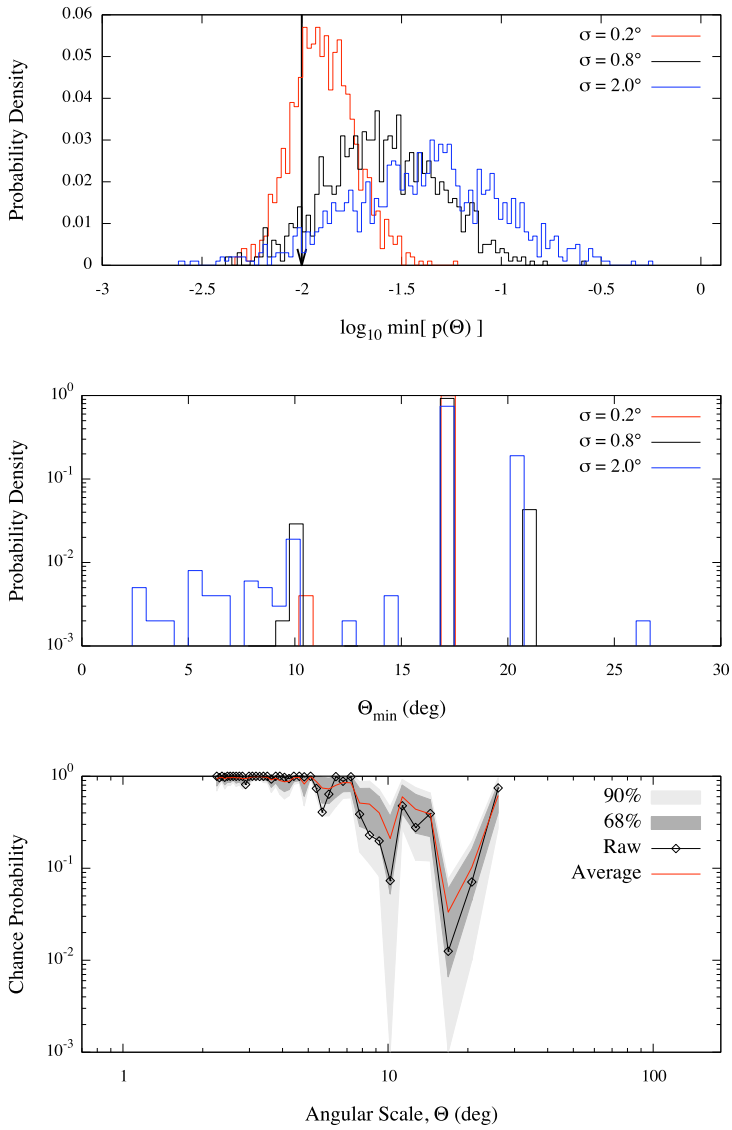


Figure 5.10: Smeared realizations of Auger data with $E \geq 52$ EeV. **Upper Panel.** Distribution of minimum chance probability estimated for the three values of the dispersion. The value (≈ 0.01) of the minimum chance probability corresponding to the original unsmeared sky is indicated by the arrow. **Middle panel.** Distribution of the angular scales where the chance probability is minimum, for the three values of the dispersion. **Lower panel.** Chance probability versus the angular scale Θ for the unsmeared sky (black line) and the average chance probability (red line) from the smeared map ($\sigma = 0.8^\circ$) with 68% and 90% regions around the mean value.

we have assigned an integer rank to each unsmeared event in the original data set, ordered descending by energy. Successively, we have repeated the ranking procedure for each smeared sky. In Fig. 5.12 (left panel) we show the smeared map of Auger data for three values of the relative dispersion: $\Delta E/E = 0.05$, $\Delta E/E = 0.10$ and $\Delta E/E = 0.15$. In Fig. 5.12 (right panel) we show the rank transitions from the unsmeared data set (initial) to the smeared one (final) for $\Delta E/E = 0.10$, and the probability for such a transition, as estimated from 10^3 smeared realizations. As expected, low rank events, corresponding to the highest energy ones, are still present in each smeared sky.

The MAF analysis has been performed on the energy smeared map, built as previously described. In Fig. 5.13 (upper panel) we show the distribution of minimum chance probability as estimated for each

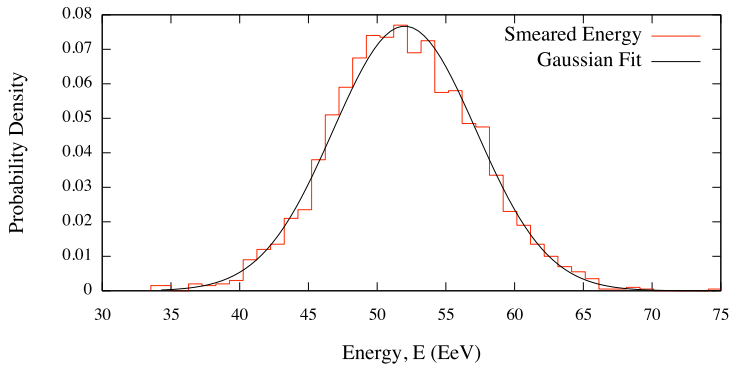


Figure 5.11: Distribution of the energy for 10^3 smeared realization ($\Delta E/E = 0.10$, $E = 52$ EeV): the solid line corresponds to a Gaussian distribution with dispersion ~ 5 EeV.

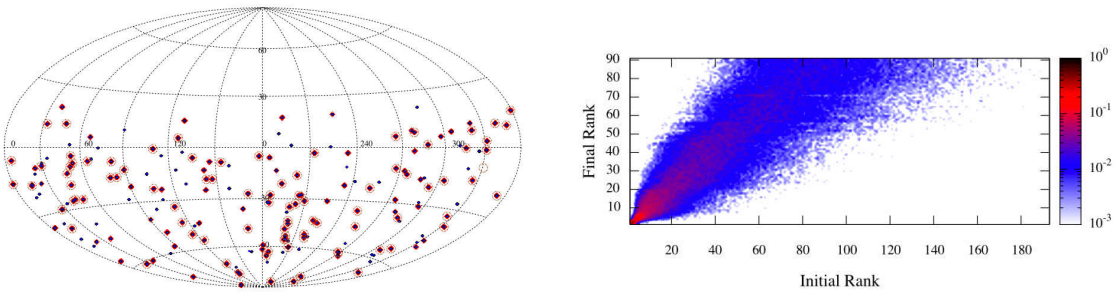


Figure 5.12: Smeared realizations of Auger data with $E \geq 52$ EeV. **Left:** Map of smeared skies for three different values of the relative dispersion: $\Delta E/E = 0.05$ (red dots), $\Delta E/E = 0.10$ (black dots) and $\Delta E/E = 0.15$ (blue dots). **Right:** Rank transition probability from the original data set (initial), of Auger events with $E \geq 40$ EeV, to the smeared one (final), for $\Delta E/E = 0.10$.

smeared sky in the map, for three values of the relative dispersion: $\Delta E/E = 0.05$, $\Delta E/E = 0.10$ and $\Delta E/E = 0.15$. The value (≈ 0.01) of the minimum chance probability corresponding to the original unsmeared sky is indicated by the arrow. In Fig. 5.13 (middle panel) we show the distribution, for the three values of the energy dispersion, of the angular scales where the chance probability is minimum: the dominant clustering scale is $\Theta_{min} = 16.9°$; however there is a greater number of small scales (increasing with the value of the dispersion) than in the case of the angular smearing, meaning that the energy smearing increases the small scale clustering. In Fig. 5.13 (lower panel) we show the chance probability versus the angular scale Θ for the unsmeared sky and the average chance probability from the smeared map ($\Delta E/E = 0.10$) with 68% and 90% regions around the mean value: from this plot it is evident that the small scale clustering is not significative. In Tab. 5.4 we report, for each relative dispersion, the probability to get a chance probability smaller than $1\%, 2\%, 5\%$ and 10%, after the smearing procedure: results show that just a small fraction of smeared skies reject the null hypothesis $\mathcal{H}_0$ with a significance of 10% when the energy resolution of the Pierre Auger experiment is taken into account.

Finally, we have performed the same analysis previously described by using 300 smeared skies. For each smeared sky, we have scanned over the energy threshold E and the angular scale Θ to estimate the minimum chance probability $P_{\text{chance}}(E, \Theta)$. In this case, for the MAF analysis, we have selected the number of events corresponding to the energy where the penalized chance probability $\tilde{P}_\Theta(E, \Theta)$ is minimum. In Fig. 5.14 (upper panel) we show the distribution of $\tilde{P}_\Theta(E, \Theta)$, obtained from such a procedure for $\Delta E/E = 0.10$: the value (≈ 0.01) of the minimum chance probability corresponding to the original unsmeared sky is indicated by the arrow. In Fig. 5.14 (middle panel and lower panel) are shown the distribution of the angular scales and of the number of events over the energy threshold,

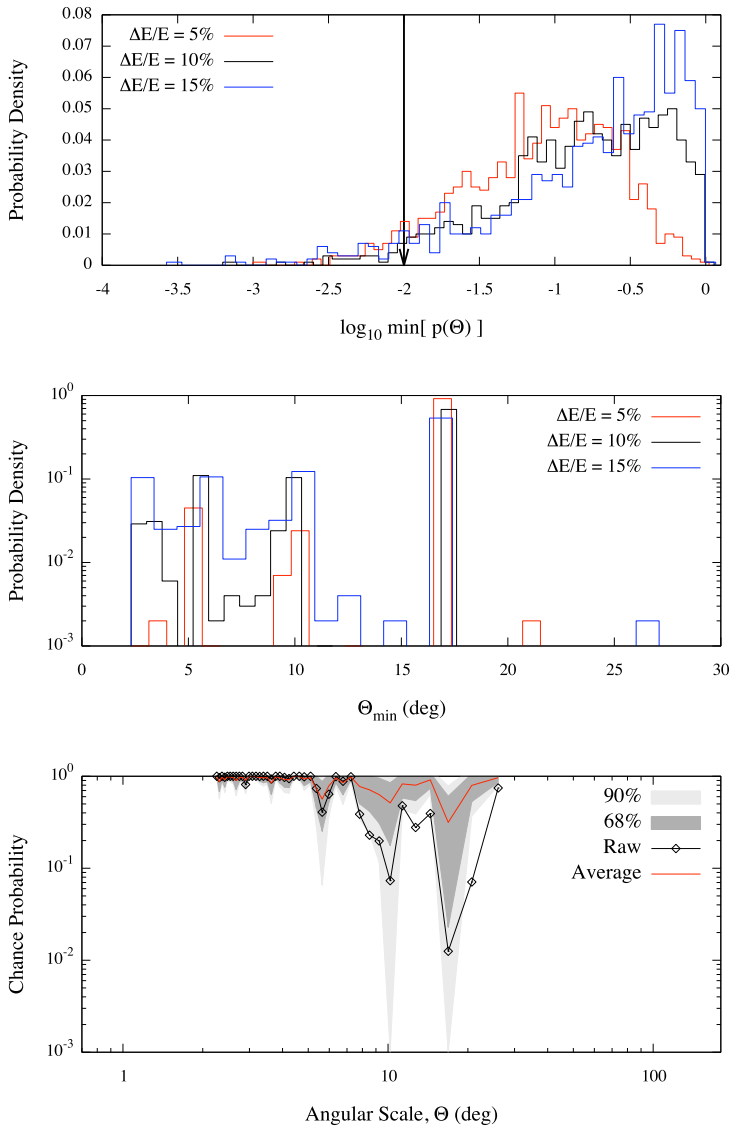


Figure 5.13: **Upper panel.** Distribution of minimum chance probability estimated for the three values of the relative dispersion. The value (≈ 0.01) of the minimum chance probability corresponding to the original unsmeared sky is indicated by the arrow. **Middle panel.** Distribution of the angular scales where the chance probability is minimum, for the three values of the dispersion. **Lower panel.** Chance probability versus the angular scale Θ for the unsmeared sky (black line) and the average chance probability (red line) from the smeared map ($\Delta E/E = 0.10$) with 68% and 90% regions around the mean value.

respectively, where the chance probability is minimum: $\Theta_{\rm min} = 16.9°$ turns to be the dominant clustering scale; scales around 10° represent about 20% of the distribution. In Tab. 5.5 we report, for this value of the dispersion, the probability to get a chance probability smaller than $1\%, 2\%, 5\%$, 10% and 20%, after both the smearing procedure and the scanning over energy and angular scale: results show that an higher fraction of smeared skies reject the null hypothesis $\mathcal{H}_0$ with a significance of 10% when the energy resolution of the Pierre Auger experiment is taken into account; indeed, the significance is smaller than 20% for the 95% of smeared skies, a better result than that one of the previous approach.

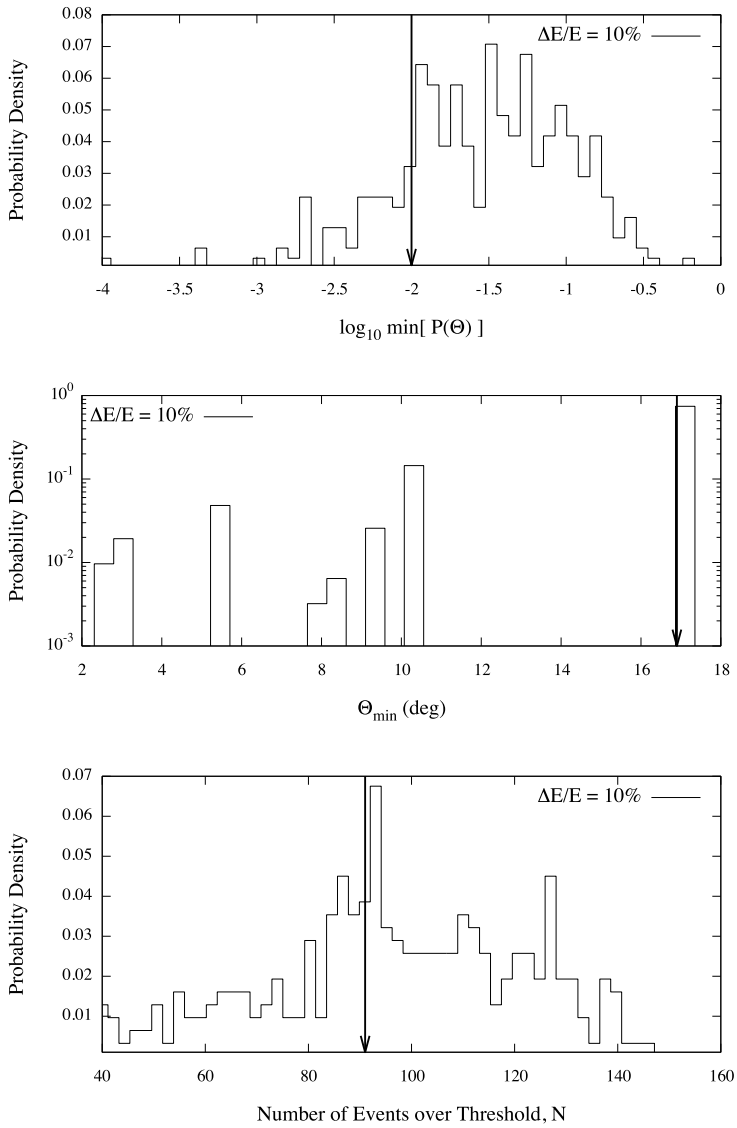


Figure 5.14: Energy-Θ scan of smeared skies, for the relative dispersion $\Delta E/E = 0.10$. **Upper panel.** Distribution of minimum chance probability: the value (≈ 0.01), corresponding to the original unsmeared sky, is indicated by the arrow. **Middle panel.** Distribution of the angular scales where the chance probability is minimum: the value ($16.9°$), corresponding to the original unsmeared sky, is indicated by the arrow. **Lower panel.** Distribution of the number of events over the energy threshold where the chance probability is minimum: the value (91), corresponding to the original unsmeared sky, is indicated by the arrow..

$\Delta E/E$	$P \leq 1\%$	$P \leq 2\%$	$P \leq 5\%$	$P \leq 10\%$	$P \leq 20\%$
0.10	18.7%	40.2%	63.3%	82.31%	96.14%

Table 5.5: Cumulative chance probability for different significance thresholds, for Auger data with smeared energy $E \geq 40$ EeV and relative dispersion $\Delta E/E = 0.10$. Chance probability for the unsmeared sky, corresponding to an energy threshold of 52 EeV, is $\approx 1\%$.

5.3.4 Discussion and conclusion

We have performed several analyses of Auger data up to June 5th 2010, based on a multiscale approach, for detecting an anisotropy signal in the arrival direction distribution of UHECR. From the multiscale autocorrelation analysis, the minimum chance probability ($\approx 1\%$) of an underlying isotropic distribution has been obtained for an energy $E = 52$ EeV, in agreement with previous studies, and a clustering angular scale $\Theta \approx 16.9°$.

Indeed, by means of a smearing procedure, we have investigated the impact on the chance probability of taking into account angular and energy resolution of the surface detector of the Pierre Auger Observatory. Our results showed that for an angular uncertainty $\sigma = 0.8°$, such a chance probability reduced down to 10%. However, for a relative energy uncertainty $\Delta E/E = 0.1$, the anisotropy signal is still present with a significance of about 19% for the 95% of smeared skies. In each case, the dominant clustering scale is still $\Theta \approx 16.9°$.

The obtained results should be interpreted from an ultra-conservative point of view: they represent an upper limit to the real chance probability for the anisotropy signal. In fact, results are affected by the robustness of the method with respect to the smearing, according to the detector resolution.

5.4 Search for correlations with AGNs with the MCF and uncertainties effects

In this section, we present an analysis similar to that one presented in the previous section, but performed by using the multiscale cross-correlation function (MCF) to investigate the impact of experimental uncertainties on the cross-correlation signal between Auger data and the SWIFT-BAT 39-month catalog of AGN in the nearby Universe [268, 297] (for three different redshift: $z < 0.0018$, $z < 0.0024$ and $z < 0.0047$). In particular, we have created ensembles of skies corresponding to:

- angular smearing of $\sigma = 0.8°$, corresponding to the angular resolution of the Pierre Auger Observatory [292];
- energy smearing of $\Delta E/E = 0.10$, corresponding to the energy resolution of the Pierre Auger Observatory [293];

In the following, we will talk about a distribution of points (data) and a distribution of sources (catalog), with no loss of generality. Indeed, we will indifferently refer to a Montecarlo realization of points not correlating with the catalog, as *uncorrelating sky*, *uncorrelating distribution* or *uncorrelating points*.

We consider the 69 events (corresponding to Period I+II+III) with energy $E \geq 55$ EeV [89] detected by the Pierre Auger Observatory. Indeed, sources with known redshift $z < 0.018$ (75 Mpc), $z < 0.024$ (100 Mpc) and $z < 0.047$ (200 Mpc) have been considered separately. In particular Seyfert galaxies, blazars, QSO, and unclassified AGN from the SWIFT catalog (Fig. 5.15, left panel), have been taken into account.

In Fig. 5.15 (right panel) we show the penalized chance probability versus the angular scale Θ, for the cross-correlation between Auger data and SWIFT: the global minimum at $\Theta = 2.53°$ ($P \approx 8.0 \times 10^{-4}$) is evident when sources within 75 Mpc and 100 Mpc are considered.

5.4.1 Effects on the correlation signal due to angular and energy resolutions

We have separately performed the MCF analysis on angular and energy smeared maps of 10^3 skies, built from the Auger data set of 69 events with $E \geq 55$ EeV. For the angular smearing analysis, we have selected $\sigma = 0.8°$ and Auger events with $E \geq 55$ EeV up to 31 December 2009. Because of the

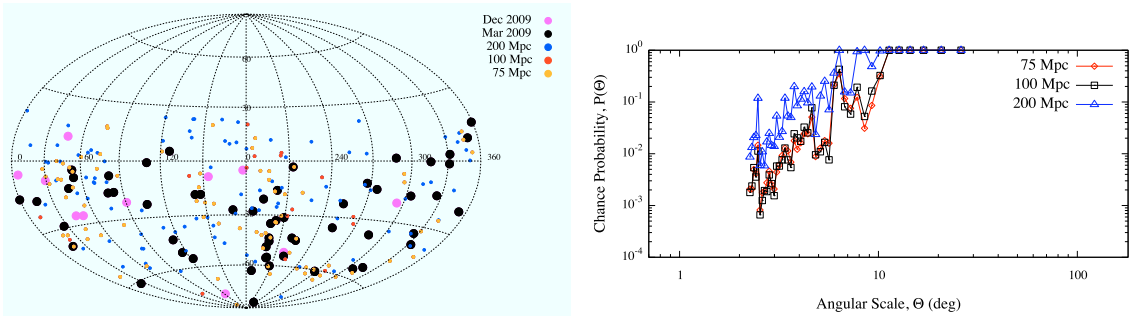


Figure 5.15: **Left:** Candidate sources (small dots of different colors are used according to redshifts: $z < 0.018$, $z < 0.024$ and $z < 0.047$) and Pierre Auger Observatory data (black and purple big dots). Equatorial coordinates are shown. **Right:** Chance probability for the cross-correlation between Auger data and SWIFT.

energy smearing, a larger data set, we extend the data set to a lower energy threshold corresponding to 40 EeV: in fact, about 1% of events with energy $E = 40$ EeV are expected to be smeared up to an energy of 55 EeV, if $\Delta E/E = 0.10$.

After each smearing, the first 69 highest energy events are used for the successive analysis. In order to give a representation of the contaminating events with an unsmeared energy smaller than 55 EeV, we assigned an integer rank to each unsmeared event in the original data set, ordered descending by energy. Successively, we have repeated the ranking procedure for each smeared sky, with results shown in Fig. 5.12.

In Fig. 5.16 we show the distribution of minimum chance probability, estimated from angular and energy smeared maps, for the cross-correlation between Auger data and objects from the SWIFT catalog within three different redshift ($z < 0.018$, $z < 0.024$ and $z < 0.047$). The values of the minimum chance probability, corresponding to the original unsmeared skies, are indicated by the arrows.

In Tab. 5.6 we report, for each redshift, the probability to get a chance probability smaller than $1\%, 2\%, 5\%$ and 10%, after the angular and energy smearing procedures.

Finally, in Fig. 5.17 and 5.18 we show the chance probability versus the angular scale Θ for the unsmeared sky and the average chance probability from the smeared maps ($\sigma = 0.8°$ and $\Delta E/E = 0.10$, respectively) with 68% and 90% regions around the mean value.

Results show that 100% of smeared skies reject the null hypothesis $\mathcal{H}_0$ with a significance of at most 2% when the angular resolution of the Pierre Auger experiment is taken into account for the cross-correlation with objects within 100 Mpc. Indeed, almost 100% of smeared skies reject the null hypothesis $\mathcal{H}_0$ with a significance of at most of 2% when the energy resolution of the Pierre Auger experiment is taken into account, again for distances within 100 Mpc.

It is worth remarking that for both smearing analyses, 100% of smeared skies kept correlated with the SWIFT catalog at an angular scale of about 2.6°.

5.4.2 Discussion and conclusion

We have performed several analyses of Auger data up to 31 December 2009, based on a multiscale approach, for detecting a correlation signal with the distribution of astrophysical objects in the SWIFT catalog.

From the multiscale cross-correlation analysis, the minimum chance probability ($\approx 0.1\%$) of an uncorrelated distribution of events has been obtained for sources within 75 Mpc and 100 Mpc, in agreement with previous studies, and a correlation angular scale $\Theta \approx 2.6°$.

Indeed, by means of a smearing procedure, we have investigated the impact on the chance probability of taking into account angular and energy resolution of the surface detector of the Pierre Auger Obser-

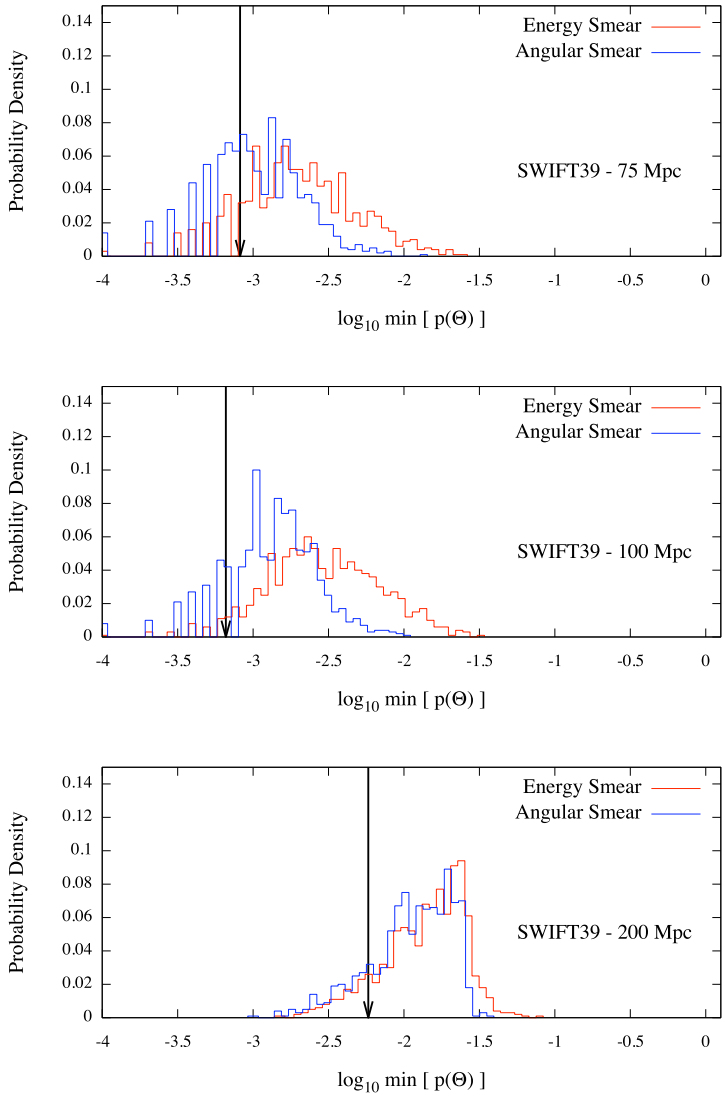


Figure 5.16: Distribution of minimum chance probability, estimated from angular and energy smeared maps, for the cross-correlation between Auger data and objects from the SWIFT catalog within three different redshift ($z < 0.018$, $z < 0.024$ and $z < 0.047$). Chance probability for the unsmeared sky, ≈ 0.0008 (for $z < 0.018$ and $z < 0.024$) and 0.006 (for $z < 0.047$), is indicated by the arrows.

vatory. Our results showed that for an angular uncertainty $\sigma = 0.8°$, such a chance probability reduced down to 2% for objects within 100 Mpc. However, for a relative energy uncertainty $\Delta E/E = 0.1$, the correlation signal is still present even if it reduced down to 2% for objects within 100 Mpc. It is worth remarking that for both smearing analyses, 100% of smeared skies kept correlated with the SWIFT catalog at an angular scale of about 2.6°.

The obtained results should be interpreted from an ultra-conservative point of view: they represent an upper limit to the real chance probability for the anisotropy signal. In fact, results are affected by the robustness of the method with respect to the smearing, according to the detector resolution.

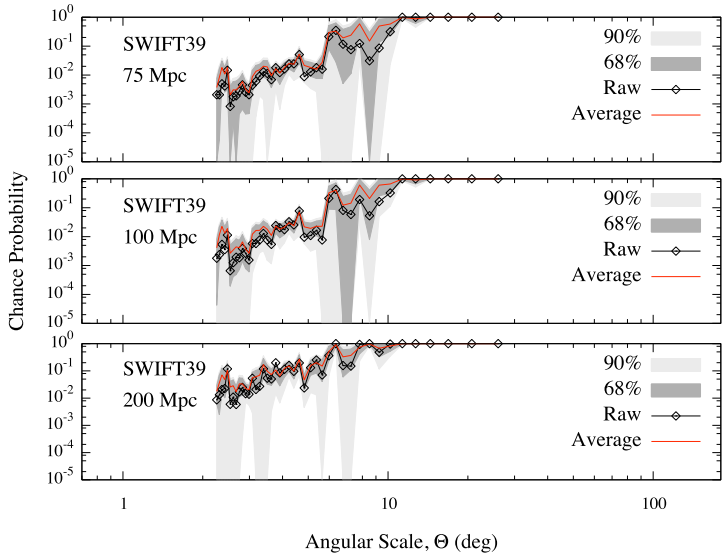


Figure 5.17: **Angular smearing.** Chance probability versus the angular scale Θ for the unsmeared sky (black line) and the average chance probability (red line) from the smeared map ($\sigma = 0.8°$) with 68% and 90% regions around the mean value. From top to bottom each plot corresponds to the cross-correlation between Auger data and the objects from the SWIFT catalog, for three different redshift: $z < 0.018$ and $z < 0.024$ and $z < 0.047$.

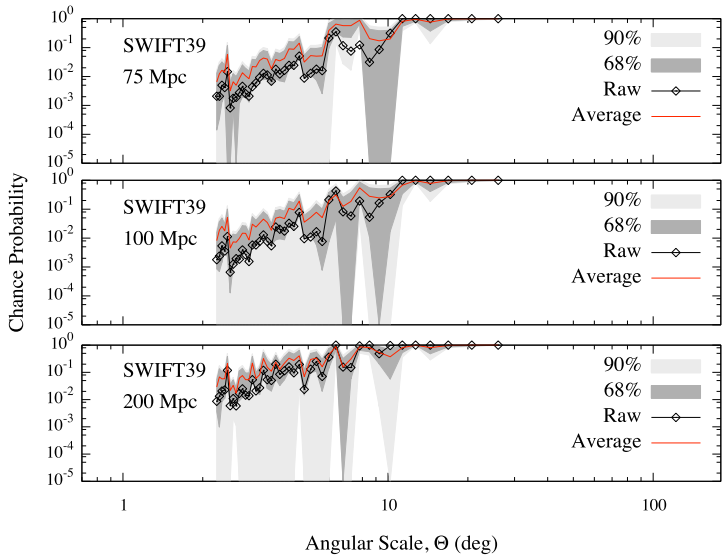


Figure 5.18: **Energy smearing.** Chance probability versus the angular scale Θ for the unsmeared sky (black line) and the average chance probability (red line) from the smeared map ($\Delta E/E = 0.10$) with 68% and 90% regions around the mean value. From top to bottom each plot corresponds to the cross-correlation between Auger data and the objects from the SWIFT catalog, for three different redshift: $z < 0.018$ and $z < 0.024$ and $z < 0.047$.

5.5 Conclusions

In this chapter, we have extensively used multiscale methods, introduced in the previous chapter, to search for an anisotropy signal in Auger UHECR data, as well as for a correlation with candidate objects in astronomical catalogues. We have investigated the influence of angular and energy uncertainties on both the anisotropy and the correlation signals of events detected with the Pierre Auger Observatory.

Angular smearing $\sigma = 0.8°$				
z (redshift)	$P \leq 1\%$	$P \leq 2\%$	$P \leq 5\%$	$P \leq 10\%$
0.018	99.9%	100%	100%	100%
0.024	97.3%	100%	100%	100%
0.047	38.6%	81.7%	100%	100%

Energy smearing $\Delta E/E = 0.10$				
z (redshift)	$P \leq 1\%$	$P \leq 2\%$	$P \leq 5\%$	$P \leq 10\%$
0.018	96.5%	99.7%	100%	100%
0.024	92.3%	99.1%	100%	100%
0.047	28.9%	67.6%	99.5%	100%

Table 5.6: Angular and energy smeared maps of Auger data with $E \geq 55$ EeV: cumulative chance probability for different significance thresholds by varying the redshift of objects in the SWIFT catalog. Chance probability for the unsmeared sky is $\approx 0.1\%$ (for $z < 0.018$ and $z < 0.024$) and 0.6% (for $z < 0.047$).

The effects of both angular and energy resolution of the detector are generally ignored in clustering studies: we have shown that they have an impact on the estimated signal. In particular, the energy resolution significantly affects the clustering, whereas the angular resolution negligibly affects such studies.

Our study has benefited of the smearing technique introduced at the beginning of this chapter, to account for both angular and energy resolution of the detector. In the next chapters, we will adopt the same methods to explore possible astrophysical scenarios able to explain data collected with the Pierre Auger Observatory.

The scientist has a lot of experience with ignorance and doubt and uncertainty, and this experience is of very great importance, I think.

R.P. Feynman

It is better to light one small candle than to curse the darkness.

Kong Fu Tzu

Chapter 6

Clustering and correlation of UHECR in Auger: astrophysical scenarios

The nature of the sources of UHECRs is still not known. In the first chapter we have discussed several candidates, characterized by different astrophysical features. In this chapter, we focus our attention to one of the global observables representing a distinctive signature of the nature of such candidate sources, namely their number density, defined as the number of objects per unit volume. An indirect estimation of the density of sources should help to restrict the number of candidates: in the following, we present the results of our studies which provide lower bounds to such a density by means of clustering analysis.

Hence, we explore the astrophysical scenarios that could be responsible for the observed clustering and correlation in Auger data, as presented in the previous chapter. We perform several studies, independent from each other but sharing the same common thread, i.e. the density of sources. In particular, we i) infer the density of sources of UHECRs; ii) put bounds on the r.m.s. strength of the extragalactic magnetic field; iii) investigate the possibility that nearby black holes are the sources of observed UHECRs; iv) investigate the impact of the choice of the catalog on the correlation signal. We describe our results in four separated sections, each one dedicated to a particular study that has been presented to the Pierre Auger Collaboration as an internal report.

Simulations with HERMES (see Chap. 2 and 3) and analyses with multiscale methods (see Chap. 4) have been extensively used for such investigations. We refer to Sec. 5.2 for the description of the selection of Auger events used in the present chapter.

6.1 Bounds on the density of UHECR sources with MAF

6.1.1 Relating the density of UHECR sources to the observation of clustering

The idea behind the studies presented in this chapter is sketched in Fig. 6.1. We show how the intensity of the clustering signal is related to the density of sources of UHECRs and to the deflections that such particles experience during their propagation, because of the intervening magnetic fields. In order to understand the figure, we have to consider some extreme astrophysical scenarios.

Let us consider the case of small deflections, i.e. the intensity of magnetic fields is weak and the mass composition of UHECRs is light. If the number of sources is small and the number of observed particles is fixed, UHECRs tend to be highly clustered around such sources and a certain number of clusters is expected. Conversely, for an increasing number of sources, a decreasing number of clusters is expected, together with a diluted clustering signal with respect to the previous case.

Let us consider now a small number of sources and increasing deflections, corresponding to the astrophysical scenario with more intense magnetic fields and/or heavier mass composition. In this case,

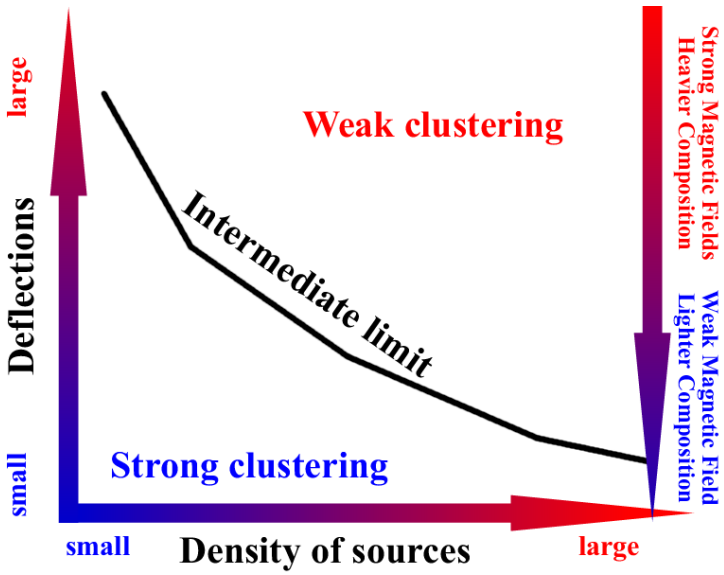


Figure 6.1: Schematic representation of the relationship between the clustering of UHECRs, the density of their sources and the deflections they experience because of the intervening magnetic fields.

clusters around sources are still detectable up to a certain angular scale. Conversely, for an increasing number of sources, the clustering signal decreases until it completely vanishes in the limit of very intense magnetic fields, heavy composition and high density of sources.

Hence, the statistical analysis of clustering may provide tight constraints on astrophysical models of UHECRs, and interesting bounds on the density of sources can be put. In the following, we will briefly report existing studies in this direction and we will present our original contribution to this hot topic in the rest of the chapter.

6.1.2 Relating our work to existing studies

Revealing and understanding the astrophysical sources of ultra high energy cosmic rays (UHECR) is still an open problem, although a significant correlation with active galactic nuclei (AGN) has been recently reported by the Auger collaboration [88, 89, 208]. Thus, it is of fundamental importance to constrain the density of candidate source of UHECR by employing several different approaches.

By assuming small deflection of UHECR during propagation, an elegant statistical analysis of clustering, based on theoretical arguments, has been proposed some years ago. By applying this method to astrophysical models involving extragalactic sources and by using the 14 UHECR events with $E > 10^{20}$ eV available at that time, the value $\approx 6 \times 10^{-3}$ Mpc^{-3} has been obtained for the source density [226]. However, this result is subjected to high variance because of the small statistics. An extension of such an analysis, providing also confidence intervals for the source density under reasonable assumptions, has been given in Ref. [266], where the interval $180^{+2730}_{-165} \times 10^{-3}$ Mpc^{-3} has been obtained at 68% confidence level. A successive study, involving a Monte Carlo approach to the problem and different astrophysical assumptions, has given 10^{-6} Mpc^{-3} as a lower bound to source density more luminous than 20.5 mag, by using the Optical Redshift Survey (ORS) galaxy sample and events with energy $E > 4 \times 10^{19}$ eV detected by the AGASA experiment [298]. The analysis on the same data set of AGASA events, involving also the study of the flux of UHECR, has provided $\approx 10^{-6}$ Mpc^{-3} as a value for source density, with large uncertainties [299]. The case of a finite number of uniformly distributed proton sources has been also investigated in Ref. [231], where the value $1 - 4 \times 10^{-5}$ Mpc^{-3} has been estimated by using AGASA data, although the estimated interval at 95% confidence level ranges from

$\approx 2 \times 10^{-6}$ Mpc^{-3} to $\approx 10^{-2}$ Mpc^{-3}. More recently, both *local* and *global* approaches involving the two-point autocorrelation function (ACF) or its variants, have been used to constrain the source density and other purposes. Representative studies can be found in Refs. [234, 235], where toy models for sources, mimicking the distribution of matter in the nearby universe or a uniform distribution, have been investigated by using the 27 Auger events available at that time [208]. Results from such analyses suggest a source density ranging from $\approx 0.2 \times 10^{-4}$ Mpc^{-3} to $\approx 5 \times 10^{-4}$ Mpc^{-3} with upper bound $\approx 10^{-2}$ Mpc^{-3} at 95% confidence level and for different values of UHECR energy threshold. Similar scenarios have been investigated in Ref. [300], where the interval $0.0015 - 4 \times 10^{-4}$ Mpc^{-3} has been given at 90% confidence level for angular scales smaller than 6° by using the same data.

The main goal of the present study is to estimate both lower and upper bounds to the source density of UHECR, by using the multiscale method [278]. We will consider two toy models: one following the large scale structure of matter and another one with uniformly distributed sources in the nearby universe (up to 250 Mpc). In any case, sources are assumed to be with uniform and equal luminosity, with a spectrum whose injection index is 2.7 and producing only UHE protons. The produced particles are successively propagated in a ΛCDM universe, taking into account non-negligible energy-loss processes in the cosmic microwave background photon field in the limit of a null extragalactic magnetic field. The clustering analysis is performed on Auger-like skies by varying both the density of sources from 10^{-6} Mpc^{-3} up to 10^{-3} Mpc^{-3} and the energy threshold of protons from 50 EeV to 60 EeV. Auger data is used to constrain the source density.

This section is organized as follows. The first part is dedicated to the description and the simulation of the astrophysical models adopted for this study. In the second part, we analyze simulated data and we use the values for clustering obtained from Auger data to constrain the source density: we adopt two *local* methods and one *global* method for this purpose, comparing and discussing the results.

6.1.3 Astrophysical models and simulations

We consider two simple toy models for investigating the ability of the MAF method to discriminate among different source densities. It is worth remarking that we are considering sources of protons, under the assumption of a null extragalactic magnetic field. In the first model, events are generated from sources distributed according to the Two Micron All Sky Survey - Redshift Survey (2MRS) extended source catalog and, in the following, we will refer to this model as "2MASS" or "2MRS". In the second model, events are generated from sources isotropically and homogeneously distributed in a sphere of a given radius, and, in the following, we will refer to this model as "ISOHOM".

The particular choice of such models is justified by the fact that for a fixed source density n_s, we are interested in investigating the clustering differences between skies of events following the distribution of matter in the nearby universe (2MASS model) and skies of events generated by randomly distributed sources (ISOHOM model). Moreover, we are interested in estimating bounds to source density by comparing the clustering features of models with the clustering features of real data. For both models we consider the following constraints:

1. **Radius.** The maximum distance to be adopted for simulations does not depend on the adopted cosmological model but it mainly depends on the energy threshold of protons, although we have found a dependence on the Hubble parameter at the present time and a weak dependence on the injection index s. In this study, we consider a ΛCDM cosmology with $H_0 = 70.0$ km/s/Mpc, $s = 2.7$ and three different values of the energy threshold, namely 50, 55 and 60 EeV. The value of the energy threshold affects the GZK horizon, defined as the distance within which 90% of the observed flux above the energy threshold is expected to be produced.

2. **Source weighting.** The probability to get an event from a given source depends only on the exposure of the detector [222] (the Pierre Auger Observatory in this study), the source distance z

and the energy of the event by $z^{-2}\omega_{\rm GZK}(z, E_{\rm thr})$, where the function $\omega_{\rm GZK}(z, E_{\rm thr})$ is defined in Eq. (3.60). According to previous studies [300], we do not take into account the weight due to the flux of the source (equal-flux assumption), and we do not consider the angular deflections due to the unknown magnetic field.

3. **Completeness of the catalog.** We make use of the 2MRS catalog because it is the most densely sampled all-sky redshift survey to date. It is a compilation [301] of the redshifts of the $K_{\rm mag} < 11.25$ brightest galaxies from the 2MASS catalog [302]. It contains approximately 22,000 galaxies within 200 Mpc, providing an unbiased measure of the distribution of galaxies in the local universe, out to a mean redshift of z = 0.02, and to within 10° of the Galactic plane. The 2MRS catalog is complete for $|b| < 10°$, i) $\{M < -24, D < 120 \text{ Mpc}\}$ (with 11969 sources) or ii) $\{M < -25.1, D < 200 \text{ Mpc}\}$ (with 2638 sources), where M is the source magnitude and D is the source distance. The density of sources is strongly constrained by both the cuts in magnitude and distance: for instance, the second cut allows to investigate only source densities smaller than $\approx 10^{-4}$ Mpc^{-3}. In order to investigate higher densities, we extend the original catalog with sources isotropically distributed in the sky and whose distances D_k are such that the catalog becomes complete above 120 Mpc, up to the GZK horizon corresponding to the energy threshold adopted. In Fig. 6.2 we show the number of sources as a function of the distance for the original 2MASS catalog with $\{M < -24, D < 200 \text{ Mpc}\}$ (complete up to 120 Mpc), a completed version[1] of the catalog up to 200 Mpc, and another completed version of the catalog up to 250 Mpc generated by our simulation code. In the following we will indicate with "2MASS+ISO" any extended version of the 2MASS original catalog with isotropic sources. In order to evaluate the amount of isotropic contamination due to the completing procedure, we compute the following quantity :

$$\mathcal{C}(n_s, E_{\rm thr}) = \frac{\int_{z_0}^{z_{\rm max}} N_{\rm iso}(z) z^{-2} \omega_{\rm GZK}(z, E_{\rm thr}) dz}{\int_{z_0}^{z_{\rm max}} N_{\rm cat}(z) z^{-2} \omega_{\rm GZK}(z, E_{\rm thr}) dz}$$

$$\int_{z_0}^{z} N_{\rm iso}(z') + N_{\rm cat}(z') dz' = \frac{4}{3}\pi n_s \left(\frac{c}{H_0} z\right)^3$$

where $N_{\rm cat}$ and $N_{\rm iso}$ are respectively the number of sources in the catalog and the number of sources added from an isotropic distribution.

The case of ISOHOM model is much simpler, because isotropic sources are homogeneously generated in the sphere of a given radius and the corresponding catalog is complete by construction. It is worth remarking here that in the cases of 2MASS+ISO and ISOHOM models, the parts of the catalogues corresponding to random sources are not kept fixed but are produced "online": each sky of events is generated from a different random realization of the model. Finally, catalogues corresponding to 2MASS model are subjected to the additional cut $|b| < 10°$, whereas those ones corresponding to ISOHOM model are not.

In Tab. 6.1 we report the parameters adopted to generate catalogues for both 2MASS and ISOHOM models, successively adopted to generate mock maps of unsmeared events as described above. It is worth noticing that skies generated by 2MASS model are expected to be anisotropic up to a certain density and that even ISOHOM skies can be anisotropic for small values of the source density. We will verify this hypothesis in the next sections. It can be shown that the amount of isotropic contamination for all 2MASS models is a simple linear function of the density as shown in Fig. 6.3.

Although some of our assumptions could appear not physical, we will see in the next sections that very interesting conclusions may be drawn to constrain the source density of UHECR.

[1] This catalog has been generated by S. Mollerach and has been made available during private communications.

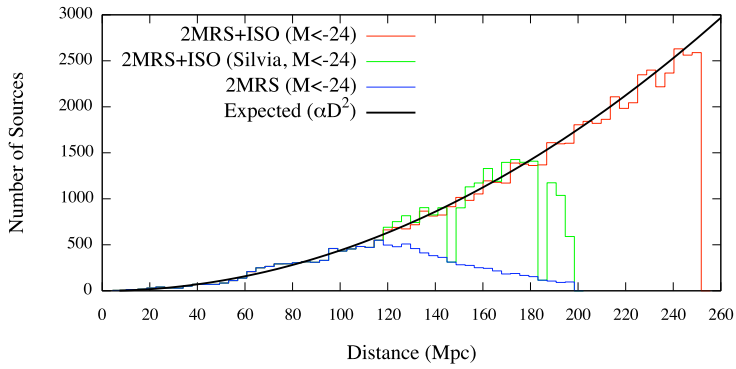


Figure 6.2: Number of sources as a function of the distance for: the original 2MASS catalog with $\{M < -24, D < 200 \text{ Mpc}\}$ (blue line), a completed version of the catalog by S. Mollerach (green line), and another completed version of the catalog generated by our simulation code (red line).

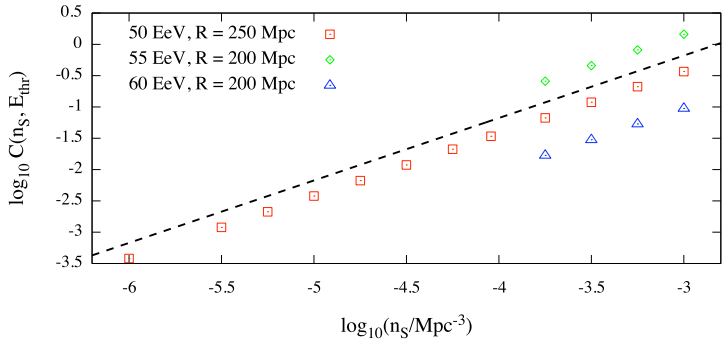


Figure 6.3: 2MASS model. Isotropic contamination with respect to the original catalog as a function of the source density n_s for the 3 different energy thresholds.

6.1.4 Clustering analysis of simulated sky maps and Auger data

In this section we will analyze the mock maps of events generated from catalogues corresponding to configurations in Tab. 6.1, for both 2MASS and ISOHOM models. We investigate the statistical power of MAF against the models considered in this study, by varying all parameters in our parameter space (i.e. the model, the energy threshold of protons and the source density). Then we investigate the ability of MAF to discriminate among different source densities by employing two "local methods" and a "global method". The comparison with results obtained from Auger data is also discussed. For the sake of completeness, we use the official data set of Auger events up to 25 January 2011 with standard quality cuts for anisotropy studies and energy $E \geq E_{\text{thr}}$. Moreover, in the case of 2MASS model only Auger events with $|b| \geq 10°$ are considered. In Fig. 6.4 we show the skymap of 2MASS sources superimposed on directions of Auger events for different energy thresholds.

Statistical power

In order to estimate the range of source densities where the MAF method is able to perform a significant discrimination from isotropy, we investigate the statistical power of MAF against the 2MASS and the ISOHOM models. In this specific study, the power $1 - \beta$ is defined as the probability of correctly rejecting the null hypothesis of isotropy when it is known to be false. It is worth remarking that the isotropy we are referring to should not be confused with the isotropy of sources in the ISOHOM model.

R	$E_{\rm thr}$	$\log_{10} n_s/{\rm Mpc}^{-3}$	Note
200	55	-3.000	2MASS+ISO with $M < -24$
		-3.250	2MASS+ISO with $M < -24$
		-3.500	2MASS+ISO with $M < -24$
		-3.750	2MASS+ISO with $M < -24$
		-4.043	2MASS with $M < -25.1$
		-4.250	2MASS with $M < -25.1$
		-4.500	2MASS with $M < -25.1$
		-4.750	2MASS with $M < -25.1$
		-5.000	2MASS with $M < -25.1$
		-5.250	2MASS with $M < -25.1$
		-5.500	2MASS with $M < -25.1$
		-6.000	2MASS with $M < -25.1$
200	60	-3.000	2MASS+ISO with $M < -24$
		-3.250	2MASS+ISO with $M < -24$
		-3.500	2MASS+ISO with $M < -24$
		-3.750	2MASS+ISO with $M < -24$
		-4.043	2MASS with $M < -25.1$
		-4.250	2MASS with $M < -25.1$
		-4.500	2MASS with $M < -25.1$
		-4.750	2MASS with $M < -25.1$
		-5.000	2MASS with $M < -25.1$
		-5.250	2MASS with $M < -25.1$
		-5.500	2MASS with $M < -25.1$
		-6.000	2MASS with $M < -25.1$
250	50	-3.010	2MASS+ISO with $M < -24$
		-3.250	2MASS+ISO with $M < -24$
		-3.500	2MASS+ISO with $M < -24$
		-3.750	2MASS+ISO with $M < -24$
		-4.000	2MASS+ISO with $M < -24$
		-4.250	2MASS+ISO with $M < -24$
		-4.500	2MASS+ISO with $M < -24$
		-4.750	2MASS+ISO with $M < -24$
		-5.000	2MASS+ISO with $M < -24$
		-5.250	2MASS+ISO with $M < -24$
		-5.500	2MASS+ISO with $M < -24$
		-6.000	2MASS+ISO with $M < -24$

Table 6.1: List of parameters adopted to generate the catalogues for 2MASS model; ISOHOM catalogues are generated with the same parameters, with the only difference that all sources are random. R is the maximum source distance in Mpc units, $E_{\rm thr}$ is the proton energy threshold in EeV units and M indicates the source magnitude.

In Fig. 6.5 we show the power vs source density for both 2MASS (left panel) and ISOHOM (right panel) mock maps. A sky is labelled as *anisotropic* if, for the fixed value $\alpha = 1\%$ of the significance, the penalized chance probability defined in Eq. (4.22) is smaller or equal than α. For applications, a power of 90% is generally required: under this threshold the method could miss to detect an existing anisotropy signal. For considered mock maps, the power increases by decreasing the source density, as expected. In fact, smaller densities reflect a small number of sources and, of consequence, a higher number of multiplets per source causing a greater deviation from isotropy.

Our analysis suggests that mock maps are likely to be isotropic, instead of anisotropic, above a certain density threshold depending on the considered model. Being non-conservative and by considering densities where $1 - \beta \geq 0.2$, the threshold is $10^{-3.75} \leq n_s^\star/{\rm Mpc}^{-3} \leq 10^{-3.5}$ for 2MASS, whereas for

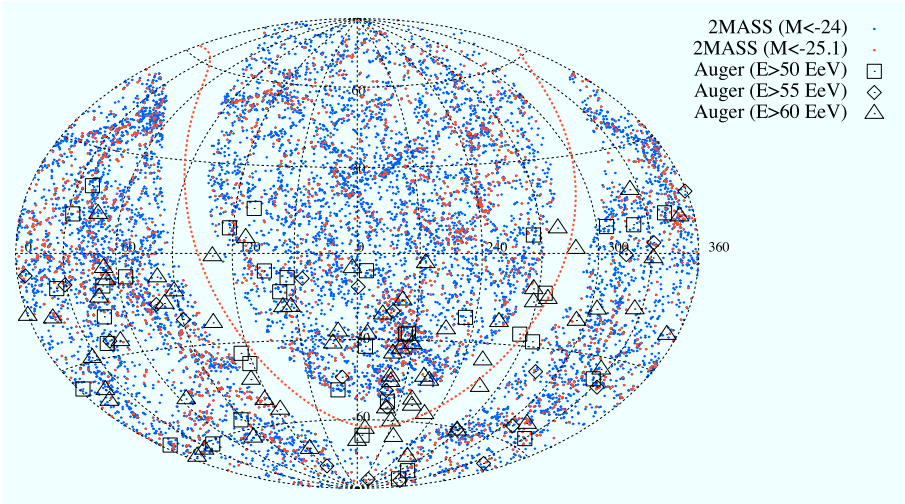


Figure 6.4: Distribution of sources for 2MASS catalog (with cuts $M < -24$ and $M < 25.1$) and Auger events for different energy thresholds.

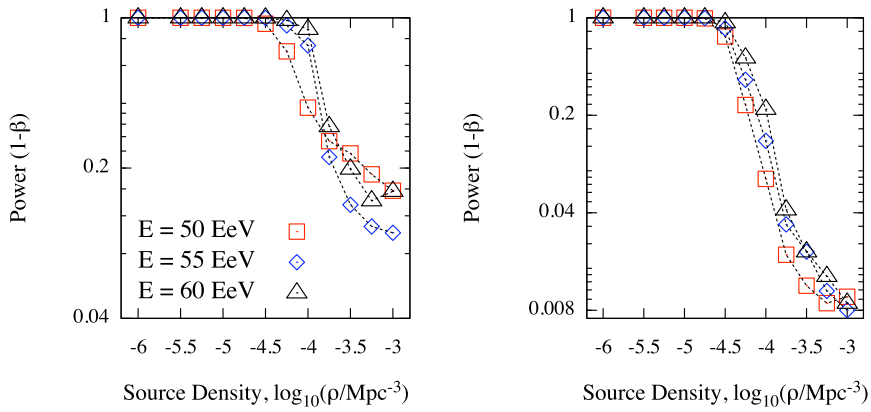


Figure 6.5: Statistical power vs source density for 2MASS (*left panel*) and ISOHOM (*right panel*) mock maps.

ISOHOM it is $10^{-4.25} \leq n_s^\star/\mathrm{Mpc}^{-3} \leq 10^{-4.0}$, depending in both cases on the energy threshold.

Source density discrimination: the "first local method"

For any catalog corresponding to a particular set of parameters in the parameters space, we build a large mock map $\mathcal{M}_m(\vec{v})$ of skies, where m = 2MASS, ISOHOM and $\vec{v}$ is a vector in the parameters space corresponding to a particular choice of parameters reported in Tab. 6.1. The size of each sky in the mock map is fixed by the number of Auger events with $E \geq E_{\mathrm{thr}}$ (and, in the case of 2MASS, falling outside the mask $|b| < 10°$). For each sky $\mathcal{S} \in \mathcal{M}_m(\vec{v})$ we estimate the amount of clustering at the scale Θ by $s(\Theta)$. Successively, we estimate the sample mean $\langle s(\Theta)\rangle$ and the 68% dispersion region around it [2]. In Fig. 6.6 and 6.7 we show for the 2MASS and ISOHOM models, respectively, the estimated value of $\langle s(\Theta)\rangle$ vs Θ for some values of the source density n_s and the three values of energy threshold investigated.

[2]We have found that the distribution of $s(\Theta)$ is not gaussian and of consequence, the estimation of the dispersion by the standard root mean square is avoided.

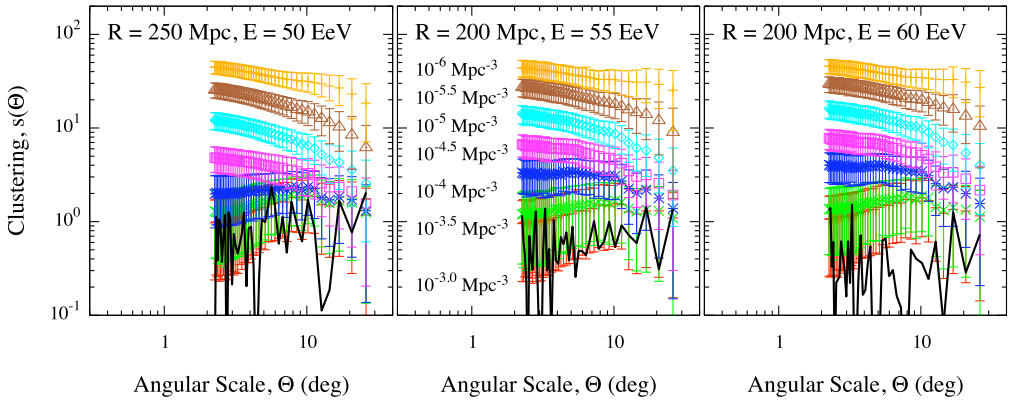


Figure 6.6: 2MASS. Average amount of clustering $\langle s(\Theta)\rangle$ vs Θ, for some values of the source density n_s and the three values of energy threshold investigated. The corresponding values obtained from the data are shown (black solid line).

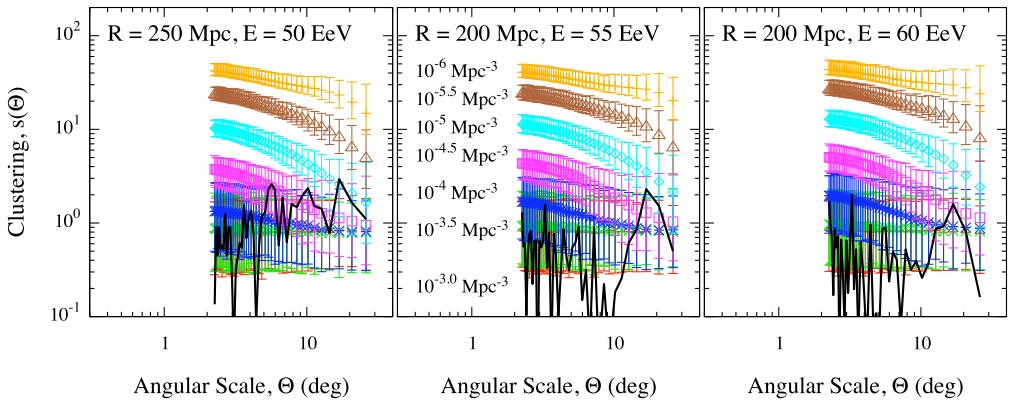


Figure 6.7: ISOHOM. Average amount of clustering $\langle s(\Theta)\rangle$ vs Θ, for some values of the source density n_s and the three values of energy threshold investigated. The corresponding values obtained from the data are shown (black solid line).

The comparison with the corresponding values obtained from the data is shown (black solid line). As expected, the amount of clustering is higher when the source density is small, whereas it is lower when the source density increases. Two main conclusions can be drawn from this simple analysis: i) curves corresponding to different values of n_s can be completely discriminated only for $\Theta < 10°$; ii) $n_s = 10^{-3}$ Mpc^{-3} represents an upper limit for the method, because the amount of clustering is compatible with that one obtained by chance for values close to this density. The second effect is less visible for the 2MASS model because of the clustering of sources at the medium angular scales, but it is clearly visible in the case of the ISOHOM model, where $s(\theta)$ is nearly constant for $n_s = 10^{-3.5}$ Mpc^{-3} and $n_s = 10^{-3}$ Mpc^{-3}, as expected for isotropic skies [278]. This result is in agreement with the result obtained more quantitatively from the statistical power. Moreover, few differences are evident among results obtained from different values of $E_{\rm thr}$: in fact, by increasing the energy threshold the average amount of clustering increases and the dispersion around it gets narrower. Another important difference emerges between the two considered models, during the comparison with the data. In fact, the MAF method is very sensible to the presence of events clustered in the Centaurus region and the exclusion of some of those points, due to the mask $|b| < 10°$ in the 2MASS analysis, produces a lower clustering signal for the data, not reflected in the ISOHOM analysis, where the mask is not considered and all Auger events are used.

The first method we apply to constrain the source density of UHECR is the same adopted in Ref.

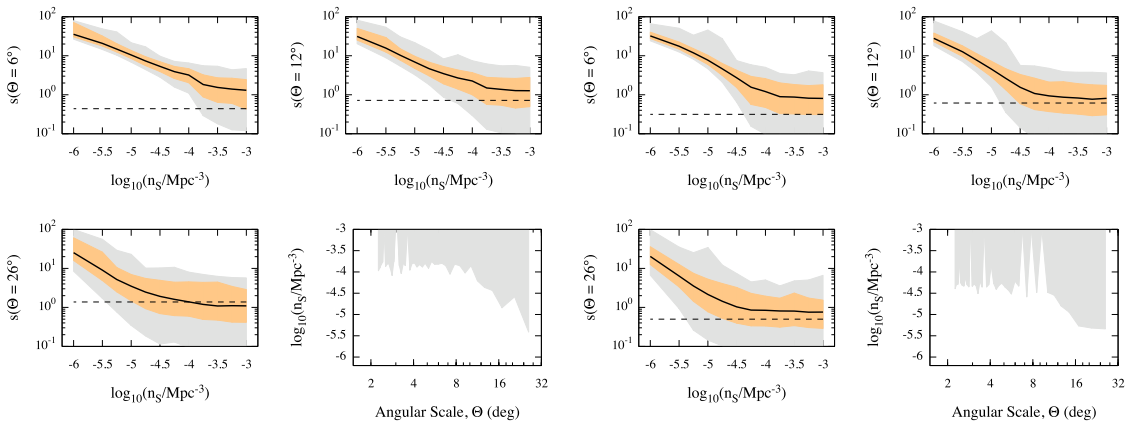


Figure 6.8: 2MASS (four panels left-hand side) and ISOHOM (four panels right-hand side) for $E_{\rm thr} = 55$ EeV. Top and left-bottom panels show the average clustering curve (solid line) with dispersion regions (68% orange, 95% gray) vs source density n_s and the value obtained from the data (dashed line) for $\Theta = 6°, 12°$ and $26°$. The right-bottom panels show the region obtained for source density bounds when 95% dispersion bands are considered, for each Θ.

[300]. It is a *local approach*. First, for each Θ separately, the mean clustering $\langle s(\Theta)\rangle$ (with 68% and 95% regions around it) is estimated for each source density: the result is an average clustering curve with dispersion bands vs source density n_s, as shown in the first three panels of Fig. 6.8 for 2MASS (left-hand side) and ISOHOM (right-hand side) models with $E_{\rm thr} = 55$ EeV and $\Theta = 6°, 12°$ and $26°$. Successively, we estimate the value of n_s where such a curve and the clustering value obtained from the data at the angular scale Θ lie across each other. The same estimation is thus employed by using the curves corresponding to the upper and the lower dispersion bands. The result is shown, for both models, in the fourth panels of Fig. 6.8: the region is obtained by using upper and lower 95% dispersion bands for the source density, for each value of the angular scale Θ separately. A discussion of the result will be provided further in the text.

Source density discrimination: the "second local method"

We propose a second *local* method to constrain source density bounds. This approach is explicitly based on hypothesis testing, for each angular scale Θ separately. Let $\mathcal{H}_0(\Theta, n_s)$ be the null hypothesis that $s_{\rm data}(\Theta)$ is a random outcome of the distribution of $s(\Theta)$ corresponding to the source density n_s. The alternative hypothesis is simply the negation of $\mathcal{H}_0(\Theta, n_s)$ and the test size α, i.e. the significance, is fixed *a priori*. $\mathcal{H}_0(\Theta, n_s)$ is rejected if the fraction of skies providing a clustering more extreme than that of data is smaller than α. In particular, when we test for the upper bound, $\mathcal{H}_0(\Theta, n_s)$ is rejected if $p(\Theta, n_s) \leq \alpha$, with

$$p(\Theta, n_s) = \mathbb{P}\left[s_{\rm data}(\Theta) \leq s_{n_s}(\Theta)\right],$$

whereas if we test for the lower bound, $\mathcal{H}_0(\Theta, n_s)$ is rejected if $p(\Theta, n_s) \leq \alpha$, with

$$p(\Theta, n_s) = \mathbb{P}\left[s_{\rm data}(\Theta) \geq s_{n_s}(\Theta)\right].$$

The lower bound to source density, at the angular scale Θ, is defined to be the lowest value of source density for which $\mathcal{H}_0(\Theta, n_s)$ has not been rejected. Analogously, the upper bound to source density is defined to be the highest value of source density for which $\mathcal{H}_0(\Theta, n_s)$ has not been rejected. It is worth remarking that, within this approach, the comparison between the distribution of $s(\Theta)$ corresponding to a particular value of the source density and the data is independently performed twice: the first time

to test the hypothesis that the density is an upper bound and the second one to test the hypothesis that the density is a lower bound.

Before discussing the results obtained from the first and the second method, we want to remark the main differences between them.

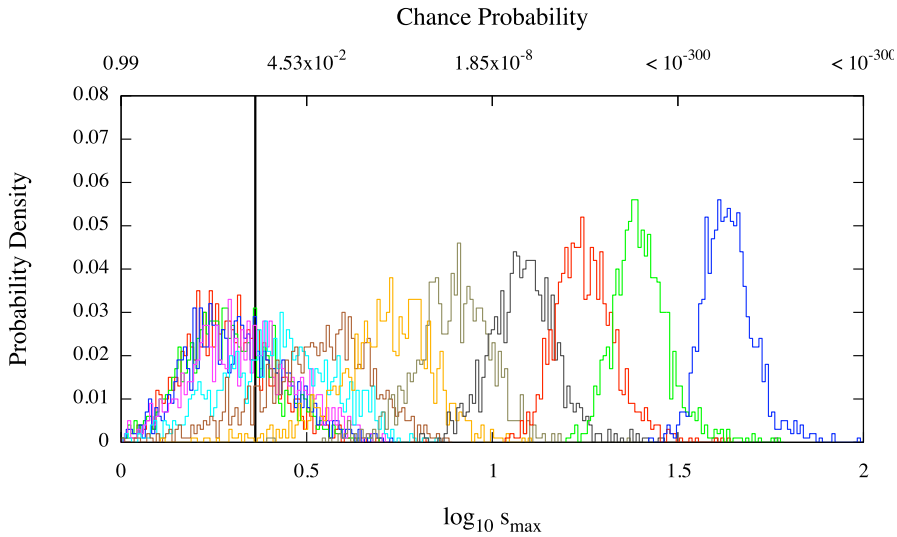


Figure 6.9: Distributions of $s^{\star}_{n_s}$ for each source density in Tab. 6.1 in the case of the ISOHOM model with $E_{\rm thr} = 55$ EeV. The solid line indicates the value $s^{\star}_{\rm data}$ obtained from the Auger data. For the highest value of the source density, $s^{\star}_{n_s}$ tends to be the lowest; viceversa, for the lowest value of the source density, $s^{\star}_{n_s}$ tends to be the highest: thus, histograms are ordered from the highest to the lowest source density from left to right. The (properly penalized) chance probability corresponding to $s_{\rm max}$ is reported in top $x-$axis for reference.

Discussion of the results from local methods

The first method provides a region delimited from a lower and an upper bound to source density: such bounds are obtained independently by estimating the value where the data cross with the 95% dispersion bands for the first time or, equivalently, where the value from the data falls into the 95% band around $\langle s_{n_s} \rangle$.

Conversely, even if the second method provides a region delimited from a lower and an upper bound to source density, such bounds are obtained independently by performing *one-sided* hypothesis testing with $\alpha = 5\%$, i.e. with 95% confidence level.

In the first method the "identity" (lower or upper) of the bound is known before performing the test, whereas in the second method there is no knowledge if the bound that is going to be tested is a lower or an upper one.

In the first case, the hypothesis tested is something like "this value of the 95% band around $\langle s_{n_s} \rangle$ for the lower (upper) bound is compatible with the data" and the test is performed only in the left (right) tail of the distribution of $s(\Theta)$, by verifying if the value from the data falls in the 95% region around the mean. In practice, two one-sided tests with 2.5% significance, corresponding to the area of one tail, are performed. When the two results are combined to give a confidence interval, the confidence level is given by the probability that the lower bound is correct (97.5%) *and* the upper bound is correct (97.5%), that is $97.5\% \times 97.5\% \approx 95\%$.

In the second case, the hypothesis tested is something like "this value of the density is a lower (upper) bound" and an one-sided test is performed with significance $\alpha_2 = 5\%$. When the two results

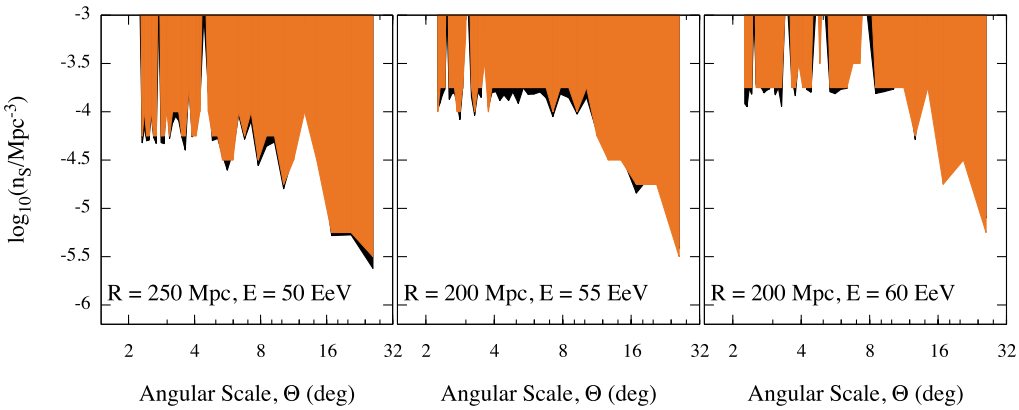


Figure 6.10: 2MASS model. Confidence regions at 95% confidence level estimated by the first (black) and the second (orange) *local* method, for each value of the angular scale Θ and of the energy threshold.

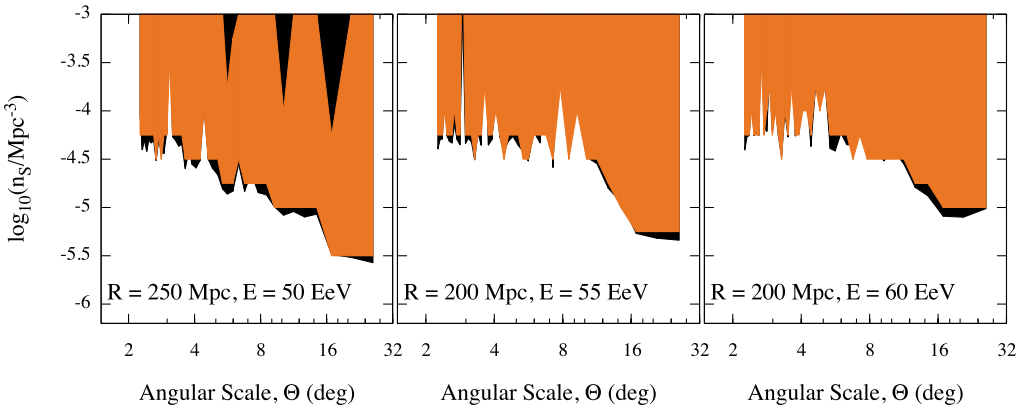


Figure 6.11: ISOHOM model. Confidence regions at 95% confidence level estimated by the first (black) and the second (orange) *local* method, for each value of the angular scale Θ and of the energy threshold.

are combined to give a confidence interval, the confidence level is $95\% \times 95\% \approx 90\%$. In order to obtain the same confidence level of the first case ($\alpha_1 \approx 5\%$), the significance of hypothesis testing in the second method should be $\alpha_2 = 1 - \sqrt{1-\alpha_1} \approx 1 - (1 - \frac{\alpha_1}{2}) \approx 2.5\%$ for each test.

For the 2MASS model, the confidence regions at 95% confidence level estimated within the first and the second method are shown superimposed in Fig. 6.10, to facilitate the comparison. In Fig. 6.11 we show the results obtained for the ISOHOM model. The two methods are equivalent under such assumptions, and we have verified that the differences in the results are only due to statistical fluctuations, because of the small number of mock skies adopted for this study (10^3).

Interestingly, for the 2MASS model the estimation of the lower bound is about $10^{-3.75}$ Mpc^{-3} ($\approx 1.8 \times 10^{-4}$ Mpc^{-3}) for almost any angular scale $\Theta < 10°$ and $E_{\rm thr} \geq 55$ EeV, with no regard for the method adopted. For $E_{\rm thr} = 50$ EeV the result is about 30% smaller, with lower bound $\approx 5.6 \times 10^{-5}$ Mpc^{-3} for $\Theta < 10°$. The ISOHOM model provides even smaller values for the lower bounds, ranging from 1.8×10^{-5} to 5.6×10^{-5} Mpc^{-3}. Upper bounds are compatible with source density $\approx 1.0 \times 10^{-3}$ Mpc^{-3}, although we can not exclude higher densities because this value represents a limit for the MAF method, where clustering is consistent with that one observed by chance.

In Fig. 6.12 we show the lower bounds $\rho_{\rm LB}$ at 95% CL on the density of sources, as a function of the angular scale, obtained for $E_{\rm thr} = 60$ EeV from both first (MAF I) and second (MAF II) methods discussed in this section, and from the autocorrelation analysis recently reported in Ref. [303]. Both

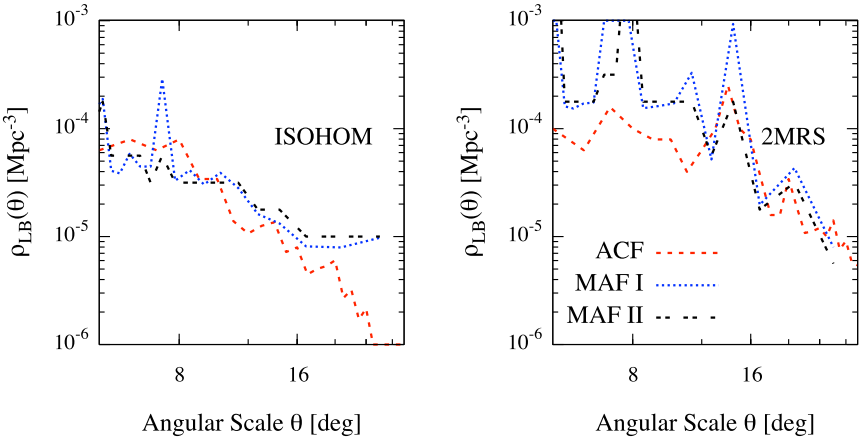


Figure 6.12: Protons with $E_{\rm thr} = 60$ EeV. Lower bounds at 95% CL on the density of sources as a function of the angular scale, obtained from both first (MAF I) and second (MAF II) methods discussed in this section, and from the autocorrelation analysis recently reported in Ref. [303] for the uniform (left) and the realistic (right) astrophysical scenarios.

uniform (ISOHOM) and realistic (2MRS) astrophysical scenarios are considered.

If the effects of intervening magnetic fields do not smooth out the clustering properties of UHECRs on scales of about 5° (as can be expected in the case of a proton composition), the measurements imply a 95%CL lower limit on the source density of order 10^{-4} Mpc^{-3}. Conversely, if magnetic deflections are larger, and such that the clustering properties observed reflect the expectation from the source scenario only at larger angular scales, then less stringent lower bounds apply. They are about one order of magnitude smaller for angular scales around 25°. Such bounds apply to specific scenarios, since they depend on the overall distribution of sources, and they are valid if the deflections of CR trajectories by intervening magnetic fields do not erase the clustering properties expected from the models at the angular scales considered.

Bounds from the MAF I and MAF II are in good agreement with each other, while an overall agreement with bounds obtained from ACF is evident. In the ISOHOM model, MAF and ACF analyses disagree at the largest angular scales: the multiscale method provides more stringent lower bounds to the density of sources with respect to the standard autocorrelation. The latter provides a more conservative estimation (i.e. smaller values of lower bounds) because, at large angular scales, the test statistics captures a clustering signal that is not present by construction. Conversely, the MAF method is characterized by a smaller rate of wrong rejections of the null hypothesis.

In the 2MRS scenario there is an overall agreement between MAF and ACF at any angular scale, although at the smaller angles, below 10°, the MAF provides again more stringent lower bounds.

Source density discrimination: the alternative "global method"

Finally, before concluding this section, we report the results obtained by a variant of the second method, a *global* approach instead of a *local* one. Our *global* method is intended to provide bounds to source density of UHECR by performing the comparison between the data and the models through maximum values $s^\star_{n_s} = \max\{s_{n_s}(\Theta)\}$. For each value of the source density n_s, we build [3] the distribution of $s^\star_{n_s}$ for testing the hypothesis that the value $s^\star_{\rm data}$, obtained from the data, is drawn from the distribution. The adopted procedure is equivalent to that one adopted for the second *local* method, with the only difference that the dependence on tha angular scale Θ vanishes. As an illustrative example, in Fig. 6.9

[3] For each sky $\mathcal{S} \in \mathcal{M}_m(\vec{v})$ a single value of $s^\star_{n_s}$ is obtained.

we show the distributions of $s^{\star}_{n_s}$ for each source density in Tab. 6.1 in the case of the ISOHOM model with $E_{\rm thr} = 55$ EeV.

Model	$E_{\rm thr}$ (EeV)	Lower / Upper (10^{-4} Mpc^{-3})
2MASS	50	0.56/10
	55	1.78/10
	60	1.78/10
ISOHOM	50	0.32/10
	55	0.32/10
	60	0.56/10

Table 6.2: Bounds to source density obtained from the *global* method for 2MASS and ISOHOM models and different values of the energy threshold.

In Tab. 6.2 we report the values obtained from the *global* method for the bounds to source density in the case of 2MASS and ISOHOM models and for different values of the energy threshold. Strikingly, although in the *global* method the information about the angular scale is lost, the results are in extraordinary agreement with those ones obtained by using *local* methods, with the evident advantage of being faster and to provide a single couple of bounds instead of many couples corresponding to each value of Θ, for each configuration in the parameters space.

In Tab. 6.3 [298] we report the densities of candidate UHECR sources: our results provide further evidence, at 95% confidence level, that AGN, GRB (gamma-ray burst) and dead quasar are suitable astrophysical object accounting for the observed clustering in Auger data.

Object	Density (Mpc^{-3})	Reference
AGN	$0.1 \div 1 \times 10^{-4}$	Loveday et al. 1992; Steffen et al. 2003
FR II radio galaxy	3×10^{-8}	Woltjer 1990
BL Lac	3×10^{-7}	Woltjer 1990
GRB	1×10^{-4}	Mao & Mo 1998
Dead quasar	5×10^{-4}	Boldt & Ghosh 1999
Colliding galaxy	7×10^{-7}	Smialkowski et al. 2002

Table 6.3: Densities of candidate UHECR sources [298].

6.1.5 Conclusion

We have considered different models of source distribution in the nearby universe (up to 250 Mpc). With HERMES, we have simulated the production of high energy protons and their propagation in a ΛCDM universe, building mock maps of events in the Auger sky. We have investigated the clustering features of such events by varying the main parameters in this study, namely the density of sources and the energy threshold of protons, by means of the novel multiscale autocorrelation function. We have constrained the source density by using three different approaches: all methods are compatible with the confidence interval $0.32 - 10 \times 10^{-4}$ Mpc^{-3} at 95% confidence level, in agreement with previous and more recent studies [303]. The given value of the upper bound is 1×10^{-3} Mpc^{-3} although we can not exclude higher densities because this value represents a limit for the MAF method, where clustering is consistent with that one observed by chance.

6.2 Bounds on source density of UHECR and EMF with MAF

The main goal of the study presented in this section is to estimate the lower (and, eventually, upper) bounds to the source density of UHECRs and to the extragalactic magnetic field (EMF), by using the multiscale method (Chap. 4 and [278]). We will consider two toy models: one following the large scale structure of matter and another one with uniformly distributed sources in the nearby universe, up to 200 Mpc. In any case, sources are assumed to be with uniform and equal luminosity, with a spectrum whose injection index is 2.7 and producing only UHE protons or irons (i.e. no mixed composition has been considered). The produced particles are successively propagated in a ΛCDM universe, taking into account non-negligible energy-loss processes in the cosmic microwave and infrared background photon fields. The effects of the irregular component of the EMF are taken into account through a smearing procedure, as well as the effect of experimental uncertainty on the angular resolution. Moreover, we have considered the cases where the irregular component of the galactic magnetic field (GMF) is also taken into account.

For details concerning the source weighting and the completing procedure for the 2MRS catalog [301], we refer to Sec. 6.1. The clustering analysis is performed by taking into account the Auger exposure. The parameter space consists of the distribution of source, the density of sources from 10^{-6} Mpc^{-3} up to 10^{-3} Mpc^{-3}, the EMF from 0.1 nG to 10 nG, the composition of UHECRs ($(Z, A) = (1, 1)$ and $(26, 56)$) and the energy threshold of particles from 60 EeV to 80 EeV. In any case, equal intrinsic luminosity for the sources is assumed. Auger data is used to constrain the source density and the EMF in different astrophysical scenarios.

6.2.1 Simulation of magnetic fields

As we saw in Chap. 2, magnetic fields can be described by two components, a regular and an irregular one. The regular component has the effect of bending the trajectories of charged particles, with the typical deflection scale provided by the Larmor radius. In this work we do not take into account the regular component. Conversely, the irregular component deals with the turbulent nature of the magnetic field on scales much smaller than the Larmor radius. Such a turbulent component is described by a r.m.s. strength B and a coherence length Λ, corresponding to the hypothesis that the power spectrum of the irregular field has a cut-off in wavenumber space at $k = 2\pi/\Lambda$ and in real space it is smooth on scales below Λ [40]. For a propagation distance greater than Λ, and in the absence of a regular component, the particle is randomly deflected with a spreading angle δ that follows the Fisher-von Mises probability distribution described in Sec. 5.1 (random walk approximation). In the case of the GMF, the spreading angle is given by

$$\delta_{\rm GMF} = 0.6^\circ Z \frac{100 \text{ EeV}}{E} \frac{B_{\rm GMF}}{4\ \mu\text{G}} \left(\frac{d}{3 \text{ kpc}}\right)^{\frac{1}{2}} \left(\frac{l_c}{50 \text{ pc}}\right)^{\frac{1}{2}} \tag{6.1}$$

where $B_{\rm GMF}$ is the strength of the irregular component of the GMF, l_c is the corresponding coherence length and d is the traversing distance of the UHECR [130]. In the case of the EMF, the smearing angle is given by [40]

$$\delta_{\rm EMF} = 0.8^\circ Z \frac{100 \text{ EeV}}{E} \frac{B_{\rm EMF}}{1 \text{ nG}} \left(\frac{D(z)}{10 \text{ Mpc}}\right)^{\frac{1}{2}} \left(\frac{L_c}{1 \text{ Mpc}}\right)^{\frac{1}{2}} \tag{6.2}$$

where $B_{\rm EMF}$ is the strength of the irregular component of the EMF, L_c is the corresponding coherence length and $D(z) \approx zc/H_0 \approx 4258.4\ z$ Mpc, if $z \ll 1$ (it is worth remarking that in our study, we have considered sources up to 200 Mpc, or, equivalently, $z \approx 0.047$).

We are interested in estimating the impact of irregular component of the EMF on the bounds of the source density and for this reason we vary its strength from 0.1 nG up to 10 nG. The coherence length is fixed to 1 Mpc. On the other hand, such an analysis for the GMF case should require a full propagation code for estimating the traversing distance of the UHECR inside our galaxy. However, we are not interested in achieving such a precision in this particular case and we use the rough approximation that the average propagation distance in the galaxy is ~ 4 kpc. Moreover, we fix the r.m.s. strength of the irregular field to 10 μG and its coherence length to 150 pc. It follows that the spreading angle becomes

$$\delta_{\rm GMF} \approx 3^\circ Z \frac{100 \text{ EeV}}{E} \tag{6.3}$$

Note. In the case of iron, the probability to get an event from a source takes into account the photo-disintegration process and the possibility that a lighter fragment may reach the Earth, with energy above the given threshold. However, when dealing with the smearing of the arrival direction, because of the turbulent magnetic fields, we will assume that the nuclei are always irons. Such an assumption is required because we do not perform a full propagation from the source to the Earth. In a more realistic scenario the average deflection should be smaller than that one considered in our study. However, we have estimated that our error has a negligible impact on the final result. For instance, for $E_{\rm thr} = 60$ EeV, the average mass of the fragment reaching the Earth is estimated to be $\langle A \rangle \approx 46$ [4], with a corresponding atomic number $\langle Z \rangle \approx 23$. The spreading angles are proportional to Z: it follows that, on average, we overestimate the deflection when considering an iron nucleus instead of the "true" fragment. Quantitatively, the overestimation factor is given by $26/\langle Z \rangle \approx 1.13$. The main consequence of such an overestimation is reflected by an underestimation of the clustering signal, due to the larger spreading angle, for angular scale within $\sim 30^\circ$. In fact, above such a threshold, the MAF is not able to discriminate the angular scale of clustering with accuracy. Moreover, such an angular scale is obtained for fragments with $Z \geq 6$, under the assumptions leading to Eq. (6.3) and for $E = 60$ EeV: it follows that we really underestimate the clustering signal in those cases where the iron nuclei have been produced by far sources (leading to an higher probability to have a fragment at Earth with $Z < 6$).

However, we have estimated that the probability, integrated over the distance, to have a fragment with $A < 12$ is smaller than $\approx 1\%$, i.e. in practice our assumption will have a negligible impact on the final results, with the great advantage of a smaller computation time required by simulations.

Experimental uncertainty

In order to take into account the experimental uncertainty on the angular resolution, we apply an additional smearing to the arrival directions of simulated events with spreading angle $\delta_\sigma \approx 0.8^\circ$, according to the resolution of the Pierre Auger Observatory at the energy considered in this study. Thus, for a given astrophysical scenario, the total smearing angle is

$$\delta = \sqrt{\delta^2_{\rm GMF}(E;Z) + \delta^2_{\rm EMF}(z,E;Z) + \delta^2_\sigma} \tag{6.4}$$

6.2.2 Clustering analysis of simulated sky maps and Auger data

We analyze the mock maps of events generated from catalogues corresponding to configurations in Tab. 6.1, for both 2MRS and ISOHOM models. We investigate the ability of MAF to discriminate among different source densities and EMFs. The comparison with results obtained from Auger data is also discussed. For the sake of completeness, we use the official data set of Auger events up to 25

[4] Such an information has been deduced from Fig. 7 in Ref. [171].

January 2011 with standard quality cuts for anisotropy studies and energy $E \geq E_{\rm thr}$. Moreover, in the case of 2MRS model only Auger events with $|b| \geq 10°$ are considered.

Source density discrimination: the "global method"

For any catalog corresponding to a particular set of parameters in the parameter space, we build a large mock map $\mathcal{M}_m(\vec{v})$ of skies, where m = 2MRS, ISOHOM and $\vec{v}$ is a vector in the parameters space corresponding to a particular choice of parameters. The size of each sky in the mock map is fixed by the number of Auger events with $E \geq E_{\rm thr}$ (and, in the case of 2MRS, falling outside the mask $|b| < 10°$). In Sec. 6.1 we have proposed a *global* method to discriminate among models. Our *global* method is intended to provide bounds to source density of UHECR, as a function of the EMF r.m.s. strength B, by performing the comparison between the data and the models through maximum values $s^\star_{n_s} = \max\{s_{n_s}(\Theta)\}$. For each value of the source density n_s and of B, we build the distribution of $s^\star_{n_s}$ for testing the hypothesis that the value $s^\star_{\rm data}$, obtained from the data, is drawn from the distribution. The adopted procedure is equivalent to that one adopted for the second *local* method in the previous section, with the only difference that the dependence on the angular scale Θ vanishes. See Sec. 6.1 for further detail.

Results obtained by assuming protons, GMF and EMF

As an illustrative example of the global method, in Fig. 6.13 we show the distributions of $s^\star_{n_s}$ for each probed source density in the case of the ISOHOM model. Protons with $E_{\rm thr}$ = 60 EeV are used and the smearing due to both the GMF and EMF is considered. In Fig. 6.14 we show the distributions obtained under the same assumptions, but in the case of the 2MRS model. It is evident that increasing the irregular component EMF has a small effect on the distribution of the clustering coefficient, when protons are considered. However, this result is not surprising. In fact, only the closest sources are significantly responsible of the clustering signal in the arrival direction distribution of UHECRs, because of the energy-loss processes. Moreover, the spreading angle due to the deflections in the irregular GMF is always the same ($\sim 5°$) and the deflections due to the irregular EMF (e.g. for a propagation distance $D \approx 10$ Mpc) span from 0.1° to 10°. Over such a range of angular scales, the MAF is very sensitive to the clustering signal, that does not significantly change its strength but only its location [278].

In Fig. 6.15 we show the results obtained for the lower bounds (99% CL) on the source density, as a function of the irregular EMF r.m.s. strength, for both ISOHOM (left panel) and 2MRS (right panel) models, in the case of protons with two different energy thresholds (60 and 80 EeV). The cases where the GMF and the EMF are both considered, and where only the EMF is taken into account, are shown.

It is evident that lower bounds do not significantly depend on B in any configuration considered here. In all cases, the values of the bounds on the density decrease for increasing spreading angles. In fact, by fixing an energy threshold, limits are the smallest in the GMF+EMF case (maximum spreading), followed by the EMF case (medium spreading) and the null magnetic field case (no spreading). On the other hand, for a fixed configuration of the magnetic field, limits decrease for increasing energy thresholds, reflecting the fact that the clustering signal is higher for smaller values of the energy, where the number of events (and, of course, the statistical power) is higher. Results are the same for both ISOHOM and 2MRS models, even if bounds on the source density are higher of a factor ~ 2 for the latter.

Results obtained by assuming protons, irons and EMF

In Fig. 6.16 we show the maximum clustering $s^\star_{n_s}$ as a function of the source density and the strength of irregular component of the EMF (the GMF is not considered here). UHECRs are protons (left panel)

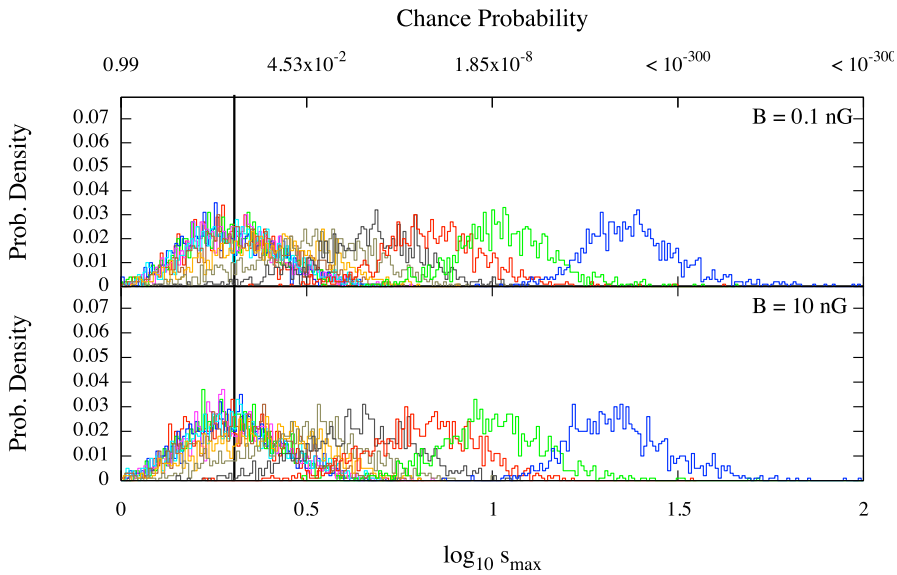


Figure 6.13: Distributions of $s^{\star}_{n_s}$ for each probed source density in the case of the ISOHOM model with $E_{\rm thr} = 60$ EeV. UHECRs are protons and both GMF and EMF are considered. The solid line indicates the value $s^{\star}_{\rm data}$ obtained from the Auger data. For the highest value of the source density, $s^{\star}_{n_s}$ tends to be the lowest; viceversa, for the lowest value of the source density, $s^{\star}_{n_s}$ tends to be the highest: thus, histograms are ordered from the highest to the lowest source density from left to right. The (properly penalized) chance probability corresponding to $s_{\rm max}$ is reported in top $x-$axis for reference. The upper panel is for $B = 0.1$ nG, whereas in the lower panel $B = 10$ nG.

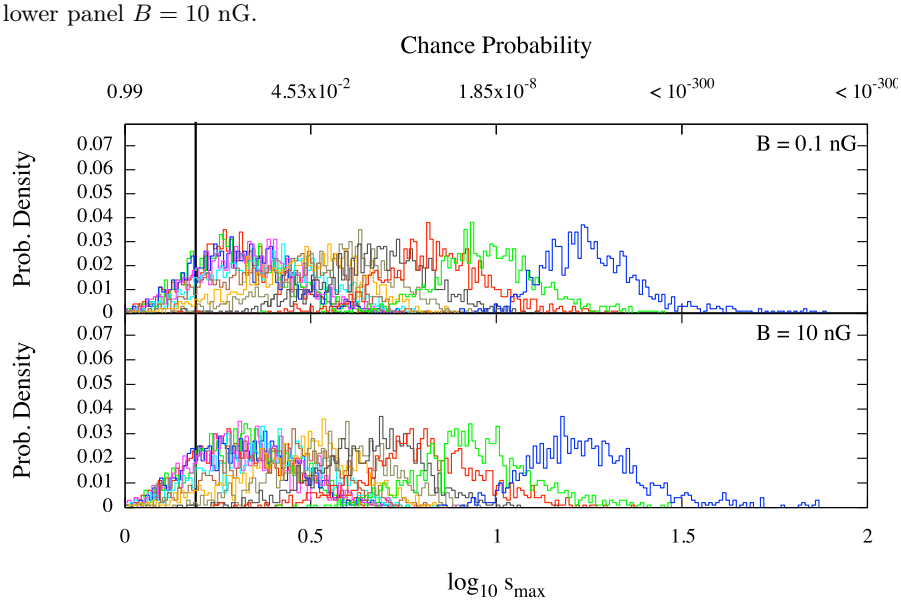


Figure 6.14: Same as in Fig. 6.13 but assuming the 2MRS model for the distribution of sources.

and irons (right panel) with $E_{\rm thr} = 60$ EeV, produced within the 2MRS model. It is evident that there is a non-significant dependence on B only in the case of protons.

In Fig. 6.17 we show the distributions of $s^{\star}_{n_s}$ for each probed source density in the case of the ISOHOM model. Protons with $E_{\rm thr} = 60$ EeV are used and only the smearing due to EMF is considered. The main effect of removing the GMF is a higher clustering signal and a more precise discrimination among different values of the source density. Again, the effect of increasing the strength of the irregular EMF is small and does not influence the estimation of the lower bound. In Fig. 6.18 we show the distributions

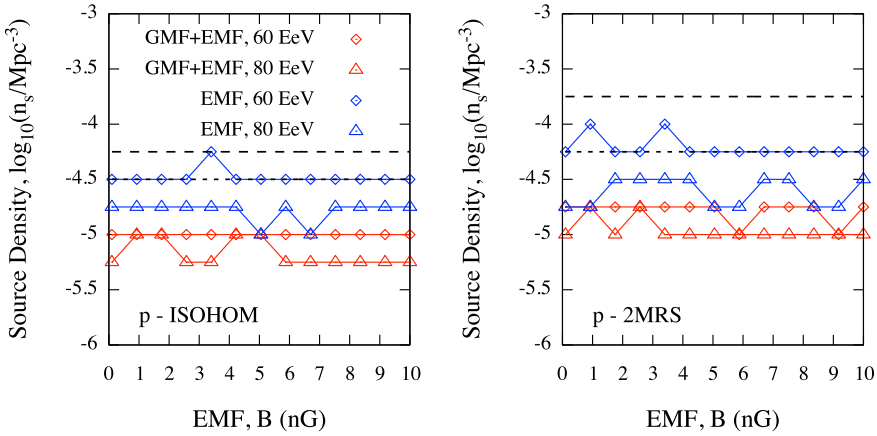


Figure 6.15: 99% CL lower bounds on the source density as a function of the irregular component of the EMF r.m.s. strength, for ISOHOM (left panel) and 2MRS (right panel) models. UHECRs are protons with $E_{\rm thr} = 60$ and 80 EeV. Dashed and short-dashed lines correspond to the estimated values in the limit of null magnetic field and $E_{\rm thr} = 60$ and 80 EeV, respectively.

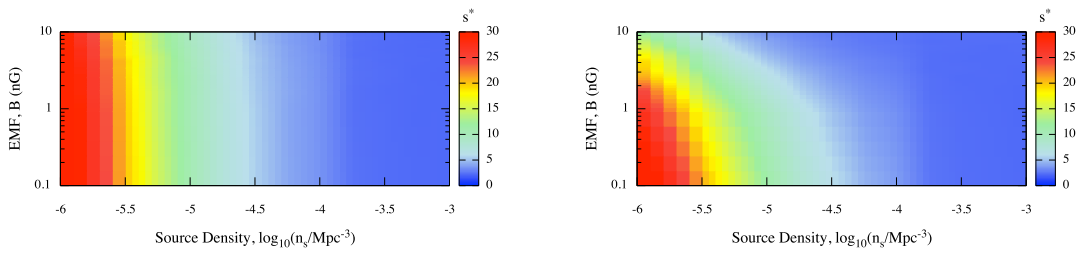


Figure 6.16: Maximum clustering $s^\star_{n_s}$ as a function of the source density and the strength of irregular component of the EMF. UHECRs are protons (left panel) and irons (right panel) with $E_{\rm thr} = 60$ EeV, produced within the 2MRS model.

obtained under the same assumptions, but in the case of irons. At variance with the case of protons, a more significant difference between distributions corresponding to the highest source density is present, and the clustering signal is reduced.

Such differences are more evident in Fig. 6.19, where results obtained for the lower bounds (99% CL) on the source density, as a function of the irregular EMF r.m.s. strength, are shown for both ISOHOM and 2MRS models, in the case of protons (left panel) and of irons (right panel), with energy threshold equal to 60 EeV.

Results are compatible with those ones reported in Sec. 6.1, with bounds of the order of $\sim 10^{-4}$ Mpc^{-3} for softer EMFs (corresponding to deflections smaller than 15°) and of the order of $\sim 10^{-5}$ Mpc^{-3} (or lower) for stronger EMFs.

In Tab. 6.3 [298] we report the densities of candidate UHECR sources: our results provide further evidence, at 99% confidence level, that AGN, GRB (gamma-ray burst) and dead quasar are suitable astrophysical objects accounting for the observed clustering in Auger data, even if mass composition and magnetic fields are taken into account. Even in the more extreme cases, candidate sources with density smaller than 10^{-6} Mpc^{-3} are excluded, unless very exotic scenarios get involved. However, it is worth remarking that we have obtained our results after making several assumptions on the magnetic fields involved, and it should be considered as a naive exercise for exploring more complicated astrophysical scenarios than those investigated in our previous studies, where only light composition and no magnetic

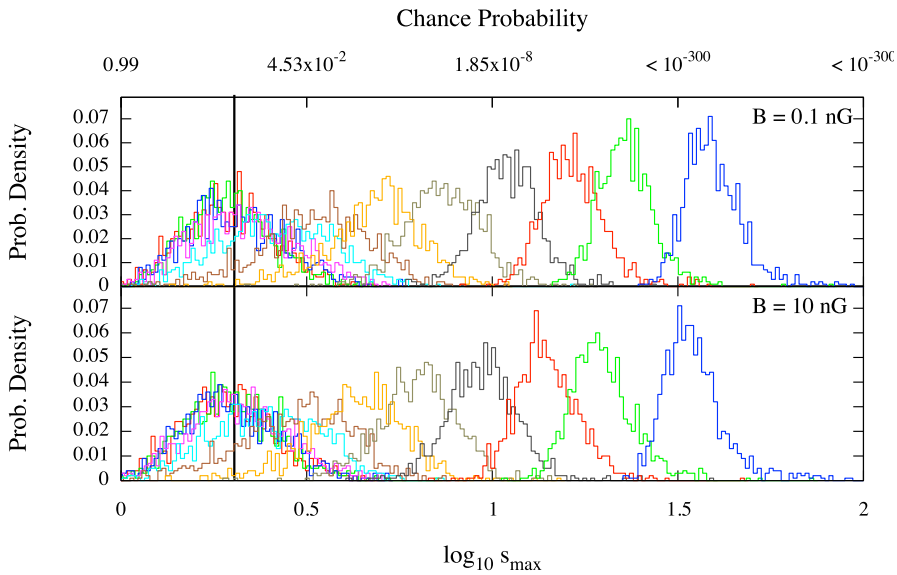


Figure 6.17: Same as in Fig. 6.13, but excluding the smearing due to the GMF. UHECRs are protons with $E_{\rm thr} = 60$ EeV.

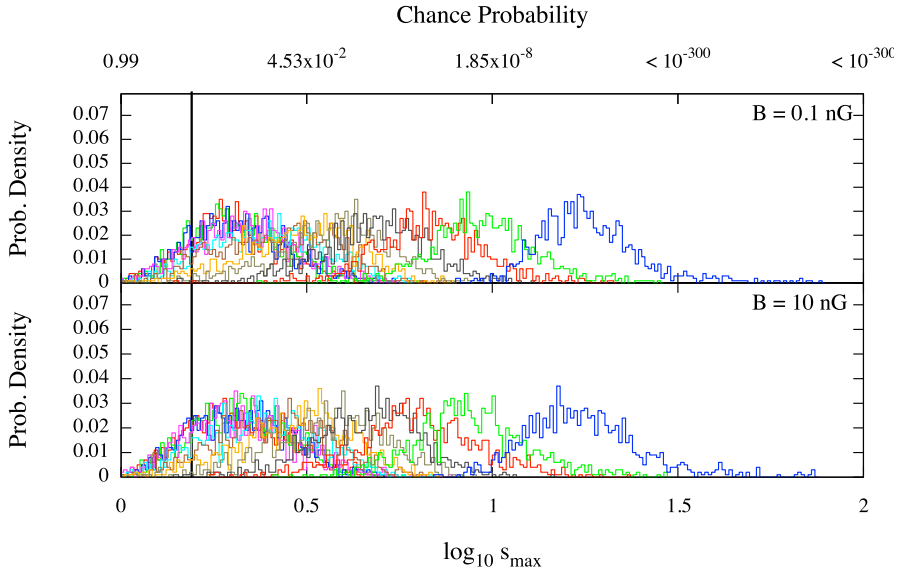


Figure 6.18: Same as in Fig. 6.17 but assuming that UHECRs are iron nuclei in the 2MRS scenario.

fields are considered.

6.2.3 Conclusion

We have considered different models of source distribution in the nearby universe (up to 200 Mpc) and we have simulated the production of high energy protons and irons and their propagation in a ΛCDM cosmology. The deflections due to different configurations of galactic and extragalactic magnetic fields (only the turbulent component) have been taken into account, and mock maps of events in the Auger sky have been built by varying the values in the parameter space.

Successively, we have investigated the clustering features of such events, for different density of sources, magnetic field strength and energy threshold of UHECRs, by means of the novel multiscale autocorrelation function and by including the effects of the experimental uncertainty on the angular

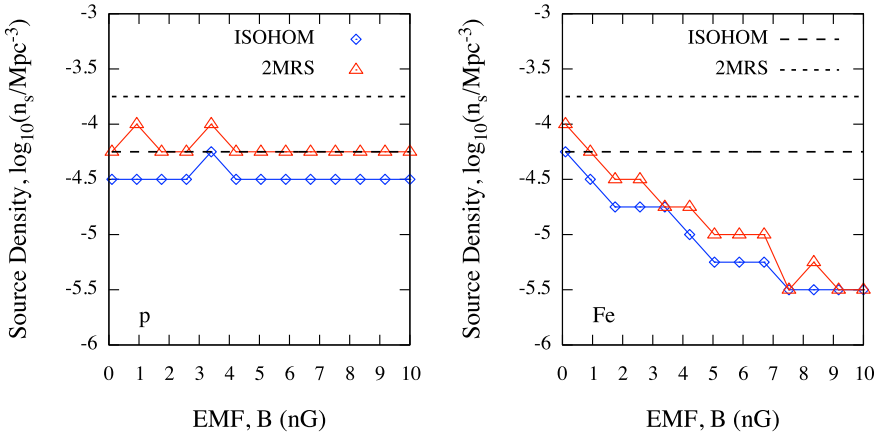


Figure 6.19: 99% CL lower bounds on the source density as a function of the irregular component of the EMF r.m.s. strength, for ISOHOM and 2MRS models. UHECRs are protons (left panel) and irons (right panel) with $E_{\rm thr} = 60$ EeV. Dashed and short-dashed lines correspond to the estimated values in the limit of null magnetic field for ISOHOM and 2MRS models, respectively.

resolution. Thus, we have investigated the dependence on the EMF of the source density, providing lower bounds for several different astrophysical scenarios at 99% CL.

We have found that the main effect of including an explicit treatment of deflections due to the magnetic field is to decrease the lower bound on the density of sources. Such results are in agreement with those ones reported in the previous section and in Ref. [303]. By increasing the strength of the field, deflection angles get larger further reducing the clustering signal. Moreover, we have found that such bounds decrease for increasing energy thresholds of UHECRs. For all configurations investigated within this study, we have observed that bounds increase if the distribution of sources is not isotropic.

Finally, for any given set of parameters we have found that lower bounds on the source density do not depend on the strength of the magnetic field if the composition is light, i.e. for protons. Conversely, for heavier nuclei lower bounds decrease in the case of stronger magnetic fields, with differences of more than one order of magnitude between the limit of a null field and a r.m.s. strength of the irregular component of the EMF of ~ 10 nG.

The provided values of the upper bounds are 1×10^{-3} Mpc^{-3}, although we can not exclude higher densities because this value represents a limit for the MAF method, where clustering is consistent with that one observed by chance. Our results suggest that, even in the more extreme cases, candidate sources of UHECRs with density smaller than 10^{-6} Mpc^{-3} are excluded, unless very exotic scenarios get involved.

6.3 Can nearby black holes be sources of UHE protons?

In the previous sections, bounds on the source density of UHECRs in the nearby universe (up to 250 Mpc) have been estimated by investigating the clustering signal in their arrival direction distribution, in absence or in presence of the extragalactic magnetic field. The lower bound on the source density has been constrained by using three different approaches: all methods are compatible with the confidence interval $0.32 - 10 \times 10^{-4}$ Mpc^{-3} at 95% CL, in agreement with previous studies in literature. The given value of the upper bound is 1×10^{-3} Mpc^{-3} although higher densities can not be excluded because this value represents a limit for the multiscale autocorrelation function (MAF) method [278], where clustering is consistent with that one observed for an isotropic sky.

As we will see in the following, the density of black holes and their luminosity are a function of their

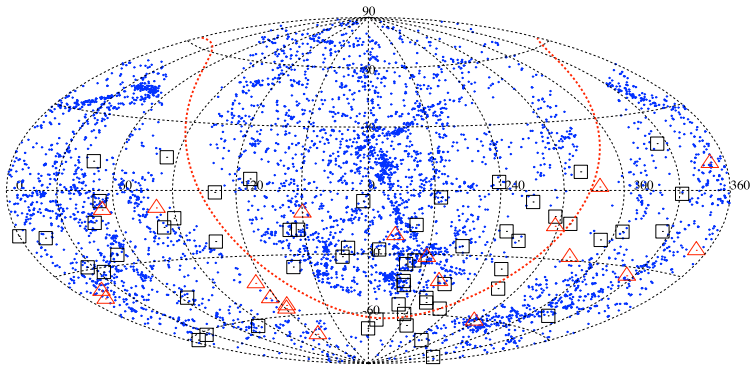


Figure 6.20: Skymap of massive black holes within ≈ 106 Mpc [304, 305] and 78 UHECRs detected by the Pierre Auger Observatory from 1 January 2004 to 9 September 2011 (blue squares indicate events with $60 \leq E/\text{EeV} < 80$, whereas red triangles indicate events with $E \geq 80$ EeV). Aitoff projection of equatorial coordinates is shown.

mass, under some astrophysical assumptions. The main goal of the present study is to estimate the lower (and, eventually, upper) bounds to the density of black holes, under the hypothesis that they are the sources of UHECRs, and to discuss the dependence of such bounds on the extragalactic magnetic field (EMF), by using the MAF. We consider the whole catalog of black holes (within ≈106 Mpc) that have been recently reported by Caramete and Biermann [304, 305]. One of the most interesting features of such a catalog, among others, is its completeness in the whole sky, which makes it suitable for clustering studies related to an underlying distribution of sources.

The intrinsic luminosity of sources is taken into account and a power-law injection of protons with spectral index 2.4 is considered. The produced particles are successively propagated in a ΛCDM universe, taking into account non-negligible energy-loss processes in the cosmic microwave background (CMB) photon field. The effects of the irregular component of the EMF are taken into account through a smearing procedure, as well as the effect of experimental uncertainty on the angular resolution.

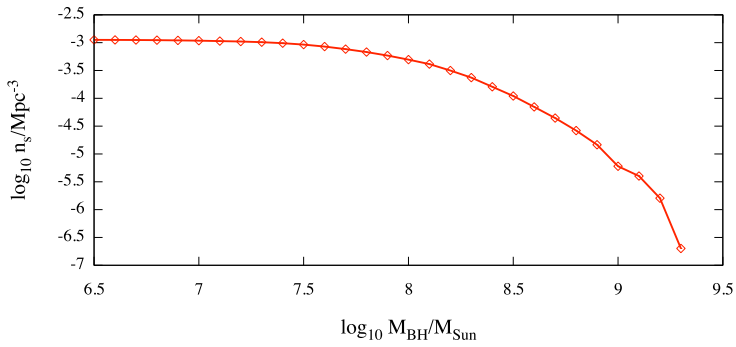
Auger data is used to constrain the black hole density and the EMF in different astrophysical scenarios, and the clustering analysis is performed by taking into account the Auger exposure. The parameter space consists of the black hole mass, ranging from $10^{6.5} M_\odot$ to $10^{9.25} M_\odot$, the r.m.s. strength of the EMF, ranging from 0 nG to 10 nG, and the energy threshold of particles at Earth, namely 60 EeV and 80 EeV. The simulation of propagating protons has been described in detail in previous chapters and will not be discussed in the study.

In the analysis presented here we adopt two different approaches for the investigation of clustering: a *global* method, making use of the maximum clustering signal as a test statistic, and a *local* method, making use of the clustering signal at any angular scale.

6.3.1 Data selection

The catalog of black holes adopted for our study has been recently shown in Refs. [304, 305] and represents an unbiased sample of about 6000 black holes within the local Universe (∼ 106 Mpc). The catalog has been built by starting from the 2 Micron All Sky Survey (2MASS) and it is complete if the region within 10° along the Galactic plane is excluded. In our study, we consider black holes with mass ranging from $10^{6.5} M_\odot$ to $10^{9.25} M_\odot$.

We use 78 UHECR events with $E \geq 60$ EeV detected by the Pierre Auger Observatory from 1 January 2004 to 9 September 2011, selected with standard quality cuts for anisotropy studies. In Fig. 6.20 we show the skymap of massive black holes in the catalog together with UHECR events used in our study. UHECRs in the region $|b| < 10°$ are excluded from the present study: the number of events surviving



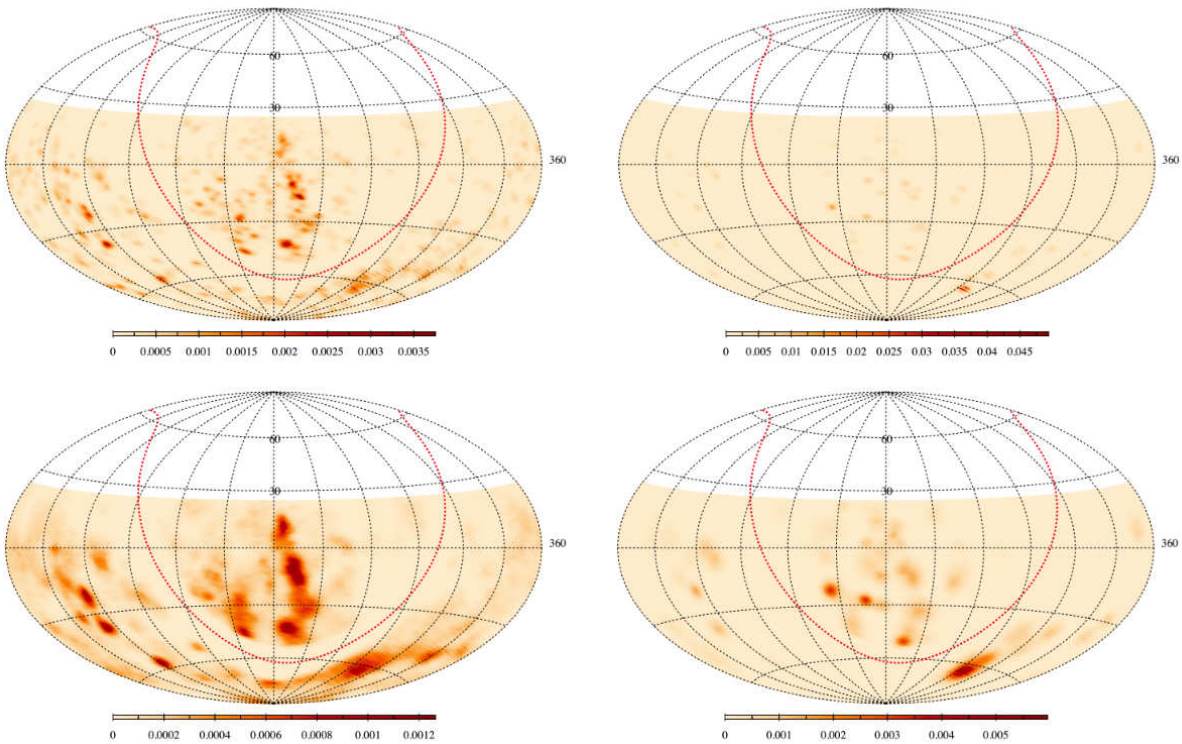


Figure 6.22: Density of 10^5 simulated protons emitted from nearby black holes with mass $M \geq M_{\rm BH}$ and propagated in the extragalactic magnetic field with r.m.s. strength $B_{\rm EMF}$. **Top:** Cases with no magnetic field, $M_{\rm BH} = 10^7 M_\odot$ (left) and $M_{\rm BH} = 10^{8.5} M_\odot$ (right). **Bottom:** Same of top panels but with $B_{\rm EMF} = 1$ nG. In any case, the angular uncertainty of the Pierre Auger Observatory is taken into account through a smearing of $\delta_\sigma = 0.8°$.

to such a cut is 59 for $E \geq 60$ EeV (squares) and 18 for $E \geq 80$ EeV (triangles).

In Fig. 6.21 we show the dependence of black holes on the mass below a given threshold $M_{\rm BH}$ and their number density n_s. Thus, by estimating the lower (upper) bounds on the density, we indirectly constrain the mass of black holes, whose distribution in the sky should account for the arrival direction distribution of UHECRs observed at Earth.

6.3.2 Simulation of UHECRs

Sources and protons

As seen in Chap. 2 and 3, UHECRs with injection energy E_i, are subjected to several effects during their propagation and they are detected with a degraded energy $E_f < E_i$, depending on the interactions they experienced and on the distance of the source. Thus, the probability to get an event from a source at redshift z is proportional to $\mathcal{L}z^{-2}\omega(z, E_{\rm thr})$ (defined by Eq. 3.60), being $\mathcal{L}$ the intrinsic luminosity of the source.

The available channels for the interactions depend on the particle and on the background radiation. In this analysis we will consider only protons. If ϵ is the energy of the ambient photon field and E is the energy of an UHECR in the observer's rest frame, the energy ϵ' of the photon in the proton rest frame is given by $\epsilon' = 2\Gamma\epsilon$, if head-on collisions are assumed. Here, $\Gamma = E/m_p c^2$ is the Lorentz factor of the proton. In the case of proton the available channels above the threshold for single pion production ($\epsilon' \approx 145$ MeV) involve baryonic resonances and direct particle production, with multi-pion production playing a significant role at the highest energies ($\epsilon' > 700$ MeV). The energy losses due to the pair production and to the adiabatic expansion occur in any case, with significant contributions only in a small range of energies. For the energies of interest for the present study, only the cosmic microwave background is relevant for protons.

The probability that an UHECR, with injection energy greater or equal than E_i, reaches the Earth with energy $E_f \geq E_{\rm thr}$, after taking into account the effects due to the energy-loss processes and the power-law injection spectrum at the source, is given by

$$\omega(z, E_{\rm thr}) = \frac{s-1}{E_{\rm thr}^{-s+1}} \int_{E_i(z,E_{\rm thr})}^{\infty} E^{-s} dE, \tag{6.5}$$

where s is the injection index. Cosmology is implicitly included in the term $E_i\,(z, E_{\rm thr})$ (see Sec. 3.6): in the following we assume the updated ΛCDM model parameters ($\Omega_b = 0.0456$, $\Omega_c = 0.227$, $\Omega_\Lambda = 0.728$ and $H_0 = 70.4$ km/s^{-1}/Mpc) [141] and $s = 2.4$.

Thus, the probability to get an event from a source at redshift z is proportional to $z^{-2}\mathcal{L}(z)\omega(z, E_{\rm thr})$, where $\mathcal{L}(z)$ indicates the evolved luminosity of the black hole. In the present study, we consider the case of no evolution ($\mathcal{L}(z) = \mathcal{L}_0(M_{\rm BH})$) and a luminosity which depends on the black hole mass. In fact, an astrophysical object of mass M which is powered by spherical accretion provides at most a luminosity $\mathcal{L}_{\rm Edd} = 1.26 \times 10^{38}(M/M_\odot)$ erg s^{-1}, i.e. the *Eddington luminosity.* If sources in our catalog are considered in the Blandford-Znajek mode, providing a mechanism for the extraction of energy from a rotating black hole [306], the resulting luminosity will be

$$\mathcal{L}_{\rm BH}(M_{\rm BH}) \propto 10^{-2.5}\mathcal{L}_{\rm Edd} \propto \frac{M_{\rm BH}}{M_\odot}. \tag{6.6}$$

Magnetic fields

The simulation of magnetic fields follows the same procedure described in Sec. 6.2. Here, the EMF is assumed to be dominated by an irregular component characterized by a r.m.s. strength B and a coherence length ℓ.

We are interested in estimating the impact of deflections on our study and for this reason we vary the r.m.s. strength of the EMF from 0 nG up to 10 nG, while keeping fixed L_c to 1 Mpc. Moreover, in order to take into account the experimental uncertainty on the angular resolution, we apply an additional smearing to the arrival directions of simulated events with spreading angle $\delta_\sigma \approx 0.8°$, according to

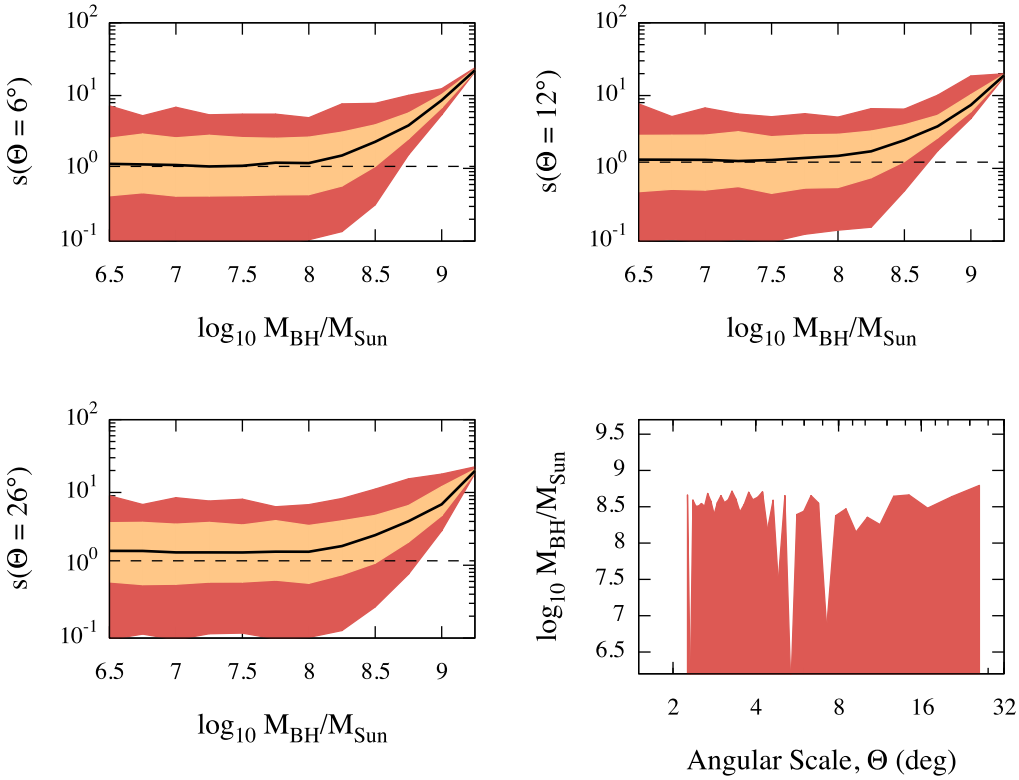


Figure 6.23: **Top and bottom-left panels:** Average clustering curve (solid line) with 68% and 95% dispersion bands obtained from simulations in the absence of the extragalactic magnetic field, as a function of black hole mass $M_{\rm BH}$ and angular scale Θ. The dashed line indicates the clustering measured in the data, in the case of $E_{\rm thr} = 80$ EeV. **Bottom-right panel:** 95% lower and upper bounds on $M_{\rm BH}$ as a function of the angular scale.

the resolution of the Pierre Auger Observatory at the energy considered in this study. Thus, the total expected deflection angle is given by

$$\delta(E, z) = \sqrt{\delta^2_{\rm EMF}(E, z) + \delta^2_\sigma}. \tag{6.7}$$

In Fig. 6.22 we show the density skymap for four realizations of 10^5 simulated protons emitted from nearby black holes with mass $M \geq M_{\rm BH}$, with $M_{\rm BH} = 10^7 M_\odot$ (left panels) and $M_{\rm BH} = 10^{8.5} M_\odot$ (right panels). Protons are then propagated by switching off (top panels) and switching on (bottom panels) the EMF.

Astrophysical assumptions

We have discussed above the simulation setup adopted for the present study. However, it is worthwhile concluding this part by remarking some limitations of this work, before discussing the analysis and the results. We are assuming i) a pure composition of protons propagating in a turbulent EMF and ii) that black holes in the catalog are the sources of such UHE particles. The implicit assumptions behind such considerations are numerous.

In fact, the distribution of matter in the limited volume of space considered in our study ($\approx 10^6$ Mpc^3) should have an impact on the intergalactic magnetic field, which in turns is expected to be structured inside and around clusters or groups of galaxies, with filaments extending over few Mpc, as shown, for instance, in recent detailed simulations [107]. The topology of such a structured magnetic field has a

non-negligible impact on the trajectories of UHE protons and, of course, on the average deflections they experience in the case of EMF with regular structures above 200 kpc [202]. If UHECRs are heavy nuclei, the impact of such regular field on their propagation could be even more dramatic.

Additionally, longitude-averaged X-ray emission observed with ROSAT near 0.65 keV and 0.85 keV towards the center of the Galaxy, are in agreement with a Galactic wind thermally-driven by cosmic rays and hot gas [203, 204], whose irregular magnetic field can also explain the isotropy of UHECR below 60 EeV observed with the Pierre Auger Observatory [89, 208]. In our study, we can neglect the effect of such a magnetic wind only in the case of protons above 60 EeV, because it is expected to have a non-negligible impact for protons below 60 EeV or heavier nuclei. Some of such limitations have been recently discussed, for instance, in Ref. [307].

It is evident that an explicit treatment of the mass composition would considerably complicate the simulations, and we reserve the exploration of such a case, together with the treatment of scenarios involving structured magnetic fields, for future works.

6.3.3 Clustering analysis of simulated sky maps and Auger data

In this section we analyze the mock maps of events generated from the catalog of black holes, as previously discussed. First, we briefly discuss the global and local methods used for clustering analysis (see Sec. 6.1 for further detail.). Successively, we investigate the ability of MAF to discriminate among different source densities (or black hole masses) and EMFs. The comparison with results obtained from Auger data is also discussed. In the following, we will simply refer to the black hole mass threshold $M_{\rm BH}$ to indicate all black holes with mass $M \geq M_{\rm BH}$, where not otherwise specified. It is worth remarking that we will use equivalently the terms "black hole mass" and "source density" because of relation in Fig. 6.21, in order to indicate the number density of black holes with mass $M \geq M_{\rm BH}$.

Description of the local method adopted for the analysis

The *local method* we apply to constrain the source density of UHECR is the same adopted in Ref. [300, 308]. First, for each angular scale Θ separately, the mean clustering $\langle s(\Theta)\rangle$ (with 68% and 95% regions around it) is estimated for each black hole mass (or, equivalently, for each value of the source density): the result is an average clustering curve with dispersion bands vs $M_{\rm BH}$, as shown in the first three panels of Fig. 6.23, in the case $B_{\rm EMF} = 0$ nG, $E_{\rm thr} = 80$ EeV, $\Theta = 6°, 12°$ and $26°$. Successively, we estimate the value of $M_{\rm BH}$ where such a curve and the clustering value obtained from the data at the angular scale Θ lie across each other. The same estimation is thus employed by using the curves corresponding to the upper and the lower dispersion bands. The result is shown in the fourth panel of Fig. 6.23: the region is obtained by using upper and lower 95% dispersion bands for the black hole mass, for each value of the angular scale Θ separately. In Fig. 6.24 we show curves as in Fig. 6.23 but for different magnetic fields intensity, at the angular scale $\Theta = 6°$, for both $E_{\rm thr} = 60$ EeV (top panels) and 80 EeV (bottom panels). At 60 EeV, the data is not compatible at 95% CL with simulation for $B_{\rm EMF} \leq 2.5$ nG, whereas it is only partially compatible for increasing EMF strength. At 80 EeV, the data is in good agreement with simulation with almost no regard for the EMF. A discussion about the lower bounds on $M_{\rm BH}$ as a function of the EMF will be provided further in the text.

Description of the global method adopted for the analysis

For any catalog corresponding to a particular set of parameters in the parameter space, we build a large mock map $\mathcal{M}(\vec{v})$ of skies, where $\vec{v}$ is a vector in the parameter space corresponding to a particular choice of parameters. The size of each sky in the mock map is fixed by the number of Auger events with $E \geq E_{\rm thr}$ and $|b| \geq 10°$. In Ref. [308] we have proposed a *global* method to discriminate among models.

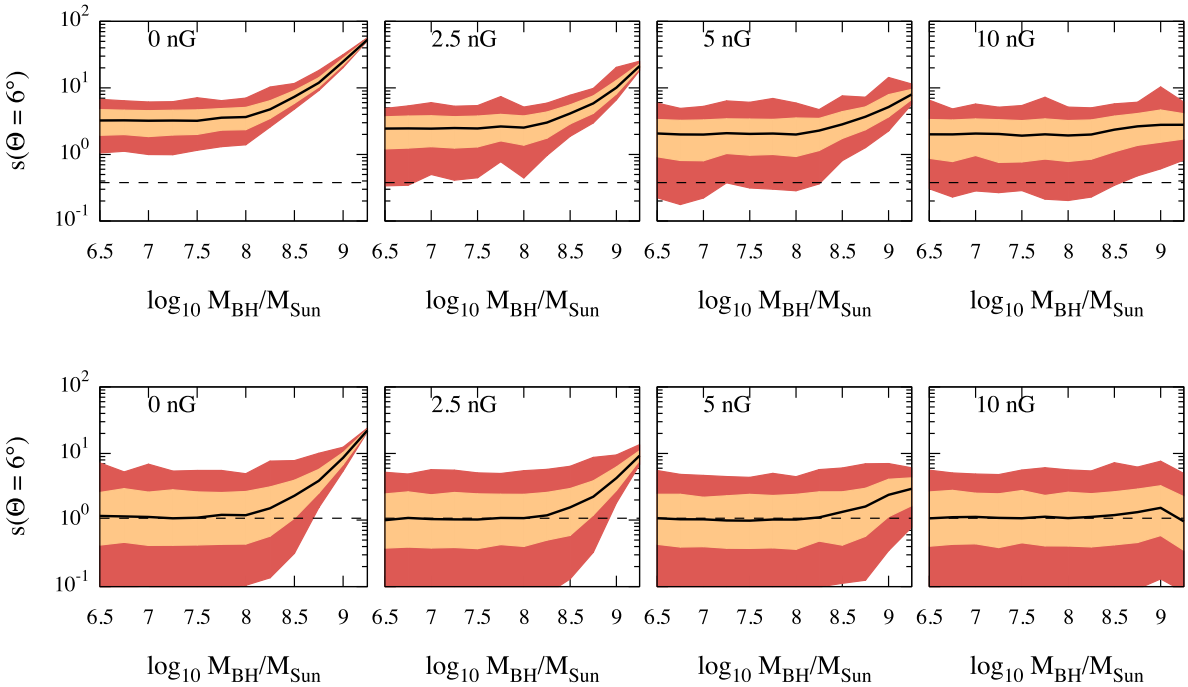


Figure 6.24: Average clustering curve (solid line) with 68% and 95% dispersion bands obtained from simulations at an angular scale $\Theta = 6^\circ$ and for different magnetic fields intensity, as a function of black hole mass $M_{\rm BH}$. The dashed line indicates the clustering measured in the data. **Top:** $E_{\rm thr} = 60$ EeV. **Bottom:** $E_{\rm thr} = 80$ EeV.

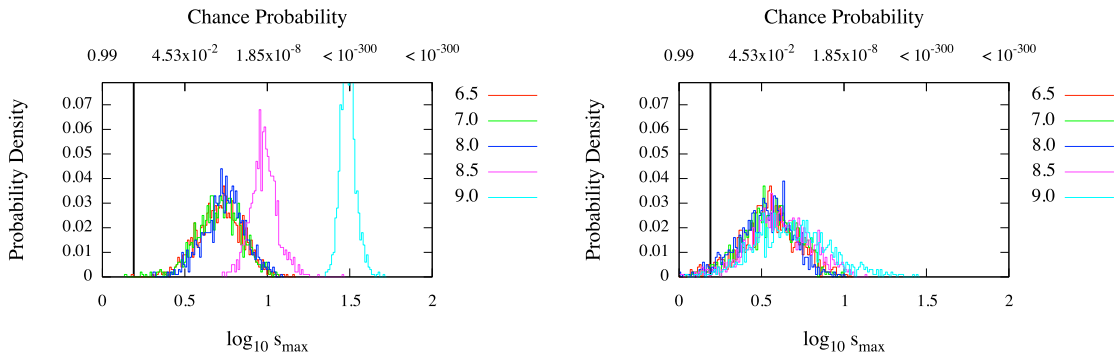


Figure 6.25: Distributions of $s^\star_{M_{BH}}$ for each probed black hole mass $M_{\rm BH}$, in the case of protons with $E_{\rm thr} = 60$ EeV propagating in the absence of magnetic field (left) and in EMF with strength 10 nG (right). The vertical solid line indicates the value $s^\star_{\rm data}$ obtained from the Auger data. For the highest value of the source density, $s^\star_{M_{BH}}$ tends to be the lowest; viceversa, for the lowest value of the source density, $s^\star_{M_{BH}}$ tends to be the highest: thus, histograms are ordered from the highest to the lowest source density from left to right. The (properly penalized) chance probability corresponding to $s_{\rm max}$ is reported in top $x-$axis for reference.

Such a method is intended to provide bounds to source density of UHECR, as a function of the r.m.s. strength $B_{\rm EMF}$ of the EMF, by performing the comparison between the data and the models through maximum values $s^\star_{M_{\rm BH}} = \max\{s_{M_{\rm BH}}(\Theta)\}$. For each value of the black hole mass $M_{\rm BH}$ and of $B_{\rm EMF}$, we build the distribution of $s^\star_{M_{\rm BH}}$ for testing the hypothesis that the value $s^\star_{\rm data}$, obtained from the data, is drawn from the distribution.

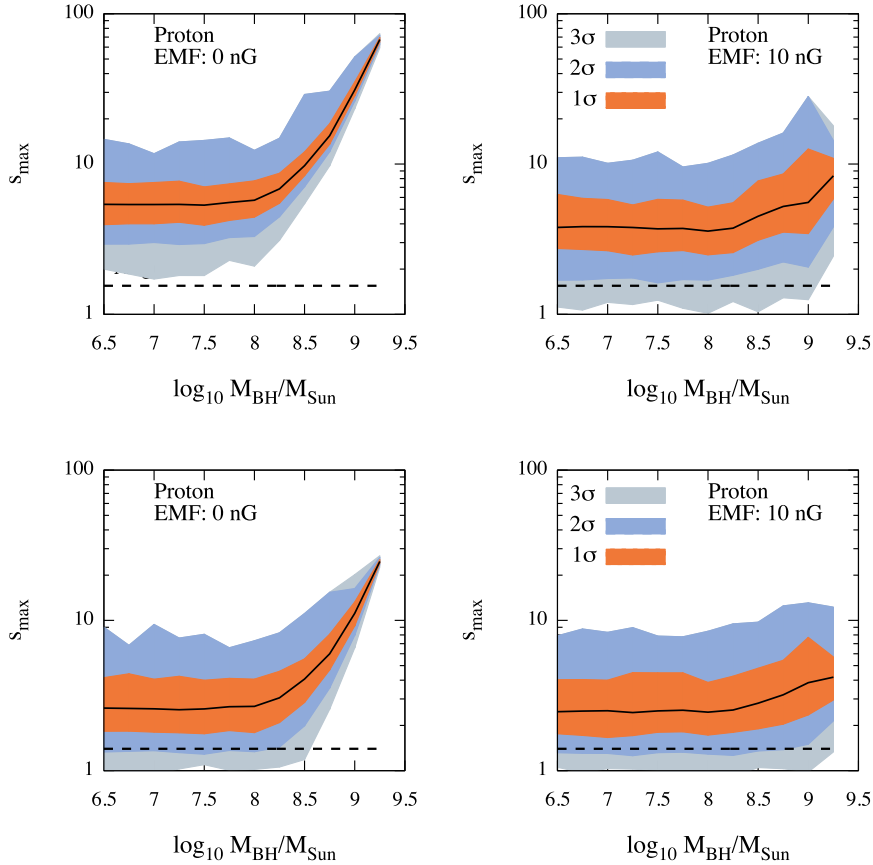


Figure 6.26: Multiscale autocorrelation of UHECRs with $E \geq 60$ EeV (upper panels) and $E \geq 80$ EeV (lower panels). It is shown the maximum clustering signal, as a function of the black hole mass threshold, expected from simulated sky maps (solid lines) with the corresponding dispersion bands (shaded areas) and the signal estimated from the data (dashed lines). **Left panels:** EMF is switched off. **Right panels:** $B_{\rm EMF} = 10$ nG is considered.

Constraining the black hole mass with the local and global method

Global approach. As an illustrative example of the global method, in Fig. 6.25 are shown the distributions of $s^{\star}_{n_s}$ for each probed black hole mass $M_{\rm BH}$ (i.e. source density). Protons with $E_{\rm thr} = 60$ EeV are considered, as well as the case with no magnetic field (left panel) and $B_{\rm EMF} = 10$ nG (right panel). It is evident that in our astrophysical scenario, the increase of the EMF intensity is responsible of decreasing the clustering coefficient, reducing its strength. In such a particular case, scenarios with no EMF are excluded with high CL.

Hence, we estimate the maximum clustering signal, as a function of $M_{\rm BH}$, expected from simulated sky maps (with the corresponding 1σ, 2σ and 3σ dispersion bands) and from the data. In Fig. 6.26 we show the results for scenarios including the EMF with strength $B_{\rm EMF} = 10$ nG (right panels) or no EMF (left panels), for $E_{\rm thr} = 60$ EeV (upper panels) and $E_{\rm thr} = 80$ EeV (lower panels). The global method provides only a rough indication of the black hole mass compatible with the data: in fact, provided confidence bands do not allow, in general, a definitive and significant conclusion. Only for $E_{\rm thr} = 60$ EeV, scenarios with small or vanishing EMF are disfavored.

Local approach. The bounds estimated with the local approach and shown in Fig. 6.23 for a particular scenario, should be interpreted with caution. In fact, when the EMF is switched off, we are in

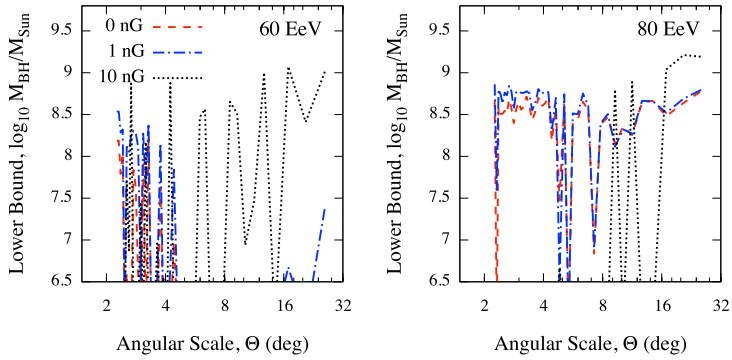


Figure 6.27: Lower bounds on the black hole mass (as in bottom-right panel of Fig. 6.23) for some different scenarios, namely with $B_{\rm EMF} = 0, 1$ and 10 nG and with $E_{\rm thr} = 60$ (left) and 80 EeV (right), as a function of the angular scale.

the same hypotheses of a recent similar analysis [303], where the angular scale Θ can be interpreted as the deflection due to intervening (unknown) magnetic fields, e.g. EMF and Galactic one. In this case the estimated bounds apply if the deflections of CR trajectories by magnetic fields do not erase the clustering properties expected from the models at the angular scales considered. If we switch on the EMF, thus including an explicit treatment of the magnetic field, the estimated bounds apply if the deflections of CR trajectories by galactic magnetic field are compatible with the angular scales considered. In Fig. 6.27 we show the lower bounds for some different scenarios, namely with $B_{\rm EMF} = 0, 1$ and 10 nG and with $E_{\rm thr} = 60$ and 80 EeV, as a function of the angular scale. The global effect of increasing the EMF is to provide less stringent lower bounds, at any angular scale. Such bounds range from $\approx 10^7 M_\odot$ to $\approx 10^9 M_\odot$ in the majority of cases, corresponding to a source density of $0.06 - 10 \times 10^{-4}$ Mpc^{-3} at 95% CL, in agreement with previous studies [285, 308]. It is worth noticing that for $E_{\rm thr} = 80$ EeV and weak EMF, the bounds are globally stable with respect to the angular scale, fluctuating around $10^{8.5} - 10^{8.75} M_\odot$, corresponding to a source density of $0.4 - 1 \times 10^{-4}$ Mpc^{-3} at 95% CL.

We have also estimated the average clustering signal, as a function of both the black hole mass and the angular scale, expected from simulated sky maps, and the clustering signal from the data. We expect more robust constraints on the allowed values of $M_{\rm BH}$ with respect to the global method because the whole information about the distribution of Θ is taken into account.

In Fig. 6.28 we show the results for $B_{\rm EMF} = 0$ nG (left panel) and $B_{\rm EMF} = 10$ nG (right panel) scenarios, with $E_{\rm thr} = 60$ EeV. In order to compare such results with the data, we make use of the χ^2 statistics defined by

$$\chi^2_{M_{\rm BH}} = \sum_\Theta \frac{[s_{\rm data}(\Theta) - \langle s_{\rm MC}(\Theta; M_{\rm BH})\rangle]^2}{\sigma_{\rm MC}(\Theta; M_{\rm BH})} \tag{6.8}$$

where $s_{\rm data}(\Theta)$ is the clustering signal estimated from the data, $\langle s_{\rm MC}(\Theta; M_{\rm BH})\rangle$ is the average clustering signal obtained from a scenario with a black hole mass threshold $M_{\rm BH}$ and $\sigma_{\rm MC}(\Theta; M_{\rm BH})$ is the corresponding dispersion.

In Fig. 6.29 are shown the values of the reduced χ^2 as a function of the black hole mass threshold for scenarios with different values of the EMF intensity. The mass $M^\star_{\rm BH}$ where the χ^2 gets a minimum is the most compatible with the observed clustering. We find that scenarios with $M > 10^{8.5} M_\odot$ (corresponding to a source density of $\approx 0.4 \times 10^{-4}$ Mpc^{-3}) are disfavored for $E_{\rm thr} = 60$ EeV. At higher energy threshold, the data is compatible with simulation with almost no regard for the EMF strength if $M < 10^{8.25} M_\odot$, whereas for $M < 10^9 M_\odot$ only scenarios with strong EMF are allowed.

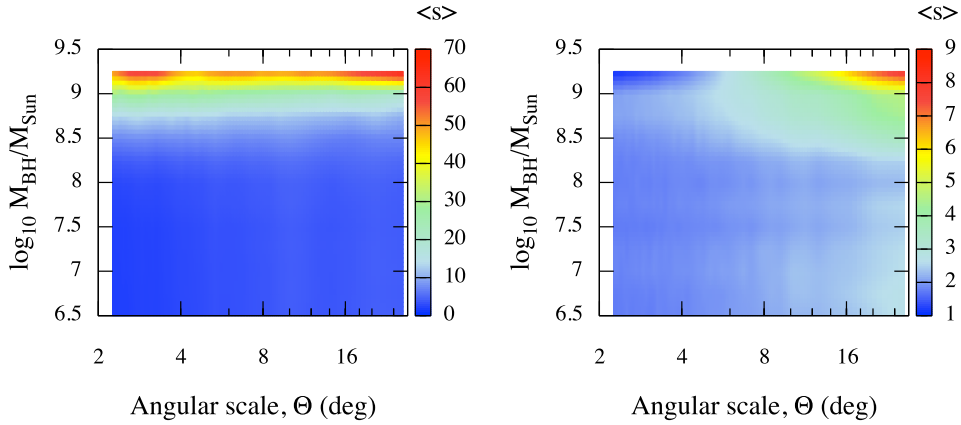


Figure 6.28: Multiscale autocorrelation of UHECRs with $E \geq 60$ EeV. It is shown the average clustering signal expected from simulated sky maps of protons, as a function of both the black hole mass and the angular scale. Scenarios with $B_{\rm EMF} = 0$ nG (left) and $B_{\rm EMF} = 10$ nG (right) are considered.

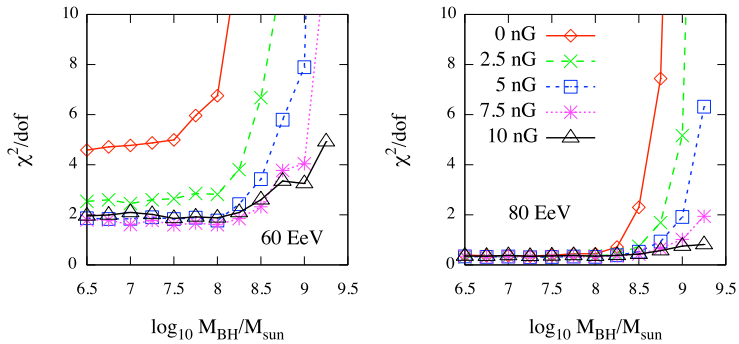


Figure 6.29: Reduced χ^2 statistics as a function of the black hole mass threshold, for different scenarios with $B_{\rm EMF}$ ranging from 0 to 10 nG, and for different energy threshold of protons, namely 60 (left) and 80 EeV (right).

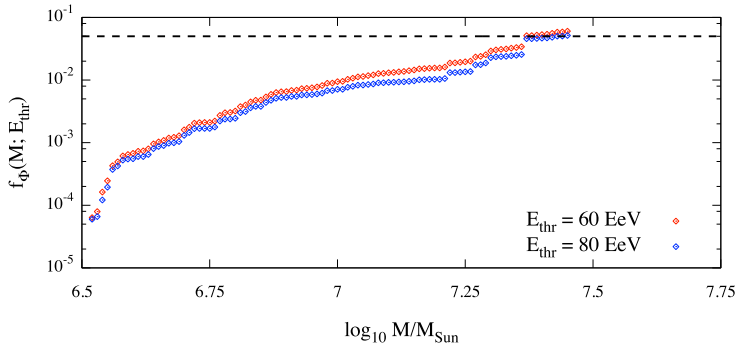


Figure 6.30: Fraction of UHE proton flux due to black holes with mass smaller than M, for different energy threshold of protons, namely 60 and 80 EeV. The 95% of the contribution to the flux is due to black holes with mass above $\approx 10^{7.3} M_{\odot}$.

It is worth noticing that a direct comparison with results reported in the previous sections (as well as in [303, 308] is not possible. In fact, the underlying assumption of those results include an equal

intrinsic luminosity of sources, whereas in the present work we have considered the dependence on the luminosity, which in turns depends on the black hole mass. Hence, heavier black hole provides a greater flux of protons and it should be remarked that the density of such objects decreases for increasing mass: such a behavior should have a non-negligible impact on the the effective density of sources. Let

$$\Omega(z, M; E_{\rm thr}) = \mathcal{L}(z, M) z^{-2} \omega(z, E_{\rm thr}) \tag{6.9}$$

be the probability that a black hole with mass M at distance z produces a UHE proton with energy above $E_{\rm thr}$ at Earth. In general, $\mathcal{L}(z, M) \propto (1+z)^m \mathcal{L}_{\rm BH}(M)$, where m is the evolution index. Under our hypothesis, $m = 0$ and $\mathcal{L}_{\rm BH} \propto M$. If $n(z, M')dz\ dM'$ is the number of black holes with mass between M' and $M' + dM'$, with distance between z and $z + dz$, the fraction of the flux Φ of UHECR that is produced by black holes with mass smaller than M, is given by

$$f_\Phi(M; E_{\rm thr}) = \frac{\int_{z_0}^{z_1} \int_{M_0}^{M} dz\ dM'\ n(z, M')\Omega(z, M; E_{\rm thr})}{\int_{z_0}^{z_1} \int_{M_0}^{M_1} dz\ dM'\ n(z, M')\Omega(z, M; E_{\rm thr})},$$

where the subscripts 0 and 1 indicates the lowest and highest values in the dataset, respectively. The flux fraction $f_\Phi(M; E_{\rm thr})$ is shown in Fig. 6.30 for both energy thresholds considered in this study: it is evident that black holes with mass below $10^{7.3} M_\odot$ do not contribute significantly (less than $\approx 5\%$) to the total flux of protons. Such a result, together with the previous one, suggest that only black holes with mass larger than $M_{\rm BH}$, with $10^{7.3} < M_{\rm BH}/M_\odot < 10^{8.75}$ are responsible for the observed clustering in the data

6.3.4 Conclusion

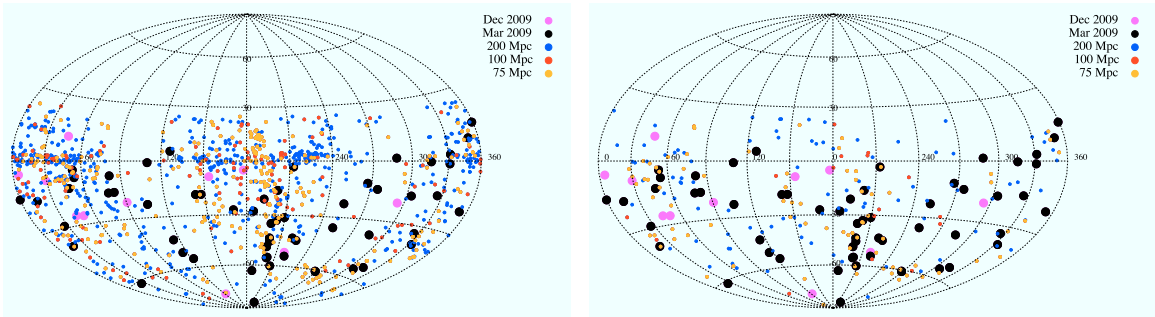
Within this study we have found that the mass of black holes plays an important role in anisotropy analysis. In fact, the luminosity of black holes is proportional to their mass, as well as the density of such objects in the nearby Universe, and both quantities have a direct impact on the clustering of UHECRs.

By varying the black hole mass threshold, we have built mock maps corresponding to different distribution of sources (up to ≈100 Mpc) and we have simulated the production of high energy protons and their propagation in a ΛCDM cosmology. The deflections due to different intensity of the turbulent extragalactic magnetic field have been taken into account, as well as detection effects as angular uncertainty and exposure of the Pierre Auger Observatory.

Successively, we have investigated the clustering features of such skymaps of events, for different density of sources (i.e. black hole mass threshold), magnetic field strength and energy threshold of UHECRs, by means of the novel multiscale autocorrelation function. Thus, we have estimated the lower bounds to black hole mass threshold for different astrophysical scenarios at 95% CL.

We have found that the main effect of including an explicit treatment of deflections due to the magnetic field is to decrease the lower bound on the density of black holes. Such results are in agreement with those ones reported in Ref. [303]. Finally, by assuming that nearby black holes are sources of UHE protons, our study constrain the minimal mass of such objects to be of the order of $10^8 - 10^9 M_\odot$, corresponding to a source density of the order $0.06 - 5 \times 10^{-4}$ Mpc^{-3} at 95% CL. However, it is worth remarking that in the case of higher energy threshold of protons at Earth, namely 80 EeV, the data is more compatible with simulations, even if the impact of the small statistic on this result at such energy is non-negligible. Moreover, scenarios with strong EMF are favored if the considered black holes are assumed to be the only sources of UHECRs.

Our results suggest that, even in the more extreme cases, black holes with mass smaller than $\approx 10^{7.3} M_\odot$ or larger than $\approx 10^{8.75} M_\odot$ are unlikely to be the only sources of UHE protons. Conversely, black holes with mass larger than $M_{\rm BH}$, with $10^{7.3} < M_{\rm BH}/M_\odot < 10^{8.75}$ are candidate sources of UHE protons observed at Earth above 60 and 80 EeV.



(a) Veron-Cetty-Veron 12th ed. catalog of AGN, QSO and BL Lac.

(b) SWIFT-BAT 39-month catalog of Seyfert galaxies, blazars, QSO, and unclassifed AGN.

Figure 6.31: Candidate sources (small dots of different colors are used according to redshifts: $z < 0.018$, $z < 0.024$ and $z < 0.047$) and Pierre Auger Observatory data (black and purple big dots). Equatorial coordinates are shown.

6.4 Influence of catalogues on the correlation between UHECRs and AGN

Within this work, we apply our multiscale procedure to Pierre Auger Observatory data (up to Dec. 2009) and catalogues of sources, to investigate the influence of the choice of the catalog on the correlation between UHECRs and AGN. In our analysis, we consider two catalogues of nearby candidate sources at different redshift ($z < 0.018$, $z < 0.024$, $z < 0.047$): Veron-Cetty-Veron 12th ed. [238] (VCV) and SWIFT-BAT 39-month [268, 297] (SWIFT). These catalogues have been used in other Auger analysis [89].

We consider the first 27 events with energy $E \geq 56$ EeV used in [208] and the 69 updated events (corresponding to the so-called Period I+II+III, see Sec. 5.2) with energy $E \geq 55$ EeV [89] detected by the Pierre Auger Observatory.

Sources with known redshift $z < 0.018$ (75 Mpc), $z < 0.024$ (100 Mpc) and $z < 0.047$ (200 Mpc) have been considered separately. In particular, AGN, QSO and BL Lac from the VCV catalog (Fig. 6.31a), and Seyfert galaxies, blazars, QSO, and unclassifed AGN from the SWIFT catalog (Fig. 6.31b), have been taken into account.

VCV Catalog. The chance probability for the correlation between Auger data and objects from the VCV catalog is shown in Fig. 6.32.

By using the first 27 events (see Ref. [208]) we find an evidence for correlation in great agreement with Auger published results [88]: $\sim 3.5\sigma$ for $\Theta^\star \approx 2.75°$ and $\sim 2.9\sigma$ for $\Theta^\star \approx 3.08°$, and sources within 75 Mpc.

Results for Period II and Period II + III (see Ref. [89]) show a less significant (but still present) correlation, at the same angular scales of the previous analysis.

Finally, by using Period I + II + III (corresponding to all of the 69 events in Ref. [89]), we found $\sim 2.6\sigma$ evidence for correlation when $\Theta^\star \approx 2.32°$, again for sources within 75 Mpc. Such results are consistent with the official analysis reported by the Pierre Auger Collaboration.

SWIFT Catalog. We perform the same analyses to investigate the correlation between Auger data and objects from the SWIFT catalog: results are shown in Fig. 6.33.

By using the 27 events we find a strong evidence for correlation: up to $\sim 3.9\sigma$ for $2.20° < \Theta^\star < 4.00°$ and $\Theta^\star \approx 4.86°, 7.81°, 10.19°$. Results are similar for both sources within 75 and 100 Mpc.

Results for Period II are still significant (up to $\sim 3.9\sigma$) for $2.50° < \Theta^\star < 3.38°$ and $\Theta^\star \approx$

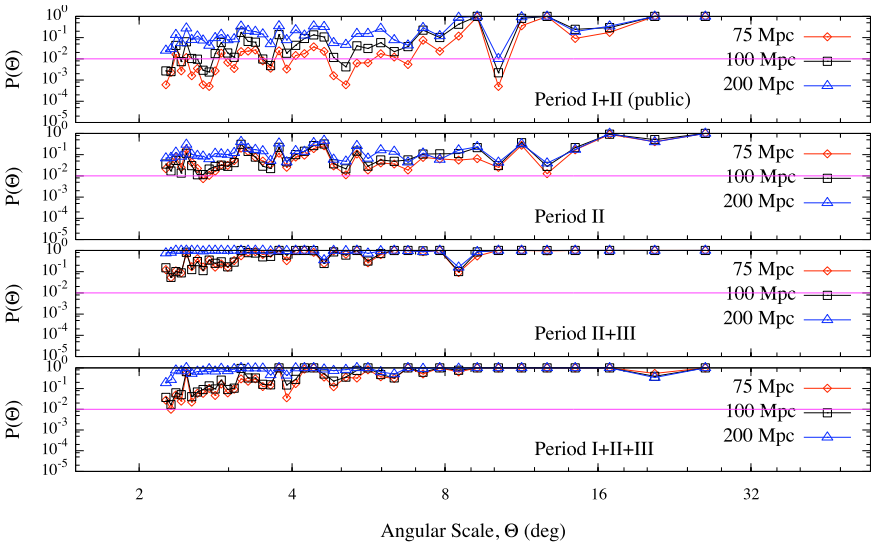


Figure 6.32: MCF: chance probability for cross-correlation between VCV catalog (varying redshift) and Auger data (varying periods). Solid straight line is the reference for $P = 1\%$.

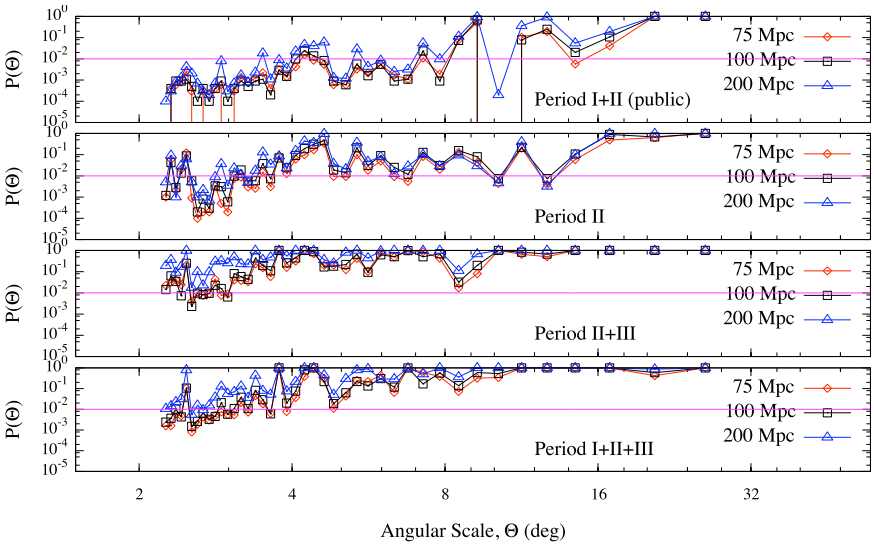


Figure 6.33: MCF: chance probability for cross-correlation between SWIFT catalog (varying redshift) and Auger data (varying periods). Solid straight line is the reference for $P = 1\%$.

$4.86°, 6.76°, 10.19°$, for sources within 75 and 100 Mpc.

For Period II + Period III the most significant angular scale is $2.53°$, with 2.9σ of evidence, for sources within 75 and 100 Mpc.

Finally, by using Period I + II + III, we found up to $\sim 3.1\sigma$ evidence for correlation when $2.25° < \Theta^\star < 3.08°$, again for sources within 75 Mpc and 100 Mpc.

Conclusion

We find evidence for correlation at different angular scales with both VCV and SWIFT catalogues, in agreement with results from other Auger analyses [88, 309, 310]. We get different evidence of correlation when using different data sub-sets (i.e periods) and the two catalogues. From our analyses, objects

within 75 Mpc and 100 Mpc are favorite candidate sources.

Angular scales ranging from 2° to 4° are, in general, statistically more significant than other ones. In the case of the SWIFT catalog, other interesting scales, ranging from 6° to 10°, emerge. It is worth remarking that Period III acts as an uncorrelated background of events, diluting the amount of correlation with both catalogues.

In addition, we note that the correlation signal with VCV catalog is always smaller than the correlation signal with SWIFT catalog. A further investigation is needed to understand if this behavior can be addressed to the incompleteness of the VCV catalog in the region where $|b| < 10°$, due to obscuration by the Milky Way, with respect to the SWIFT catalog providing the most uniform all-sky hard X-ray survey to date.

The correlation signal with AGN does not imply that AGN are the sources of UHECRs, but only that their directions are in a non-negligible relationship with those ones of such astrophysical objects that, however, may only act as tracers. Under the strong hypothesis that AGN are the real sources of UHECRs, the smaller angles at which the correlation is found suggests a light composition of UHECR above 55 EeV and small intervening magnetic fields.

6.5 Conclusions

In this chapter, we have investigated several astrophysical scenarios that could be responsible for the observed clustering and correlation in Auger data. We have performed several different studies to estimate the density of UHECR sources by varying the underlying hypotheses on the mass composition, the intensity of magnetic fields and distribution of sources.

We have found that the lower bound on the source number density is of the order of 10^{-4} Mpc^{-3} at 95% confidence level, hence excluding some classical candidate astrophysical objects as BL Lac and colliding galaxies, in favor of active galactic nuclei (AGN) and gamma-ray bursts (GRB).

Additionally, we have explored the possibility that nearby black holes are the sources of UHECRs detected with the Auger Observatory. We have found that the mass of black holes plays an important role in anisotropy analysis. In fact, the luminosity of black holes is proportional to their mass, as well as the density of such objects in the nearby Universe, and both quantities have a direct impact on the clustering of UHECRs. Our results suggest that, even in the more extreme cases, black holes with mass smaller than $\approx 10^{7.3} M_\odot$ or larger than $\approx 10^{8.75} M_\odot$ are unlikely to be the only sources of UHE protons. Conversely, black holes with mass larger than $M_{\rm BH}$, with $10^{7.3} < M_{\rm BH}/M_\odot < 10^{8.75}$ are candidate sources of UHE protons observed at Earth above 60 and 80 EeV.

Finally, we have got different evidence of correlation between Auger data and two catalogues of candidate sources, the Veron-Cetty-Veron 12th ed and the SWIFT-BAT 39-month. From our analyses, AGN within 75 Mpc and 100 Mpc are favorite candidate sources, although the correlation signal with AGN does not imply that they are the sources of UHECRs, but only that their directions are in a non-negligible relationship with those ones of such astrophysical objects that may only act as tracers of the actual UHECR sources.

Every great and deep difficulty bears in itself its own solution. It forces us to change our thinking in order to find it.

N. Bohr

We cling to our own point of view, as though everything depended on it. Yet our opinions have no permanence; like autumn and winter, they gradually pass away.

Chuang Tzu

Chapter 7

A phenomenological global picture of Auger data above 60 EeV

In this conclusive chapter we aim at a global interpretation of Auger data above 60 EeV. In the previous chapters we have focussed on searching (and finding) anisotropy and correlation with astrophysical objects in the arrival directions of UHECR detected by Auger. Here we include also the results of the measurement of the elongation rate and of the energy spectrum (see Chap. 1). We use these measurements to constrain astrophysical scenarios compatible at the same time with the observed spectrum, elongation rate, clustering of the arrival directions and correlation with AGNs. We do so by considering different models of source distribution, by including the production and propagation of UHECR in the universe, and by considering different models of composition at the source.

7.1 Introduction

The recent results reported by the Pierre Auger Collaboration apparently show opposite conclusions on the nature of UHECRs. In fact, the correlation of events with energy above 55 EeV with nearby extragalactic matter within 3.1° [89] suggests the predominance of a light composition. On the other hand, the measurements of the elongation rate end below 40 EeV even if already above few EeV the data shows a clear trend towards a heavier composition when compared with simulations corresponding to different hadronic models [92].

The main goal of the study presented here is to show that, under certain assumptions, observations and measurements are not contradictory with each other and, moreover, with the observed flux of UHECRs and their intrinsic clustering above 60 EeV.

In the previous chapter we have considered different models of source distribution in the nearby universe (up to 250 Mpc), we have simulated the production of high energy protons and their propagation in a ΛCDM universe, building mock maps of events in the Auger sky. We have investigated the clustering features of such events by varying the main parameters in that study, namely the density of sources and the energy threshold of protons, by means of the novel multiscale method [278]. We have shown that the clustering analysis of Auger data allows to constrain the source density of UHECRs. Successively, we have shown that such an approach allows to put constraints on both the source density and the magnetic field (see Sec. 6.1 and Sec. 6.2).

In this chapter, we adopt the same procedure to constrain representative astrophysical scenarios with different features (strength of the magnetic field, distribution of sources) by varying the fraction of protons and iron nuclei in the simulated sky maps. Moreover, we use the value of $X_{\max}$ above 60 EeV, extrapolated from the most up-to-date Auger observations [91] for different hadronic models under the

hypothesis of a two-species mixed composition model, to select the astrophysical scenario(s) compatible with the data.

We will consider two toy models for the distribution of sources: one with active galactic nuclei (AGN) from the SWIFT-BAT 58-months catalog [282] and another one with uniformly distributed sources in the nearby universe, up to 200 Mpc. We will consider both the cases where sources are assumed to be with uniform and equal luminosity, and where sources are assumed with their intrinsic luminosity. In any case, the injection spectrum is assumed to follow a power law. As in studies presented in previous chapters the produced particles (protons and iron nuclei) are successively propagated in a ΛCDM universe, taking into account non-negligible energy-loss processes in the cosmic microwave and infrared background photon fields. The effects of the irregular component of both galactic magnetic field (GMF) and extragalactic magnetic field (EMF) are taken into account through a smearing procedure, as well as the effect of experimental uncertainty on the angular resolution.

The clustering analysis is performed on Auger-like skies by varying all the variables in our parameter space: the distribution of source, the EMF from 0.1 nG to 10 nG, a soft and a hard GMF, the composition of UHECRs ($(Z, A) = (1, 1)$ and $(26, 56)$). The energy threshold of particles reaching the Earth is fixed to 60 EeV.

7.2 Simulation of sources and UHECRs

We consider two simple toy models for investigating the ability of the MAF method to discriminate among different source densities and magnetic fields. In the first model, events are generated from sources distributed according to the SWIFT-BAT 58-month catalog of AGN and, in the following, we will refer to this model as "SWIFT". In the second model, events are generated from sources isotropically and homogeneously distributed in a sphere of a given radius, and, in the following, we will refer to this model as "ISOHOM". The particular choice of such models is justified by the fact that for a fixed source density n_s and EMF strength B, we are interested in investigating the clustering differences between skies of events following the distribution of matter in the nearby universe (2MRS model) and skies of events generated by randomly distributed sources (ISOHOM model). The source density of the selected sample of AGN within 200 Mpc is $\approx 1.1 \times 10^{-5}$ Mpc^{-3}. In the following, we will consider the following cases for the sources:

1. ISOHOM model: equal intrinsic luminosity;
2. SWIFT model I: equal intrinsic luminosity;
3. SWIFT model II: observed intrinsic luminosity.

For the simulation of UHECRs we follow the procedure described in Sec. 6.2, for both protons and heavier nuclei. Thus, the probability to get an event from a source at redshift z is proportional to $\mathcal{L} z^{-2} \omega(z, E_{\rm thr})$ (defined by Eq. 3.60), being $\mathcal{L}$ the intrinsic luminosity of the source.

7.3 Simulation of magnetic fields

The simulation of magnetic fields follows the same procedure described in Sec. 6.2. In the following, in order to put in evidence the main differences in the clustering features, we will consider some models involving different configuration of magnetic fields:

1. **Hard GMF+EMF**: where $B_{\rm GMF} = 15\ \mu$G and $B_{\rm EMF} = 1$ nG;
2. **Soft GMF+EMF** I: where $B_{\rm GMF} = 6\ \mu$G and $B_{\rm EMF} = 0.1$ nG;

3. **Soft GMF+EMF** II: where $B_{\rm GMF} = 6\ \mu$G and $B_{\rm EMF} = 1$ nG;

4. **Soft GMF+EMF** III: where $B_{\rm GMF} = 6\ \mu$G and $B_{\rm EMF} = 10$ nG;

The coherence length is fixed to 100 pc for all GMFs and to 1 Mpc for all EMFs. Such an analysis, at least in the GMF case, should require a full propagation code for estimating the traversing distance of the UHECR inside our galaxy. However, we are not interested in achieving such a precision in this particular case and we use the rough approximation that the average propagation distance in the galaxy is ~ 4 kpc in the HGMF+EMF scenario and ~ 2 kpc in the SGMF+EMF scenario.

Note. The probability to get an event from a source has been described in Sec. 6.1. In the case of iron, such a probability takes into account the photo-disintegration process and the possibility that a lighter fragment may reach the Earth, with energy above the given threshold. However, when dealing with the smearing of the arrival direction, because of the turbulent magnetic fields, we will assume that the nuclei are always irons. Such an assumption is required because we do not perform a full propagation from the source to the Earth. In a more realistic scenario the average deflection should be smaller than that one considered in our study. However, we have estimated that our error has a negligible impact on the final results. For instance, for $E_{\rm thr} = 60$ EeV, the average mass of the fragment reaching the Earth is estimated to be $\langle A \rangle \approx 46$ [1], with a corresponding atomic number $\langle Z \rangle \approx 23$. The spreading angles are proportional to Z: it follows that, on average, we overestimate the deflection when considering an iron nucleus instead of the "true" fragment. Quantitatively, the overestimation factor is given by $26/\langle Z \rangle \approx 1.13$. The main consequence of such an overestimation is reflected by an underestimation of the clustering signal, due to the larger spreading angle, for angular scale within $\sim 30°$. In fact, above such a threshold, the MAF is not able to discriminate the angular scale of clustering with accuracy. Deflections smaller than such an angular scale are obtained for fragments with $Z \geq 7$ and $Z \geq 17$, in HGMF+EMF and SGMF+EMF scenarios, respectively, for $E = 60$ EeV: it follows that we really underestimate the clustering signal in those cases where the iron nuclei have been produced by far sources (leading to an higher probability to have a fragment at Earth with $Z < 17$ or $Z < 7$). However, we have estimated that the probability, integrated over the distance, to have a fragment with $A < 34$ ($A < 14$) is smaller than $\approx 1\%$ ($\approx 7\%$), i.e. in practice our assumption will have a negligible impact on the final results.

Deflections due to magnetic fields and experimental uncertainty are taken into account as described in Sec. 6.2.1.

7.4 Description of the methods adopted for the clustering analysis

We analyze the mock maps of events generated for different values of a set of parameters defining a particular astrophysical scenario. First, we briefly discuss the global and the local approaches used for clustering and correlation analyses. Successively, we investigate the ability of our estimators to discriminate among different astrophysical scenarios. The comparison with results obtained from Auger data is also discussed. For the sake of completeness, we use the official data set of Auger events up to 31 Dec 2010 with standard quality cuts for anisotropy and correlation studies, and energy $E \geq 60$ EeV. The data set is shown in Fig. 7.1 superimposed to AGN within 200 Mpc from the SWIFT-BAT 58-months catalog.

[1] Such an information has been deduced from Fig. 7 in Ref. [171].

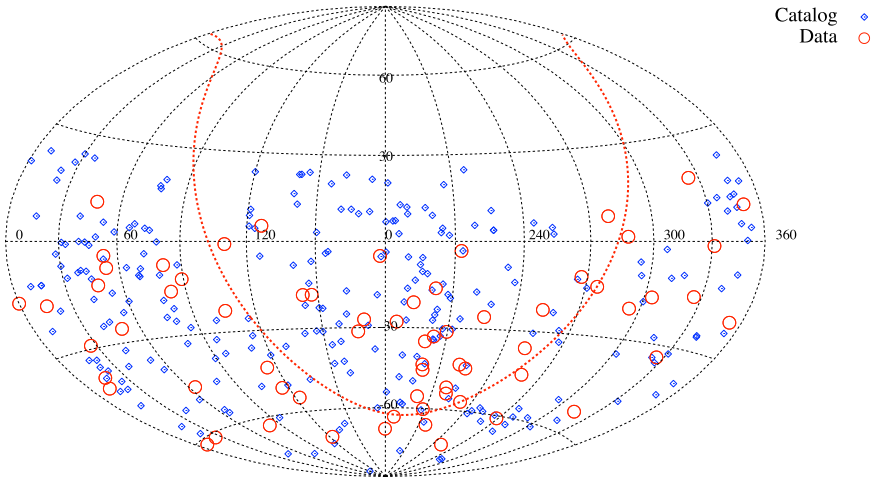


Figure 7.1: Auger events up to 31 Dec 2010 with energy $E \geq 60$ EeV (circles) and AGN within 200 Mpc from the SWIFT-BAT 58-months catalog (diamonds). The dashed line indicates the galactic plane (equatorial coordinates are shown).

7.4.1 The global approach

As for the case of the analysis on the bounds of source density and EMF (see Sec. 6.1 and Sec. 6.2), we adopt a *global* method to discriminate among models. In this study, it is intended to provide bounds to the fraction of protons, for different astrophysical scenarios, by performing the comparison between the data and the models through maximum values $s^{\star}_{f_p} = \max\{s_{f_p}(\Theta)\}$. For each astrophysical scenario and fraction of protons f_p, we build the distribution of $s^{\star}_{f_p}$ for testing the hypothesis that the value $s^{\star}_{\rm data}$, obtained from the data, is drawn from the distribution. The adopted procedure is equivalent to that one adopted in the previous analyses, where the dependence on the angular scale Θ vanishes. It is worth noticing that such a method can be applied to both intrinsic clustering (where s is the MAF) and correlation (where s is the MCF) analyses.

7.4.2 The local approach

We have already proposed a *local* method to constrain the parameters of the study in Sec. 6.1. This approach is explicitly based on hypothesis testing, for each angular scale Θ separately. Let $\mathcal{H}_0(\Theta, f_p)$ be the null hypothesis that $s_{\rm data}(\Theta)$ is a random outcome of the distribution of $s(\Theta)$ corresponding to the fraction of protons f_p, for a given astrophysical scenario. The alternative hypothesis is simply the negation of $\mathcal{H}_0(\Theta, f_p)$ and the test size α, i.e. the significance, is fixed *a priori*. $\mathcal{H}_0(\Theta, f_p)$ is rejected if the fraction of skies providing a clustering more extreme than that of data is smaller than α. In particular, when we test for the upper bound, $\mathcal{H}_0(\Theta, f_p)$ is rejected if $p(\Theta, f_p) \leq \alpha$, with

$$p(\Theta, f_p) = \mathbb{P}\left[s_{\rm data}(\Theta) \leq s_{f_p}(\Theta)\right],$$

whereas if we test for the lower bound, $\mathcal{H}_0(\Theta, f_p)$ is rejected if $p(\Theta, f_p) \leq \alpha$, with

$$p(\Theta, f_p) = \mathbb{P}\left[s_{\rm data}(\Theta) \geq s_{f_p}(\Theta)\right].$$

The lower bound to the fraction of protons, at the angular scale Θ, is defined to be the lowest value of f_p for which $\mathcal{H}_0(\Theta, f_p)$ has not been rejected. Analogously, the upper bound to the fraction of protons is defined to be the highest value of f_p for which $\mathcal{H}_0(\Theta, f_p)$ has not been rejected. It is worth remarking that, within this approach, the comparison between the distribution of $s(\Theta)$ corresponding

to a particular value of the fraction of protons and the data is independently performed twice: the first time to test the hypothesis that the fraction is an upper bound and the second one to test the hypothesis that the fraction is a lower bound.

Within such a procedure, the dependence on the angular scale Θ does not vanish (for further details on this method we refer to Sec. 6.1). It is worth noticing that such a method can be applied to both intrinsic clustering (where s is the MAF) and correlation (where s is the MCF) analyses.

7.5 Clustering analysis of simulated sky maps and of Auger data

In this section, we estimate the intrinsic clustering signal of UHECRs from both simulated sky maps and data, by using the MAF. We consider the two scenarios involving different configurations of magnetic fields (namely, HGMF+EMF and SGMF+EMF) and we separately consider some cases where the strength of the EMF is varied. All events are simulated as previously discussed. We make use of both global and local methods to obtain information about the fraction of protons compatible with clustering signal estimated from the data. It worth remarking that the chance probability, properly penalized because of the scan over the angular scale is $\tilde{P} \approx 40\%$, thus compatible with the isotropic expectation and in agreement with the results reported in Ref. [303].

In the following, intrinsic luminosity of sources is always considered, if not specified otherwise.

Results obtained from the global method

We estimate the maximum correlation signal, as a function of the fraction of protons, expected from simulated sky maps (with the corresponding 1σ, 2σ and 3σ dispersion bands) and the signal from the data. In Fig. 7.2 we show the results for ISOHOM (left panel) and SWIFT (right panel) models, in the HGMF+EMF scenario ($B_{\rm EMF} = 1$ nG). The global method provides only a rough indication of the fraction of protons compatible with the data: in fact, provided confidence bands do not allow a definitive conclusion. In each case, the allowed fraction of protons is estimated to be around 50% within 95% CL.

In Fig. 7.3 we show the same plots of Fig. 7.2 but for the SGMF+EMF scenario. In this case, the two models are better discriminated than the previous case, although confidence bands are still wide. It is worth noticing that the SGMF+EMF scenario generally provides a smaller value for the allowed fraction of protons.

Results obtained from the local method

We estimate the average clustering signal, as a function of both the fraction of protons and the angular scale, expected from simulated sky maps, and the clustering signal from the data. We expect more robust constraints on the allowed fraction of protons with respect to the global method because the whole information about the distribution of Θ is taken into account.

In Fig. 7.4 we show the results for HGMF+EMF (left panel) and SGMF+EMF (right panel) scenarios. In order to compare such results with the data, we make use of the χ^2 statistics defined by

$$\chi^2_{f_p} = \sum_{\Theta} \frac{\left[s_{\rm data}(\Theta) - \langle s_{\rm MC}(\Theta; f_p)\rangle\right]^2}{\sigma_{\rm MC}(\Theta; f_p)} \tag{7.1}$$

where $s_{\rm data}(\Theta)$ is the clustering signal estimated from the data, $\langle s_{\rm MC}(\Theta; f_p)\rangle$ is the average clustering signal obtained from a scenario with a fraction of protons f_p and $\sigma_{\rm MC}(\Theta; f_p)$ is the corresponding dispersion.

In Fig. 7.5 we show the values of the reduced χ^2 as a function of the fraction of protons for the HGMF+EMF (left panel) and the SGMF+EMF (right panel) scenarios. The fraction $f_p^\star$ where the χ^2

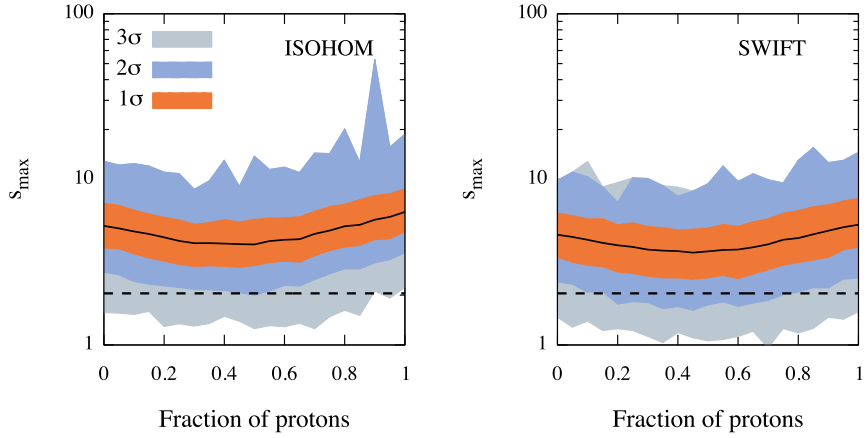


Figure 7.2: Multiscale autocorrelation of UHECRs with $E \geq 60$ EeV. It is shown the maximum clsutering signal, as a function of the fraction of protons, expected from simulated sky maps (solid lines) with the corresponding dispersion bands (shaded areas) and the signal estimated from the data (dashed lines). *Left panel:* ISOHOM model; *Right panel:* SWIFT model. HGMF+EMF scenario ($B_{\rm EMF} = 1$ nG) is considered.

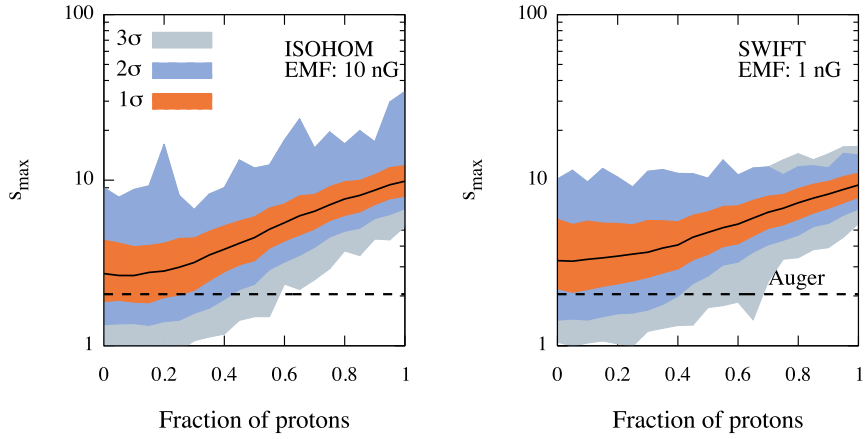


Figure 7.3: Same as in Fig. 7.2 but for the SGMF+EMF scenario.

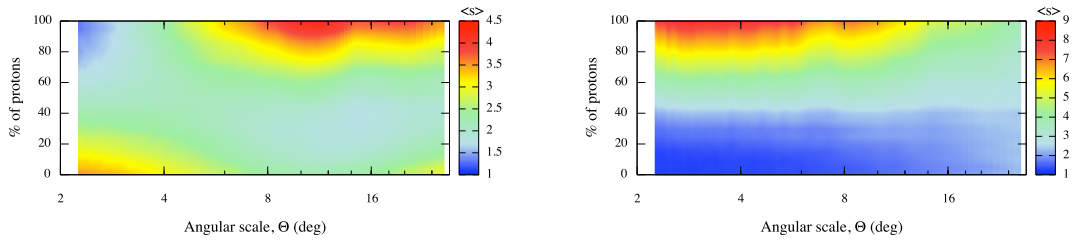


Figure 7.4: Multiscale autocorrelation of UHECRs with $E \geq 60$ EeV. It is shown the average clustering signal, as a function of the fraction of protons and of the angular scale, expected from simulated sky maps. HGMF+EMF (left) and SGMF+EMF (right) models are considered.

gets a minimum is the most compatible with the observed correlation. We find fractions around 50% and $10 - 20\%$ for HGMF+EMF and SGMF+EMF scenarios, respectively, with no differences between ISOHOM and SWIFT models. However, the latter result is not surprising: the density of sources adopted for the simulations is of the order of 10^{-5} Mpc^{-3}, whereas significant differences, between the

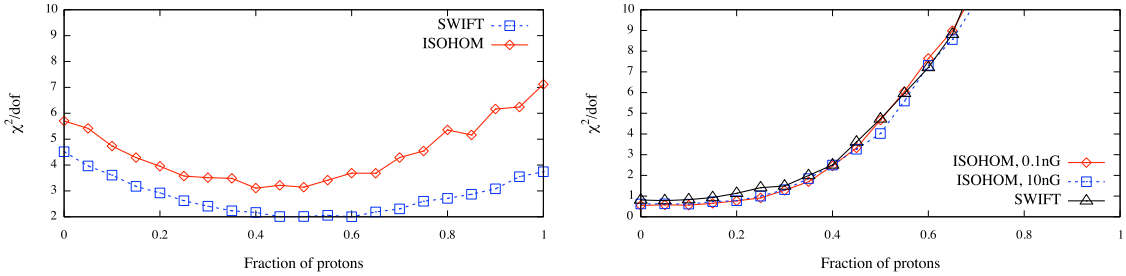


Figure 7.5: Reduced χ^2 statistics as a function of the fraction of protons, for the ISOHOM and SWIFT models. **Left:** HGMF+EMF scenario ($B_{\rm EMF} = 1$ nG). **Right:** SGMF+EMF scenario ($B_{\rm EMF} = 1$ nG for SWIFT).

clustering of events from a uniform distribution of sources and from a distribution following the large scale structure of matter in the nearby universe, arise at higher densities (see Sec. 6.1 and Sec. 6.2).

However, it is worth remarking that we have the best discrimination power only in the SGMF+EMF model. Conversely, in the other cases, the values of reduced χ^2 vary in a small range (namely, from ≈ 2 to ≈ 7), making more difficult a statistically significant discrimination. Such results are compatible with that ones obtained within the global method, but provide a more precise estimation of the allowed fraction of protons corresponding to each scenario. If a comparison between all reduced χ^2 curves is considered, the SGMF+EMF scenario is the most compatible model with the observed clustering.

The general conclusion of this particular analysis is that in the case of weaker magnetic fields (i.e. small deflections) the data is compatible with a small fraction of protons, whereas for stronger magnetic fields (i.e. large deflections) a higher fraction of protons, around 50%, is required to reproduce the clustering signal. In order to understand which astrophysical scenario is more likely, in the following we will perform independent analyses based on other physical observables as the elongation rate, the spectrum and the correlation with AGN.

7.6 The elongation rate predictions from different hadronic models

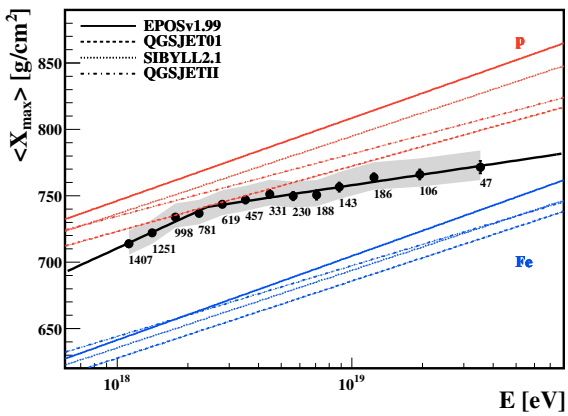


Figure 7.6: Average $X_{\rm max}$ estimated from the data (black points) and from different hadronic models (lines). The black solid line intersecting the data indicates a fit [91].

The most up-to-date measurements of the elongation rate end around 40 EeV [91], and it is not possible to draw definitive conclusions on the composition of Auger events above such an energy scale. In Fig. 7.6 we show the elongation rate estimated from the data up to 31 December 2010, together

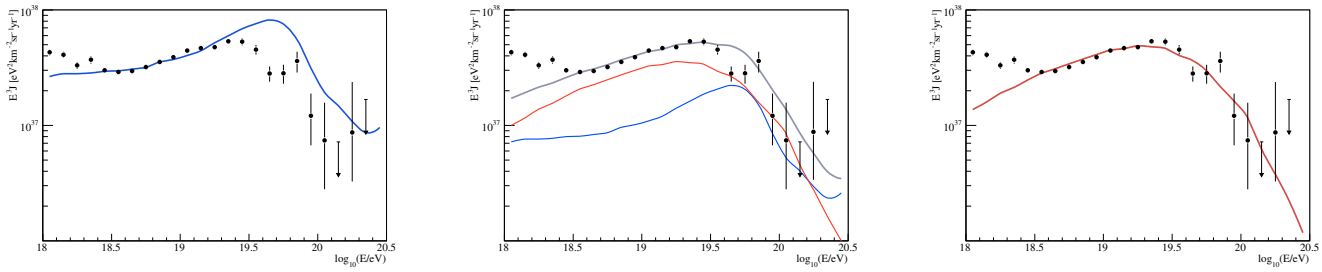


Figure 7.7: Flux scaled by E^3 (power law with $s = 2.4$ and no cutoff) for 100% protons (left panel), 50% protons and 50% iron (central panel), 100% iron (right panel). The red and blue lines refer to the all-particle spectrum at Earth coming from the two primaries; the gray line in the case of mixed composition refer to the all-particle spectrum from the sum of the two all-particle spectra. Black solid points indicates the flux reported by the Pierre Auger Collaboration [28].

with values estimated from different hadronic models. It is evident a clear trend towards a heavier composition, although the value of $\langle X_{\max} \rangle$ can not definitively shed light on the composition of UHECRs. The predicted value of $\langle X_{\max} \rangle$ for $E = 60$ EeV is obtained from the fit in the figure, and it corresponds to ≈ 778.7 g cm^{-2}. In order to know the predicted values from the hadronic model, we assume a mixed composition with only two species (namely, protons and irons):

$$\langle X_{\max} \rangle_{\text{fit}} = f_{\text{p}} \langle X^{\text{p}}_{\max} \rangle_{\text{hm}} + (1 - f_{\text{p}}) \langle X^{\text{Fe}}_{\max} \rangle_{\text{hm}} \tag{7.2}$$

where "hm" indicates the hadronic model. The predicted fractions f_p are reported in Tab. 7.1: values obtained from EPOS are in agreement with allowed fractions estimated from intrinsic clustering analysis in the SGMF+EMF scenario, whereas the other hadronic models are in agreement with allowed fractions in the HGMF+EMF scenario. Such a results provides an interesting relationship between the elongation rate and the deflections of UHECRs in the GMF, that needs further investigation.

Model	EPOSv1.99	SIBYLLv2.1	QGSJETII	QGSJET01
%protons	24	39	50	60

Table 7.1: Predicted fractions of protons from different hadronic models for $E = 60$ EeV.

It is worth remarking that, at this first step, we are neglecting information carried by the other popular shower profile observable, namely the dispersion on $X_{\max}$. In fact, it has been recently shown that such an observable reflects the propagation of nuclei and then induces fluctuations on $X_{\max}$ that are not compatible with a pure composition at Earth [311]. In the future developments of the present work, we will include the dispersion among the other observables adopted in our framework.

7.7 The energy spectrum in different scenarios

In this section we discuss some particular astrophysical scenarios from the point of view of the UHECR spectrum. The spectra we show are obtained with a dedicated simulation code, named *SimProp*, developed in collaboration between the groups of L'Aquila and Roma II [211]. The code has been designed to simulate the mono-dimensional propagation of nuclei and is mainly based on the analytical scheme by Aloisio *et al.* [187, 312].

We simulate the generation of pure compositions, in particular protons and iron nuclei, and also a mixed composition of these two species (see Fig. 7.7). Here we decide to choose the case of a single

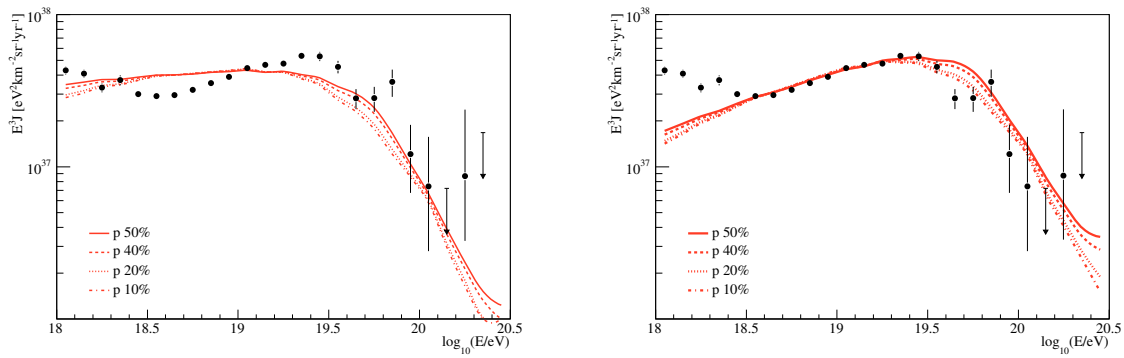


Figure 7.8: Flux scaled by E^3 for different composition, no energy cutoff, power law $s = 2.7$ (left panel) and $s = 2.4$ (right panel), superimposed to the flux measured with the Pierre Auger Observatory (black solid points).

spectral index, namely $s = 2.4$. This choice does not come from any optimization procedure, but refers to a simulated spectrum [183] calculated for comparison with Auger data in Ref. [313]. For simplicity, we assume the same spectral index for protons. Moreover, in this study we do not assume an energy cutoff at the source. However, it should be noticed that even if the energy cutoff at the source has no significative impact on the all-particle spectrum, it can have implications on other observables, as $\langle X_{\max} \rangle$ and $\sigma(X_{\max})$, as pointed out in [311]. Finally, the calculated spectra are normalized to the Auger spectrum above $10^{18.8}$ eV.

In Fig. 7.8 we compare the all-particle spectra for different values of the proton fraction and different values of the spectral index. Even if this work is not the result of a fit, the spectral index $s = 2.7$ can be easily excluded by looking at the left panel of Fig. 7.8, for different source compositions. Moreover, regarding the single proton composition at the source, this hypothesis (with spectral index $s = 2.4$) can be excluded by the comparison with the Auger spectrum in the left panel of Fig. 7.7.

In conclusion, except these extreme hypotheses, the energy spectrum alone is not able to discriminate between different source composition, as demonstrated in the right panel of Fig. 7.8, although it is still able to discriminate between different injection spectra. In the future developments of the present work, we will perform a more precise fine tuning of the main parameters involved in this particular study, by using a more advanced optimization procedure.

As it has been pointed out in Ref. [314], the understanding of the spectrum is still debated. It is generally assumed that below the "ankle" the main contribution to the spectrum is provided by CRs of galactic origin, whereas above the "ankle" the contribution is considered to be extragalactic. However, above the "ankle", the galactic component could still supply a significant number of CRs to the overall flux. The slope in this region of the spectrum, presumed end of the galactic component, is the key parameter that should shed light on such a contribution to the flux. Our results suggest that $s = 2.4$ represents an upper limit to the allowed values of the spectral index: in fact, smaller values could be also valid if a significant contribution from the galactic component is present.

7.8 The correlation with AGN

In this section, we estimate the correlation signal between UHECRs and AGN within 75 Mpc from the SWIFT-BAT 58-month catalog, by using the MCF. The choice of such a catalog is justified by its completeness, even in the region around the galactic plane. We consider the two scenarios involving different configurations of magnetic fields (namely, HGMF+EMF and SGMF+EMF, with $B_{\rm EMF} = 1$ nG)

and we separately consider the cases where the intrinsic luminosity of the sources is taken into account or not. All events are simulated as previously discussed. Again, we make use of both global and local methods, already discussed in the case of the intrinsic clustering analysis, to obtain information about the fraction of protons compatible with correlation signal estimated from the data, shown in Fig. 7.9.

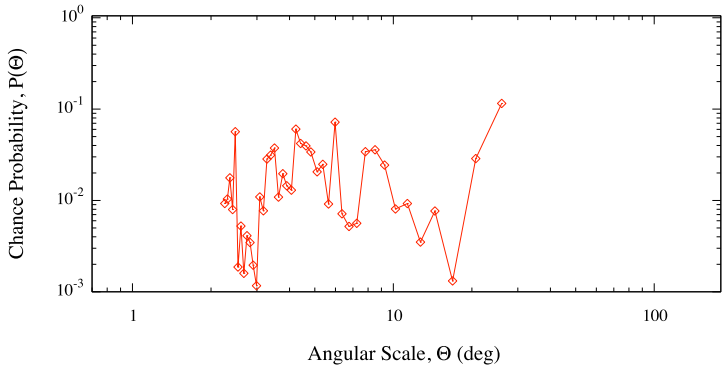


Figure 7.9: Multiscale cross-correlation between Pierre Auger events with $E \geq 60$ EeV and AGN within 75 Mpc from the SWIFT-BAT 58-month catalog. The chance probability, properly penalized because of the scan over the angular scale is $\tilde{P} \approx 4.2\%$ ($\Theta^\star \sim 3°$).

The chance probability, properly penalized because of the scan over the angular scale is $\tilde{P} \approx 4.2\%$ for $\Theta^\star \sim 3°$, although a similar probability ($\approx 4.5\%$) is obtained for $\sim 17°$. Such a result suggests that correlation is expected to be found in $\approx 95\%$ of cases with SWIFT-like scenarios, hence, in the following, we will consider only events generated from the SWIFT model for the distribution of sources.

Results obtained from the global method

We estimate the maximum correlation signal, as a function of the fraction of protons, expected from simulated sky maps (with the corresponding 1σ, 2σ and 3σ dispersion bands) and the signal from the data. In Fig. 7.10 we show the results for the HGMF+EMF scenario when the intrinsic luminosity of sources is taken into account (left panel) or not (right panel). The global method provides only a rough indication of the fraction of protons compatible with the data: in fact, provided confidence bands do not allow a definitive conclusion. Moreover, the assumption of considering the number of events coming from a source to be proportional to the intrinsic luminosity of the source itself, is reflected in a increase of 15% in the allowed fraction of protons (where the data coincides with the average maximum clustering of the model). In each case, the fraction is estimated to be larger than 50%.

In Fig. 7.11 we show the same plots of Fig. 7.10 but for the comparison between HGMF+EMF (left panel) and SGMF+EMF (right panel) scenarios. In this case, the two models are better discriminated than the previous case, although confidence bands are still wide. It is worth noticing that the SGMF+EMF scenario provides a smaller value for the allowed fraction of protons, around 20%. Such a result is partially expected because, for deflections smaller than those ones corresponding to the HGMF+EMF scenario, the MCF is more sensitive to correlating events, hence requiring a smaller number of protons coming from the sources.

Results obtained from the local method

We estimate the average correlation signal, as a function of both the fraction of protons and the angular scale, expected from simulated sky maps, and the correlation signal from the data. We expect more stringent constraints on the allowed fraction of protons with respect to the global method because the information about the distribution of Θ is taken into account.

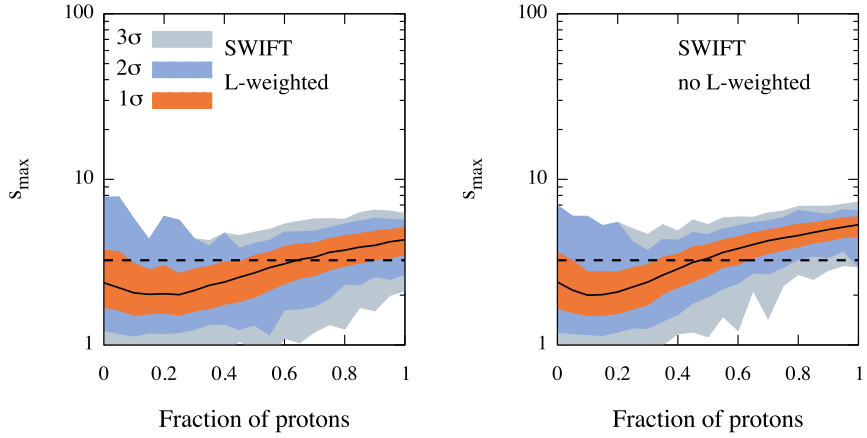


Figure 7.10: Multiscale cross-correlation between UHECRs and AGN within 75 Mpc from the SWIFT-BAT 58-month catalog. It is shown the maximum correlation signal, as a function of the fraction of protons, expected from simulated sky maps (solid lines) with the corresponding dispersion bands (shaded areas) and the signal estimated from the data (dashed lines). *Left panel:* real intrinsic luminosity $\mathcal{L}$ of sources are used; *Right panel:* sources are assumed to be with equal intrinsic luminosity ($\mathcal{L} = 1$). HGMF+EMF model is considered.

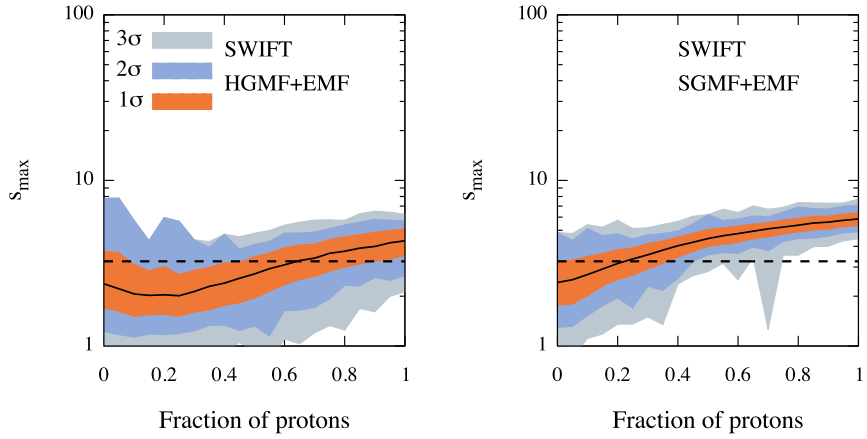


Figure 7.11: Same as in Fig. 7.10 but for the comparison between HGMF+EMF (left panel) and SGMF+EMF (right panel) models. Intrinsic luminosity of sources is considered.

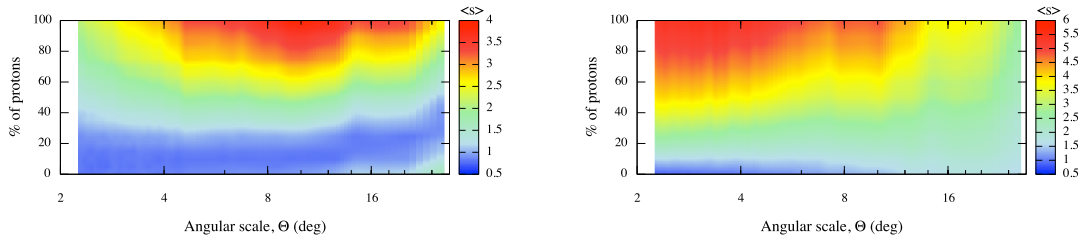


Figure 7.12: Multiscale cross-correlation between UHECRs and AGN within 75 Mpc from the SWIFT-BAT 58-month catalog. It is shown the average correlation signal, as a function of the fraction of protons and of the angular scale, expected from simulated sky maps. HGMF+EMF (left) and SGMF+EMF (right) model are considered.

In Fig. 7.12 we show the results for HGMF+EMF (left panel) and SGMF+EMF (right panel) scenarios, when the intrinsic luminosity of sources is taken into account. As in the case of the intrinsic

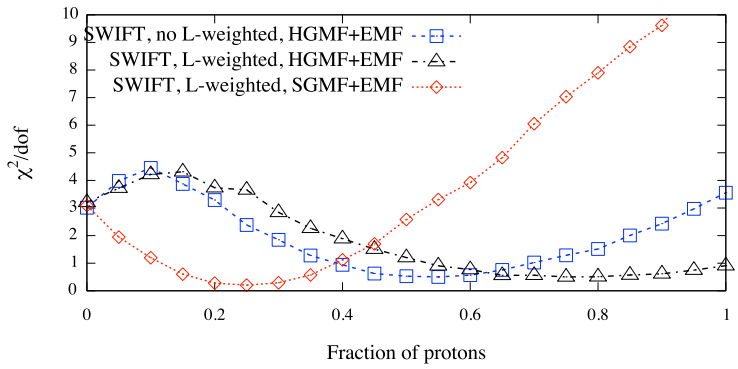


Figure 7.13: Reduced χ^2 statistics as a function of the fraction of protons, for the three scenarios considered in this section.

clustering analysis, in order to compare such results with the data, we make use of the χ^2 statistics defined by Eq. 7.1, where $s_{\rm data}(\Theta)$ is the correlation signal estimated from the data, $\langle s_{\rm MC}(\Theta)\rangle$ is the average correlation signal obtained from a scenario with a fraction of protons f_p and $\sigma_{\rm MC}(\Theta)$ is the corresponding dispersion.

In Fig. 7.13 we show the values of the reduced χ^2 as a function of the fraction of protons. The fraction $f_p^\star$ where the χ^2 gets a minimum is the most compatible with the observed correlation. When intrinsic luminosity is taken into account, we find fractions around 80% and 25% for HGMF+EMF and SGMF+EMF scenarios, respectively. If sources are considered with equal intrinsic luminosity, the fraction of protons is around 50% in the HGMF+EMF scenario. However, it is worth remarking that we have the best discrimination power only in the $\mathcal{L}-$weighted - SGMF+EMF model. Conversely, in the other cases, the values of reduced χ^2 vary in a small range (namely, from ≈ 1 to ≈ 4), making more difficult a statistically significant discrimination. Such results are compatible with that ones obtained within the global method, but provide a more precise estimation of the allowed fraction of protons corresponding to each scenario. As in the case of the intrinsic clustering analysis, if a comparison between all reduced χ^2 curves is considered, the SGMF+EMF scenario is the most compatible model with the observed correlation.

The cases where all sources are considered with equally intrinsic luminosity reflect the implicit assumption that AGN closer to Earth have a larger impact on the flux of UHECRs than more distant AGN. In those scenarios, the allowed fraction of protons is expected to be smaller than the fraction obtained from scenarios where the intrinsic luminosity is taken into account. In fact, in no$-\mathcal{L}-$weighted models, closer AGN have the highest weight (because of the z^{-2} factor) and the number of emitting sources is small. Conversely, in $\mathcal{L}-$weighted models, more distant AGN play a role and the number of emitting sources is larger than the previous case. From very simple and elegant arguments it can be shown that, for a fixed number of UHECRs in the sky, the clustering (as well as the correlation) signal increases for decreasing number of sources and vice versa [226]. By reversing such arguments, for a fixed clustering (or correlation) signal a small fraction of clustered events (i.e. protons) is required if the number of emitting sources is small, whereas a larger fraction is required if the number of sources gets larger.

The general conclusion of this particular analysis is that in the case of weaker magnetic fields (i.e. small deflections) the data is compatible with a small fraction of protons, whereas for stronger magnetic fields (i.e. large deflections) a higher fraction of protons, generally larger than 50% (depending on the assumptions about intrinsic luminosity of sources), is required to reproduce the correlation signal. Both conclusions are still compatible with previous analyses, in particular with results obtained from intrinsic clustering and $X_{\rm max}$ predictions.

7.9 Conclusion

In this chapter, we have performed an extended study which takes into account multiple Auger observations, as the elongation rate, the energy spectrum, the intrinsic clustering and the correlation with AGN. By varying the underlying assumptions, as for instance those ones on the mass composition and the intensity of magnetic fields, we have outlined an astrophysical scenario able to explain Auger data from a phenomenological point of view.

We have considered different models of source distribution in the nearby universe (up to 200 Mpc) and we have simulated the production of UHECRs and their propagation in a ΛCDM cosmology, including different configurations of magnetic fields (only the turbulent component, galactic and extragalactic) and composition of events. We have adopted a mixed composition model, by considering skies populated with a fraction f_p of protons and a fraction $1 - f_p$ of iron nuclei. By varying the parameter f_p, we have estimated, through the multiscale autocorrelation function, the intrinsic clustering signal in the arrival directions of events generated according to models and we have exploited the Pierre Auger data to constrain the simulations. The value of $X_{\max}$ above 60 EeV, extrapolated from the data, has been used to select the compatible astrophysical scenario(s) for different hadronic models, whereas the observed flux of UHECRs has been used for probing two plausible injection spectral indices. Finally, the correlation with active galactic nuclei (AGN) within 75 Mpc (from the SWIFT-BAT 58-months catalog) has been considered to further restrict the number of models compatible with the data.

Results can be summarized as follows. We have found two plausible scenarios, corresponding to two different models for galactic and extragalactic magnetic field: more intense GMFs requires a higher fraction of protons (around 50%) in the data to obtain the observed clustering and correlation signals, whereas less intense GMFs requires a smaller fraction of protons (around 20%) to achieve the same result. In the first case, results are in agreement with predictions from SIBYLL and QGSJET hadronic models, whereas in the second case, there is agreement only with predictions from EPOS. In any case, the more likely spectral injection index is 2.4. However, if a global comparison is performed, the SGMF+EMF scenario, together with the corresponding compatible models, is the most suitable candidate to reproduce the observed data.

Moreover, we have found an interesting relationship between the elongation rate and the deflections of UHECRs in the GMF, that needs further investigation.

Conclusions and outlook

The subject of this thesis has been the investigation of the arrival directions of ultra-high energy cosmic rays (UHECRs) detected by the Pierre Auger Observatory. The study of UHECR at Earth cannot prescind from the study of their propagation in the Universe. We have developed an ad hoc Monte Carlo code (HERMES) for the realistic simulation of UHECR propagation (Chaps. 2 and 3). For the search for clustering in the observed arrival directions and for a correlation with astrophysical objects considered as possible UHECR accelerators, we have introduced a novel method, a multi-scale one (Chap. 4). Besides for the actual search of anisotropies, we have used such method to infer the nature of the sources and their density (Chaps. 5 and 6). Finally, we have considered the obtained results on anisotropy in the more general frame of Auger results (including the observed energy spectrum and the inferred mass composition) to constrain astrophysical scenarios compatible at the same time with these three main observables (Chap. 7). Into more details:

- In chapter 2 and 3 we have described the HERMES propagation code, presenting the modeling we have adopted for i) the cosmological framework, ii) the cosmic background radiation (microwave, infrared/optical and radio), iii) the regular component of the Galactic magnetic field and the irregular component of both the Galactic and the extragalactic magnetic fields, iv) the cross sections describing the interactions between UHE nuclei and photons of extragalactic background radiation, v) the production of secondary particles because of such interactions.

- In the same chapters we have shown several simulations of nuclei propagating in a magnetized Universe and in our Galaxy, by varying the relevant parameters, putting in evidence impact of magnetic fields and energy-loss processes. We have included all the relevant energy losses, as the adiabatic loss (due to the expansion of the Universe), the pair and photo-pion production, and, in the particular case of heavy nuclei, the photo-disintegration processes. We have estimated the surviving probability of UHECRs as a function of their energy and their propagation distance, and we have investigated the impact of energy-loss processes on the propagation of nuclei and, in particular, on the GZK horizon of UHE protons. The agreement between our results and those ones obtained with other simulator available to the community is remarkable.

- In chapter 4 we have introduced a new multiscale method for the investigation of the anisotropy in the arrival direction distribution of UHECR events. It has been designed to perform both catalog-independent and catalog-dependent analyses, and it is based on information theory and extreme value statistics. We have shown that the multiscale method is a competing tool for the study of both small and large scale anisotropies and correlations, providing a great discrimination power even in presence of a strong background contamination and for quite different astrophysical scenarios. As a possible application of our method, we have shown how to probe the Hubble parameter with clustering analysis, by assuming that AGN in the nearby Universe are the sources of UHECRs.

- In chapter 5 we have extensively used multiscale methods to search for an anisotropy signal in

Auger UHECR data, as well as for a correlation with candidate objects in astronomical catalogues. We have investigated the influence of angular and energy uncertainties on both the anisotropy and the correlation signals of events detected with the Pierre Auger Observatory. We have shown that such uncertainties have an impact on the estimated signal. In particular, the energy resolution significantly affects the clustering, whereas the angular resolution negligibly affects such studies.

- In chapter 6 we have investigated several astrophysical scenarios that could be responsible for the observed clustering and correlation in Auger data. We have performed several different studies to estimate the density of UHECR sources by varying the underlying hypotheses on the mass composition, the intensity of magnetic fields and distribution of sources. We have found that the lower bound on the source number density is of the order of 10^{-4} Mpc^{-3} at 95% confidence level, hence excluding some classical candidate astrophysical objects as BL Lac and colliding galaxies, in favor of active galactic nuclei (AGN) and gamma-ray bursts (GRB). Additionally, we have explored the possibility that nearby black holes are the sources of UHECRs detected with the Auger Observatory. We have found that the mass of black holes plays an important role in anisotropy analysis. In fact, the luminosity of black holes is proportional to their mass, as well as the density of such objects in the nearby Universe, and both quantities have a direct impact on the clustering of UHECRs. Our results suggest that, even in the more extreme cases, black holes with mass smaller than $\approx 10^{7.3} M_\odot$ or larger than $\approx 10^{8.75} M_\odot$ are unlikely to be the only sources of UHE protons. Conversely, black holes with mass larger than $M_{\rm BH}$, with $10^{7.3} < M_{\rm BH}/M_\odot < 10^{8.75}$ are candidate sources of UHE protons observed at Earth above 60 and 80 EeV.

- Finally in chapter 7 we have performed a more extended study, taking into account the two other most relevant observations of the Pierre Auger Observatory, namely the average maximum of the shower development and the energy spectrum. By varying the underlying assumptions, as for instance those ones on the mass composition and the intensity of magnetic fields, we have simulated the production of UHECRs and their propagation in a ΛCDM cosmology. Hence, we have outlined the astrophysical scenarios able to explain Auger data from a phenomenological point of view. We have found two plausible scenarios, corresponding to two different models for galactic and extragalactic magnetic field: more intense GMFs requires a higher fraction of protons (around 50%) in the data to obtain the observed clustering and correlation signals, whereas less intense GMFs requires a smaller fraction of protons (around 20%) to achieve the same result. In the first case, results are in agreement with predictions from SIBYLL and QGSJET hadronic models, whereas in the second case, there is agreement only with predictions from EPOS. In any case, the more likely spectral injection index is 2.4. However, if a global comparison is performed, the scenario with a weak turbulent GMF, together with the corresponding compatible models, is the most suitable candidate to reproduce the observed data.

As widely discussed in the present thesis, a final answer about the origin and the composition of UEHCRs is still missing. The analyses presented in this thesis have had the aim of shedding more light on the problem of UHECR origin. This has been possible especially thanks to the access to the data of the largest ever built cosmic ray experiment, the Pierre Auger Observatory. Our studies have provided results about the propagation of UHECRs, their clustering and their possible sources in the energy region of highest attraction in astroparticle physics, namely the trans- and super-GZK region (approximatively above $10^{19.7}$ eV). They have also shown that the modeling of both sources and propagation, coupled to clustering and correlation analysis, may be able to open new insights to the solution of the still unsolved UHECR puzzle.

Appendix A: Relativistic kinematics

In this appendix we will briefly derive the main parameters involved in the kinematics of the propagation of UHECRs, calculating the inelasticity factors in collisions and decays in the framework of special relativity. In the following we will use standard units (if not specified otherwise), and a 4-momentum is indicated by $P^\mu \equiv (E/c, \vec{p})$, even if we neglect the subscript μ for simplicity. We will indicate quantities in the center of mass frame (CMF) with superscript $\star$, in the nucleus rest frame (NRF) with superscript $'$ and in the laboratory frame (LF) with no superscript. It is worth remarking that the LF is denoted by the metric which is co-moving with the expanding Universe.

Energy threshold for particle production in $p + \gamma$ interactions

The interaction between a UHE proton and a background photon may be responsible for the production of new particles, if the energy of the proton is large enough. Such an interaction is an efficient mechanism for energy losses leading, for instance, to the GZK effect, where the production of pions, through a Δ resonance, is the main channel (the branching ratio is $\approx 99.4\%$): $p + \gamma \longrightarrow \Delta \longrightarrow N + \pi$. Let E_p and ϵ_γ be the energies in the LF of the proton and the background photon, respectively, and let E_0 be the energy of the proton at the threshold for the Δ resonance. In the LF, we calculate the relativistic invariant

$$s = (P_p + P_\gamma)^2 \overset{\text{head-on}}{=} M_p^2 c^2 + 4 \frac{E_0}{c} \frac{\epsilon_\gamma}{c}, \tag{A.3}$$

which is also equal to $P_\Delta^2 = M_\Delta^2 c^2$, because of the conservation of 4-momentum and we are considering the interaction at the energy threshold. It follows that photo-pion production occurs for $E_p \geq E_0$, i.e. if

$$E_p \geq \frac{M_\Delta^2 c^4 - M_p^2 c^4}{4\epsilon_\gamma}. \tag{A.4}$$

The average energy of a CMB photon is $\approx 6 \times 10^{-4}$ eV: even by considering the most energetic photons, whose energy $\approx 3 \times 10^{-3}$ eV lies in the tail of the Planck distribution, and the lightest $\Delta-$baryon with mass ≈ 1.232 GeV, the photo-pion production is allowed for $E_p \geq 5.3 \times 10^{19}$ eV.

A similar calculation provides the energy threshold for the interaction $p+\gamma \longrightarrow p+e^+ + e^-$, involving the production of a positron/electron pair.

Case $p + \gamma \longrightarrow \Delta \longrightarrow N + \pi$

We will provide some results in the case of protons interacting with background photons, even if the generalization to the case of heavier nuclei is straightforward. In this case N denotes a nucleon, whereas π indicates any pion of the corresponding isospin triplet. Of course, we are assuming that interaction may occur, i.e. that the energy of the primary proton is over threshold.

In the CMF the system behaves like a particle of mass $M^\star = M_p + m_\gamma = M_p$, energy $E_{\rm CM}$, 4-momentum $P^\star = P_p^\star + P_\gamma^\star \equiv (E_p/c + \epsilon_\gamma/c, 0)$ and

$$s \overset{\rm CMF}{=} \left(\frac{E_p}{c} + \frac{\epsilon_\gamma}{c}\right)^2 \overset{\rm NRF}{=} M_p^2 c^2 + 2M_p\epsilon'_\gamma, \tag{A.5}$$

with $\epsilon'_\gamma = \Gamma_p \epsilon_\gamma (1 - \beta\cos\theta)$. Here, Γ_p is the Lorentz factor of the proton in the LF and θ is the collision angle. On the other hand, in the CMF we have $s = |P^\star|^2 = M_p^2 c^2$ and

$$\Gamma_{\rm CM} = \frac{E_{\rm CM}}{M_p c^2} = \frac{E_p + \epsilon_\gamma}{c\sqrt{s}} \simeq \frac{E_p}{c\left(M_p^2 c^2 + 2M_p\epsilon'_\gamma\right)^{\frac{1}{2}}}. \tag{A.6}$$

The energy of each product, namely the nucleon and the pion, can be determined in the CMF because of 4-momentum conservation $P^\star = P_p^\star + P_\gamma^\star = P_N^\star + P_\pi^\star$, from which we obtain

$$\begin{aligned} P_N^\star &= P^\star - P_\pi^\star \\ P_\pi^\star &= P^\star - P_N^\star. \end{aligned}$$

Now, we square both sides of the above equations, obtaining

$$M_{N,\pi}^2 c^2 = s + M_{\pi,N}^2 c^2 - 2\sqrt{s}\frac{E_{\pi,N}^\star}{c}\left(1 - \beta_{N,\pi}\cos\theta_{N,\pi}\right), \tag{A.7}$$

in a compact notation. By averaging over $\theta_{N,\pi}$ in the interval $[0, \pi]$, we obtain

$$\langle E_{N,\pi}^\star \rangle = c\left(\frac{s + M_{N,\pi}^2 c^2 - M_{\pi,N}^2 c^2}{2\sqrt{s}}\right), \tag{A.8}$$

where

$$s = M_p^2 c^2 + 2M_p\Gamma_p\epsilon_\gamma(1 - \beta\cos\theta) \quad \text{and} \quad \Gamma_p = \frac{E_p}{M_p c^2}, \tag{A.9}$$

that is in perfect agreement with the result reported by Stecker [181]. If we use the approximation that all collisions are head-on (that is not true, on average), i.e. $\theta = \pi$, in the LF we obtain

$$\begin{aligned} \langle E_{N,\pi} \rangle &= \Gamma_{\rm CM}\langle E_{N,\pi}^\star \rangle \\ &= \frac{1}{2}E_p\left(1 + \frac{M_{N,\pi}^2 c^4 - M_{\pi,N}^2 c^4}{M_p^2 c^4 + 4E_p\epsilon_\gamma}\right), \end{aligned} \tag{A.10}$$

that is the average energy carried by the resulting nucleon ($\langle E_N \rangle$) and pion ($\langle E_\pi \rangle$). The inelasticity factor $\kappa(E_p, \theta)$ is defined as the average fractional energy $\Delta E_p/E_p$ lost during the interaction:

$$\kappa(E_p, \theta) = \frac{1}{2}\left(1 - \frac{M_{N,\pi}^2 c^4 - M_{\pi,N}^2 c^4}{M_p^2 c^4 + 2E_p\epsilon_\gamma(1 - \cos\theta)}\right). \tag{A.11}$$

Case $N + \gamma \longrightarrow \Delta + \pi \longrightarrow N\tilde{\pi}\pi$

This interaction is directly involved in the multi-pion production process, providing more than a pion at the end of the collision. As in the previous case, from similar arguments we obtain the energy carried out by the $\Delta-$baryon and the pion in the LF:

$$\langle E_{\Delta,\pi} \rangle = \frac{1}{2} E_N \left(1 + \frac{M_{\Delta,\pi}^2 c^4 - M_{\pi,\Delta}^2 c^4}{M_N^2 c^4 + 4 E_N \epsilon_\gamma} \right). \tag{A.12}$$

After such an interaction, the $\Delta-$baryon decays to a nucleon $\tilde{N}$ and a pion $\tilde{\pi}$, where the $\sim$ symbol allows us to distinguish the pion produced in such a decay from the pion produced in the original interaction. In the CMF we have $s = M_\Delta^2 c^2$, $\Gamma_{\rm CM} = E_\Delta / c\sqrt{s}$ and

$$P_{\tilde{N},\tilde{\pi}}^\star = P_\Delta^\star - P_{\tilde{\pi},\tilde{N}}^\star, \tag{A.13}$$

from which, after some algebra, we obtain the average energy of the decay products in the LF:

$$\langle E_{\tilde{N},\tilde{\pi}} \rangle = \frac{1}{2} E_\Delta \left(1 + \frac{M_{\tilde{N},\tilde{\pi}}^2 - M_{\tilde{\pi},\tilde{N}}^2}{M_\Delta^2} \right). \tag{A.14}$$

Case $\pi^0 \longrightarrow \gamma + \gamma$

This channel is relevant for the production of GZK photons. By following a procedure similar to that previously described, in the CMF we have $P^\star \equiv \left(E_{\pi^0}^\star / c, 0 \right)$, $\sqrt{s} = E_{\pi^0}^\star / c = M_{\pi^0}^2 c^2$ and from conservation of 4-momentum:

$$P_{\gamma_1,\gamma_2}^\star = P^\star - P_{\gamma_2,\gamma_1}^\star, \tag{A.15}$$

from which $\langle E_{\gamma_1,\gamma_2}^\star \rangle = \frac{c}{2}\sqrt{s}$. In the LF we simply obtain $\langle E_\gamma \rangle = E_\pi / 2$ for both photons produced by the decay of the neutral pion. Of course, it is worth remarking that such a result is only valid on average.

Case $\pi \longrightarrow \mu + \nu_\mu$

This channel is relevant for the production of GZK neutrinos. By following the same procedure described in the case $\pi^0 \longrightarrow \gamma\gamma$, and by assuming a negligible mass for the neutrino ($M_\nu \approx 0$), in the LF we simply obtain

$$\langle E_{\mu,\nu_\mu} \rangle = \frac{1}{2} E_\pi \left(1 \pm \frac{M_\mu^2}{M_\pi^2} \right). \tag{A.16}$$

Successively, the muon undergoes a 3-body decay, with the additional production of 2 neutrinos: $\mu^- \longrightarrow e^- \overline{\nu}_e \nu_\mu$ or $\mu^+ \longrightarrow e^+ \nu_e \overline{\nu}_\mu$, depending on muon charge. This is a pure leptonic decay involving no mixing, and if the electron (positron) energy is assumed to be much larger than its rest mass, the corresponding approximated energy spectrum can be estimated by

$$\mathcal{S}(\varepsilon) d\varepsilon = \frac{G_F^2 m_\mu^2}{192\pi^3} (3 - 2\varepsilon) \varepsilon^2 d\varepsilon, \tag{A.17}$$

where G_F is the Fermi coupling constant and $\varepsilon = 2E_e / m_\mu$, in natural units. However, the correct energy spectrum, known as Michel spectrum, is given by

$$\mathcal{S}(\varepsilon) d\varepsilon = \frac{G_F^2 m_\mu^2}{16\pi^3} \left[1 - \varepsilon - \frac{2}{9} \rho (3 - 4\varepsilon) \right] \varepsilon^2 d\varepsilon, \tag{A.18}$$

taking into account the most general interaction among the four fermions and where ρ is the Michel parameter [315, 316].

Appendix B: Probability, Statistics and Information theory

The present appendix should be useful for a better understanding of the mathematical and physical meaning of information quantities involved in the multiscale approaches introduced in this work, as well as the basic definitions generally adopted in probability and statistics.

B.1 Probability and Statistics hints

Definition 1 (Expectation value) *The* **expectation value** *of a stochastic variable X is defined as*

$$E(X) = \int_{\Omega} X dPr$$

If X has a PDF $f(x)$, then

$$E(X) = \int_{\mathbb{O}} x f(x) dx$$

Definition 2 (Moments) *The $n-th$* **moment** *of a stochastic variable X about $\hat{x} \in \mathbb{O}$, is defined as*

$$\mu_p(X) = E(X^n) = \int_{\Omega} D(X(w), \hat{x})^n dPr$$

where D is a metric. If X has a PDF $f(x)$, then

$$\mu_p(X) = E(X^p) = \int_{\mathbb{O}} (x - \hat{x})^n f(x) dx$$

Definition 3 (Test statistics) *Given a random sample $\{X_n\} = X_1, X_2, ..., X_n$ of a $d-$dimensional stochastic variable $\mathbb{X}^{(d)}$, a* **test statistics** *$\mathcal{E}$ is the mapping*

$$\mathcal{E} : \mathbb{X}^{(d)} \to \mathbb{D} \subseteq \mathbb{R} : \{\{X_n\} \to \mathcal{E}\left(\{D_m\}\right)\}$$

where $\mathbb{R}$ is the set of real numbers and $\{D_m\} = D_1(\{X_n\}), D_2(\{X_n\}), ..., D_m(\{X_n\})$ are real parameters defined in the spaces $\mathbb{D}_1, \mathbb{D}_2, ..., \mathbb{D}_m$, respectively. Parameters define the $m-$dimensional parameter space

$$\mathbb{S} = \mathbb{D}_1 \otimes \mathbb{D}_2 \otimes ...\mathbb{D}_m$$

In general $\mathbb{D}$ is an interval of real numbers, open or closed, bounded or unbounded.

Definition 4 (Bias) *Let $\hat{\mathcal{E}}$ be an estimator of the test statistics $\mathcal{E}$: the* **bias** *on $\mathcal{E}$ is defined as*

$$b\left[\hat{\mathcal{E}}\right] = E\left[\hat{\mathcal{E}}\right] - \mathcal{E}$$

An estimator is unbiased *if $b[\hat{\mathcal{E}}] = 0$.*

Sellke *et al.* [221] presented a rigorous estimation of the lower limit on significance (type I error) from a p-value, which does not require assumptions on the distribution of the p-value and an explicit alternative hypothesis.

Theorem 1 (Mapping p-value to significance) *If, for the p-value, the relation $P < e^{-1}$ holds, then the lower limit to the significance of the test $\mathcal{T}$ is given by*

$$\alpha(P) \geq \left(1 - \frac{1}{eP\log P}\right)^{-1}$$

Proof 1 *We refer to [221] for the proof.*

Often, several tests are performed on the same data. The following two theorems establish how to combine results to obtain the corresponding significance.

Theorem 2 (Fisher's method to combine p-values) *Let $P_i(\mathcal{E})$ be the p-value corresponding to the statistical test T_i $(i = 1, 2, ..., s)$ of a null hypothesis $\mathcal{H}_0$ through the test statistics $\mathcal{E}$. The combined test statistics*

$$\mathcal{F}^2 = -2\ln\left(\prod_{i=1}^{s} P_i(\mathcal{E})\right) = -2\sum_{i=1}^{s}\ln P_i(\mathcal{E}) \tag{B.19}$$

has a χ^2 distribution with $2s$ degrees of freedom if $\mathcal{H}_0$ is true for each test.

Proof 2 *The sketch of the proof is as follows [317]. If $\mathcal{H}_0$ is true for the $i-$th test, then $P_i(\mathcal{E}) \in \mathcal{U}(0,1)$, $-\ln P_i(\mathcal{E})$ follows an exponential distribution and $-2\ln P_i(\mathcal{E})$ follows a χ^2 distribution with 2 degrees of freedom.*

The sum of s independent χ^2 distributions, each with 2 degrees of freedom, yields a χ^2 distribution with $2s$ degrees of freedom.

Theorem 3 (Stouffer's Z-scores method) *Let $P_i(\mathcal{E})$ be the p-value corresponding to the statistical test T_i $(i = 1, 2, ..., s)$ of a null hypothesis $\mathcal{H}_0$ through the test statistics $\mathcal{E}$, and define $Z_i = F_{\mathcal{N}}^{-1}(1 - P_i(\mathcal{E}))$, where $F_{\mathcal{N}}$ is the standard normal CDF. The combined test statistics*

$$\mathcal{Z} = \frac{1}{\sqrt{s}}\sum_{i=1}^{s} Z_i$$

is called Z-score *and it follows a standard normal distribution if $\mathcal{H}_0$ is true for each test. If w_i is the weight corresponding to Z_i, the* weighted Z-score *is defined as*

$$\tilde{\mathcal{Z}} = \frac{1}{S}\sum_{i=1}^{s} w_i Z_i, \quad S = \sqrt{\sum_{i=1}^{s} w_i^2} \tag{B.20}$$

and, again, it follows a standard normal distribution if $\mathcal{H}_0$ is true for each test.

Definition 5 (Maximum Likelihood Estimation) *Let $\mathbb{X}^{(d)}$ be a $d-$dimensional stochastic variable with continuous density $f(\mathbf{x};\mathbf{t})$, where $\mathbf{x}$ represents a $d-$dimensional outcome of $\mathbb{X}^{(d)}$ and $\mathbf{t}$ is the set of m real parameters (to be estimated) of a statistical parameterized model. If n indipendent outcomes are available, the* **likelihood function***, defined as*

$$\mathcal{L}(\mathbf{t}) = \prod_{i=1}^{n} f(\mathbf{x}_i|\mathbf{t}) \tag{B.21}$$

is maximized by the estimates $\hat{\mathbf{t}}$ of the parameters $\mathbf{t}$. In many practical applications, the negative log-likelihood *function*

$$L(\mathbf{t}) = -\ln \mathcal{L}(\mathbf{t}) = -\sum_{i=1}^{n} \ln f(\mathbf{x}_i|\mathbf{t})$$

is used instead of $\mathcal{L}(\mathbf{t})$, where minimization is used instead of maximization.

The likelihood method (or MLE) allows to estimate, through the likelihood function, the unknown parameters of a statistical model if the outcomes are known. It follows that, in general, $\mathbf{t} = \mathbf{t}(\mathbf{x}_1, \mathbf{x}_2, ..., \mathbf{x}_n)$.

Example. If n times t_i are measured, they should be distributed with density $f(t;\tau) = 1/\tau \exp(-t/\tau)$ whose likelihood function is

$$L(\tau) = -\sum_{i=1}^{n} \left(\ln \frac{1}{\tau} - \frac{t_i}{\tau} \right)$$

By minimizing $L(\tau)$, the estimate

$$\hat{t} = \frac{1}{n} \sum_{i=1}^{n} t_i$$

which is unbiased because of $\mathrm{E}[\hat{t}] = \tau$. If density is $f(t;\tau) = \lambda \exp(-\lambda t)$, then the estimate is $\hat{\lambda} = 1/\hat{\tau}$, but $\mathrm{E}[\hat{\lambda}] = n/(n-1)\lambda$, i.e. the estimate is biased [318].

It can be shown that, under very general assumptions, $\hat{t}$ are asymptotical normally distributed with minimal variance Indeed, the maximum likelihood solution is invariant under the change of parameters but the bias for finite n may be different for different transformations of the parameter [318].

Theorem 4 (Fisher's information) *Given a parameterized statistical model, if $\hat{\mathbf{t}} = \hat{\mathbf{t}}(\mathbf{x}_1, \mathbf{x}_2, ..., \mathbf{x}_n)$ are unbiased estimations of parameters, the* **Fisher's information***, defined as*

$$I(t_i, t_j) = E\left[\frac{\partial \ln \mathcal{L}(\mathbf{t})}{\partial t_i} \frac{\partial \ln \mathcal{L}(\mathbf{t})}{\partial t_j} \right] = -E\left[\frac{\partial^2 \ln \mathcal{L}(\mathbf{t})}{\partial t_i \partial t_j} \right]$$

gives a lower limit to the variance of $\hat{\mathbf{t}}$ through the Rao-Cramer-Frechet inequality

$$\boldsymbol{V}\left[\hat{\mathbf{t}}\right] \geq \boldsymbol{I}^{-1}$$

Axiom 1 (Likelihood principle) *In the statistical inference about a set of parameters $\mathbf{t}$ of a given parameterized statistical model, after the observation of $\mathbf{x}$, all relevant information about the sample is contained in the likelihood function for the observed $\mathbf{x}$. Two different likelihood functions contain the same information about $\mathbf{x}$ if they are proportional to each other.*

The MLE satisfies the likelihood principle.

Definition 6 (Likelihood ratio) *Given a statistical test $\mathcal{T}$ of a null hypothesis $\mathcal{H}_0$ (with g_0 degrees of freedom) against an alternative hypothesis $\mathcal{H}_1$ (with g_1 degrees of freedom), the* **likelihood ratio**

$$\Lambda(\mathbf{x}) = -2\left(\ln \mathcal{L}(\mathbf{t}|\mathcal{H}_0) - \ln \mathcal{L}(\mathbf{t}|\mathcal{H}_1)\right) = -2\ln\left[\frac{\mathcal{L}(\mathbf{t}|\mathcal{H}_0)}{\mathcal{L}(\mathbf{t}|\mathcal{H}_1)}\right]$$

is a test statistics following a χ^2 distribution with $g_0 - g_1$ d.of.

In general, the likelihood ratio test statistics must be treated as a common test statistics and the estimation of p-values is required for hypothesis testing.

Theorem 5 (Neyman-Pearson lemma) *Given a statistical test $\mathcal{T}$ of a null hypothesis $\mathcal{H}_0$ against an alternative hypothesis $\mathcal{H}_1$, the likelihood ratio test*

$$\Lambda(\mathbf{x}) \leq \lambda, \quad Pr\left(\Lambda(\mathbf{x}) \leq \lambda|\mathcal{H}_0\right)$$

rejecting $\mathcal{H}_0$ (with probability ξ if $\Lambda = \lambda$) in favor of $\mathcal{H}_1$, is the most powerful test of size α for the threshold λ, where

$$\alpha = \xi \times Pr(\Lambda = \lambda|\mathcal{H}_0) + Pr(\Lambda < \lambda|\mathcal{H}_0)$$

An alternative to classical hypothesis testing described so far, is the Bayesian decision making process, based on Bayes factors.

Definition 7 (Bayes factor) *A random sample $\mathbf{X} = \{X_n\}$, of a $d-$dimensional stochastic variable $\mathbb{X}^d$, is assumed to arise or under a null hypothesis $\mathcal{H}_0$ either an alternative hypothesis $\mathcal{H}_1$ with probabilities $p(\mathbf{X}|\mathcal{H}_0)$ and $p(\mathbf{X}|\mathcal{H}_1)$, respectively. Given* a priori *probabilities such that $p(\mathcal{H}_0) = 1 - p(\mathcal{H}_1)$, the sample produces* a posteriori *probabilities $p(\mathcal{H}_0|\mathbf{X}) = 1 - p(\mathcal{H}_1|\mathbf{X})$. The trasformation of the prior opinion to the posterior opinion is given by the* **Bayes factor** $\mathcal{B}$.

From Bayes' theorem:

$$p(\mathcal{H}_k|\mathbf{X}) = \frac{p(\mathbf{X}|\mathcal{H}_k)p(\mathcal{H}_k)}{p(\mathbf{X}|\mathcal{H}_0)p(\mathcal{H}_0) + p(\mathbf{X}|\mathcal{H}_1)p(\mathcal{H}_1)}, \quad k = 0, 1$$

it follows:

$$\frac{p(\mathcal{H}_0|\mathbf{X})}{p(\mathcal{H}_1|\mathbf{X})} = \frac{p(\mathbf{X}|\mathcal{H}_0)}{p(\mathbf{X}|\mathcal{H}_1)}\frac{p(\mathcal{H}_0)}{p(\mathcal{H}_1)}$$

where

$$\mathcal{B} = \frac{p(\mathbf{X}|\mathcal{H}_0)}{p(\mathbf{X}|\mathcal{H}_1)}$$

is the Bayes' factor which relates prior and posterior odds. If $\mathcal{H}_0$ and $\mathcal{H}_1$ are simple hypotheses (i.e. they are single distributions with no free parameters), $\mathcal{B} = \Lambda(\mathbf{X})$. If hypotheses depend on unknown parameters, the probabilities in the Bayes' factor can be obtained by integrating over the parameter space $\mathbb{S}$:

$$p(\mathbf{X}|\mathcal{H}_k) = \int_{\mathbb{S}} p(\mathbf{X}|\mathbf{t}_k, \mathcal{H}_k)\pi(\mathbf{t}_k|\mathcal{H}_k)d\mathbf{t}_k, \quad k = 0, 1$$

where $\mathbf{t}_k$ are the parameters for the hypothesis $\mathcal{H}_k$, $\pi(\mathbf{t}_k|\mathcal{H}_k)$ is their prior probability and $p(\mathbf{X}|\mathbf{t}_k, \mathcal{H}_k)$ is the likelihood function. We refer to [319] (and ref. therein) for further information.

From the above equation, it follows that at variance with classical hypothesis testing, the Bayesian decision making process considers the probability of a model by taking into account all possible values of parameters. Interpretation of the Bayes factor can be found in [320], where the strength of evidence is reported versus $\mathcal{B}$. The definition of the Bayes' factor is elegant and takes into account information carried out by all parameters; on the other hand setting priors is a well known problem and the main reason of criticism.

B.2 Information theory hints

Information theory is the branch of mathematics dealing with the efficient and accurate storage, transmission, and representation of information [321]. The key quantities of information theory are information and entropy, defined in the following.

Definition 8 (Information) *Let $\mathbb{X}^{(d)}$ be a $d-$dimensional stochastic variable with density $f(\mathbf{x})$, where $\mathbf{x}$ represents a $d-$dimensional outcome of $\mathbb{X}^{(d)}$. The* **information** *associated with $\mathbf{x}$ is defined by*

$$I(\mathbf{x}) = \ln\left(\frac{1}{f(\mathbf{x})}\right) = -\ln f(\mathbf{x})$$

The information unit depends on the logarithm base used: for base 2 unit is bit, for unit e unit is nat.

Definition 9 (Shannon entropy) *Let $\mathbb{X}^{(d)}$ be a $d-$dimensional stochastic variable with density $f(\mathbf{x})$, where $\mathbf{x}$ represents a $d-$dimensional outcome of $\mathbb{X}^{(d)}$. The* **Shannon entropy** *(or information entropy) of $\mathbb{X}^{(d)}$ is defined by*

$$H(\mathbb{X}^{(d)}) = E\left[I(\mathbb{X}^{(d)})\right] = -\int_{\mathbb{O}} f(\mathbf{x}) \ln f(\mathbf{x}) d\mathbf{x}$$

Because of the above definition, in the following we will refer to the entropy of a stochastic variable $H(\mathbb{X}^{(d)})$ or to the entropy of its density $H(f(\mathbf{x}))$ with no difference.

Definition 10 (Joint entropy) *Let $\mathbb{X}^{(d)}, \mathbb{Y}^{(d)}, ..., \mathbb{Z}^{(d)}$ be $d-$dimensional stochastic variables with joint density $f(\mathbf{x}, \mathbf{y}, ..., \mathbf{z})$, where $\mathbf{x}$, $\mathbf{y}$, ..., $\mathbf{z}$ represent a $d-$dimensional outcome of $\mathbb{X}^{(d)}$, $\mathbb{Y}^{(d)}$, ..., $\mathbb{Z}^{(d)}$, respectively. The* **joint entropy** *of the stochastic variables is defined by*

$$H(\mathbb{X}^{(d)}, \mathbb{Y}^{(d)}, ..., \mathbb{Z}^{(d)}) = -\int_{\mathbb{O}} f(\mathbf{x}, \mathbf{y}, ..., \mathbf{z}) \ln f(\mathbf{x}, \mathbf{y}, ..., \mathbf{z}) d\mathbf{\Omega}$$

where $\mathbf{\Omega} = d\mathbf{x}d\mathbf{y}...d\mathbf{z}$ and $\mathbb{O}$ is the total observation space.

Definition 11 (Statistical divergence) *Given a probability space $\mathcal{P}$, the* **statistical divergence** *of the density f to the density g is the mapping $\mathcal{D} : \mathcal{P}^2 \longrightarrow \mathbb{R}$ such that*

- $\mathcal{D}(f||g) \geq 0 \; \forall f, g \in \mathcal{P}$;
- $\mathcal{D}(f||g) = 0 \Leftrightarrow f = g$.

In general the divergence is not simmetric and a dual divergence *can be defined as $\mathcal{D}^*(g||f) = \mathcal{D}(f||g)$.*

In many practical applications, the statistical divergence is defined by mean of a function φ

$$\mathcal{D}_{\varphi}(f||g) = \int_{\mathbb{O}} f(\mathbf{x})\phi\left[f(\mathbf{x}), g(\mathbf{x})\right] d\mathbf{x}$$

and it is used to establish how much two densities are *statistically similar* each other.

Definition 12 (Kullback-Leibler divergence) *Given two densities f and g of a $d-$dimensional stochastic variable $\mathbb{X}^{(d)}$, both defined on the probability space $\mathcal{P}$, the* **Kullback-Leibler divergence** *(or* information gain*) is defined as*

$$\mathcal{D}_{KL}(f||g) = \int_{\mathbb{O}} f(\mathbf{x}) \ln \frac{f(\mathbf{x})}{g(\mathbf{x})} d\mathbf{x}$$

Information, Shannon entropy, conditional and cross entropy can be easily derived from the Kullback-Leibler divergence.

The Kullback -Leibler divergence is the *coding penalty* of selecting the density g to approximate the density f [275]: it is strictly related to the average likelihood $\overline{\mathcal{L}}(f|g)$ which expresses the probability of observing a density f assumed that a model g is true.

Theorem 6 *Let $\mathbb{X}^{(d)}$ be a $d-$dimensional stochastic variable with continuous density $g(\mathbf{x};\mathbf{t})$, where $\mathbf{x}$ represents a $d-$dimensional outcome of $\mathbb{X}^{(d)}$ and $\mathbf{t}$ is the set of m real parameters of a statistical parameterized model. If n indipendent outcomes are available and $\mathbf{t}_0$ is the set of true density (presumed unknown and to be estimated through MLE) for the true density $f(\mathbf{x};\mathbf{t}_0)$, then*

$$\mathcal{D}_{KL}(f||g) = -\left(\overline{L}(\mathbf{t}) + H(f)\right)$$

where $H(f) = H(\mathbb{X}^{(d)}|\mathbf{t}_0)$ is the Shannon entropy of the true density and

$$\overline{L}(\mathbf{t}) = \lim_{n\longrightarrow\infty} \frac{1}{n} \ln \mathcal{L}(\mathbf{t})$$

is the average log-likelihood.

Proof 3 *By definition of likelihood function, it follows:*

$$\begin{aligned}
\frac{1}{n} \ln \mathcal{L}(\mathbf{t}) &= \frac{1}{n} \sum_{i=1}^{n} \ln g(\mathbf{x}_i|\mathbf{t}) \\
\lim_{n\longrightarrow\infty} \frac{1}{n} \ln \mathcal{L}(\mathbf{t}) &= \int_{\mathbb{O}} f(\mathbf{x}|\mathbf{t}_0) \ln g(\mathbf{x}|\mathbf{t}) d\mathbf{x}
\end{aligned}$$

where we heuristically assumed that in the limit of an infinite number of experiments, the sampling density g equals the relative frequencies of $\mathbf{x}_i$. The Shannon entropy of the true density is

$$H(f) = H(\mathbb{X}^{(d)}|\mathbf{t}_0) = -\int_{\mathbb{O}} f(\mathbf{x}|\mathbf{t}_0) \ln f(\mathbf{x}|\mathbf{t}_0) d\mathbf{x}$$

and it follows:

$$\begin{aligned}
\lim_{n\longrightarrow\infty} \frac{1}{n} \ln \mathcal{L}(\mathbf{t}) + H(f) &= \int_{\mathbb{O}} f(\mathbf{x}|\mathbf{t}_0) \ln \frac{g(\mathbf{x}|\mathbf{t})}{f(\mathbf{x}|\mathbf{t}_0)} d\mathbf{x} \\
&= -\int_{\mathbb{O}} f(\mathbf{x}|\mathbf{t}_0) \ln \frac{f(\mathbf{x}|\mathbf{t}_0)}{g(\mathbf{x}|\mathbf{t})} d\mathbf{x} \\
&= \mathcal{D}_{KL}(f||g)
\end{aligned}$$

that ends the proof.

Axiom 2 (Maximum Entropy Principle) *Given some testable information, the best density representing the current state of knowledge is the one with largest information entropy.*

We refer to [322] for further information about the above theorem and to [323–325] for an extensive discussion about the maximum entropy principle.

Definition 13 (Mutual information) *Given two $d-$dimensional stochastic variables $\mathbb{X}^{(d)}$ and $\mathbb{Y}^{(d)}$ on the same probability space and the density f, their* **mutual information** *is the simmetric measure defined as*

$$\begin{aligned}\mathcal{I}(\mathbb{X}^{(d)};\mathbb{Y}^{(d)}) &= \mathcal{D}_{KL}\left[f(\mathbf{x},\mathbf{y})||f(\mathbf{x})f(\mathbf{y})\right] \\ &= \int_{\mathbb{O}} f(\mathbf{x},\mathbf{y})\ln\frac{f(\mathbf{x},\mathbf{y})}{f(\mathbf{x})f(\mathbf{y})}d\mathbf{\Omega}\end{aligned}$$

The mutual information has the following properties:

- $\mathcal{I}(\mathbb{X}^{(d)};\mathbb{Y}^{(d)}) = 0$ for uncorrelated variables;
- $\mathcal{I}(\mathbb{X}^{(d)};\mathbb{Y}^{(d)}) = \mathcal{I}(\mathbb{Y}^{(d)};\mathbb{X}^{(d)})$;
- $\mathcal{I}(\mathbb{X}^{(d)};\mathbb{Y}^{(d)}) = H(\mathbb{X}^{(d)}) + H(\mathbb{Y}^{(d)}) - H(\mathbb{X}^{(d)},\mathbb{Y}^{(d)})$.

The Jensen-Renyi divergence

$$\mathcal{D}_q(f||g) = -\frac{1}{1-q}\ln\int_{\mathbb{O}}[f(\mathbf{x})]^q[g(\mathbf{x})]^{1-q}d\mathbf{x}$$

where

$$\lim_{q\longrightarrow 1}\mathcal{D}_q(f||g) = \mathcal{D}_{KL}(f||g)$$

is the most general information theoretical measure: with the right assumptions, every other information measure can be derived. We refer to [326–329] for further information.

Bibliography

[1] D. Pacini *Il Nuovo Cimento* **3** (1912) 93–100.

[2] V. Hess *Phys. Zeits.* **13** (1912) 1084.

[3] W. Kolhörster *Phys. Zeits.* **14** (1913) 1153.

[4] W. Kolhörster *Ber. deutsch. Phys. Qes.* **16** (1914) 719.

[5] R. Millikan *Proc. Nat. Acad. Sci. USA* **12** (1926) 48.

[6] R. Millikan and I. Bowen *Phys. Rev.* **27** (1926) 353.

[7] R. Millikan and G. Cameron *Phys. Rev.* **28** (1926) 851.

[8] R. Millikan and R. Otis *Phys. Rev.* **27** (1926) 645.

[9] W. Bothe and W. Kolhörster *Phys. Zeits.* **56** (1929) 751.

[10] J. Clay *Amsterdam Proc.* **33** (1930) 711.

[11] B. Rossi *Nature* **125** (1930) 636.

[12] B. Rossi *Phys. Rev.* **36** (1930) 606.

[13] T. Johnson *Phys. Rev.* **41** (1934) 834.

[14] L. Alvarez and A. Compton *Phys. Rev.* **41** (1930) 835.

[15] B. Rossi *Ricerca Scientifica* **5** (1934) 579.

[16] B. Rossi *Phys. Zeits.* **82** (1933) 151.

[17] G. Clark, *The Contributions of Bruno B. Rossi to Particle Physics and Astrophysics*, in *Atti del XXV Congresso Nazionale di Storia della Fisica e dell'Astronomia*, pp. R1.1–R1.16, 2008.

[18] B. Rossi *Ricerca Scientifica* **5** (1934) 589.

[19] A. Compton *Phys. Rev.* **43** (1933) 387.

[20] R. Millikan and G. Cameron *Phys. Rev.* **32** (1928) 533.

[21] A. Compton and R. Stephenson *Phys. Rev.* **45** (1934) 441.

[22] A. Compton *Proc. Phys. Soc.* **47** (1935) 747.

[23] W. Kolhörster *et al Naturwissenschaften* **26** (1938) 576.

[24] P. Auger *Comptes rendus* **206** (1938) 1721.

[25] A. Letessier-Selvon and T. Stanev *Rev. Mod. Phys.* **83** (2011) 907–942, [arXiv:1103.0031].

[26] J. Blümer, R. Engel, and J. Hörandel *Progr. Part. Nucl. Phys.* **63** (2009) 293–338, [arXiv:0904.0725].

[27] K. Nakamura *et al J. Phys. G* **37** (2010) 075021.

[28] J. Abraham *et al Phys. Lett. B* **685** (2010) 239–246, [arXiv:1002.1975].

[29] K. Greisen *Phys. Rev. Lett.* **16** (1966) 748–750.

[30] G. Zatsepin and V. Kuz'Min *JETP Lett.* **4** (1966) 78.

[31] J. Hoerandel *Astrop. Phys.* **19** (2003) 193–220.

[32] V. Berezinsky, S. Grigorieva, and B. Hnatyk *Astrop. Phys.* **21** (2004) 617–625, [astro-ph/0403477].

[33] P. Biermann and V. de Souza *arXiv* (2011) [arXiv:1106.0625].

[34] V. Ginzburg and S. Syrovatskii *Soviet Phys. JETP* **19** (1964) 1255–1264.

[35] V.L. Ginzburg and S.I. Syrovatskii, *The origin of cosmic rays.* Pergamon Press, Oxford, 1964.

[36] A. Putze, L. Derome, D. Maurin, L. Perotto, and R. Taillet *Astron. Astroph.* **497** (2009) 991–1007.

[37] E. Fermi *Phys. Rev. E* **75** (1949) 1169.

[38] A. Bell *Mont. Not. R. Astron. Soc.* **182** (1978) 147–156.

[39] A. Bell *Mont. Not. R. Astron. Soc.* **182** (1978) 443–455.

[40] P. Bhattacharjee and G. Sigl *Phys. Rep.* **327** (2000) 109–247, [astro-ph/9811011].

[41] A. Hillas *Ann. Rev. Astr. Astrop.* **22** (1984) 425–444.

[42] K. Ptitsyna and S. Troitsky *Phys. Uspekhi* **53** (2010) 691–701, [arXiv:0808.0367].

[43] D. Torres and L. Anchordoqui *Reports on Progress in Physics* **67** (2004) 1663, [astro-ph/0402371].

[44] T. Weiler *Astrop. Phys.* **11** (1999) 303–316, [hep-ph/9710431].

[45] W. Heitler *Proc. R. Soc. Lon. A Math. Phys. Sci.* **166** (1938) 529.

[46] H. Bethe and W. Heitler *Proc. R. Soc. Lon. A Math. Phys. Sci.* **146** (1934) 83.

[47] H. Bhabha and W. Heitler *Proc. R. Soc. Lon. A Math. Phys. Sci.* **159** (1937) 432.

[48] J. Matthews *Astrop. Phys.* **22** (2005) 387–397.

[49] J. Alvarez-Muniz, R. Engel, T. Gaisser, J. Ortiz, and T. Stanev *Phys. Rev. D* **66** (2002) 033011, [astro-ph/0205302].

[50] R. Ulrich, J. Blümer, R. Engel, F. Schüssler, and M. Unger *New J. Phys.* **11** (2009) 065018, [arXiv:0903.0404].

[51] J. Boyer, B. Knapp, E. Mannel, and M. Seman *Nucl. Instr. Meth. A* **482** (2002) 457–474.

[52] T.K. Gaisser and A.M. Hillas *Proc. 15th ICRC* (1977).

[53] J. Linsley *Phys. Rev. Lett.* **10** (1963) 146–148.

[54] D.M. Edge *et al J. Phys. A* **6** (1973) 1612.

[55] D.D. Krasilnikov *et al J. Phys. A* **7** (1974) 176.

[56] R.M. Baltrusaitis *et al Nucl. Instr. Meth. A* **240** (1985) 410–428.

[57] C.J. Bell *et al J. Phys. A* **7** (1974) 990.

[58] N. Chiba *et al Nucl. Instr. Meth. A* **311** (1992) 338–349.

[59] J. Abraham *et al Nucl. Instr. Meth. A* **523** (2004) 50–95.

[60] H. Tokuno *et al Nucl. Instr. Meth. A* **601** (2009) 364–371.

[61] J. Blümer *et al New J. Phys.* **12** (2010) 035001.

[62] Y. Takahashi *et al New J. Phys.* **11** (2009) 065009.

[63] M. Takeda *et al Astrop. Phys.* **19** (2003) 447–462, [astro-ph/0209422v3].

[64] R.U. Abbasi *et al Phys. Rev. Lett.* **92** (2004) 151101.

[65] R.U. Abbasi *et al Astrop. Phys.* **23** (2005) 157–174.

[66] M. Teshima *et al Proc. 27th ICRC, Hamburg* (2001).

[67] M. Takeda *et al Proc. 27th ICRC, Hamburg* (2001).

[68] R.U. Abbasi *et al Astrop. Phys.* **22** (2004) 139–149, [astro-ph/0404366].

[69] R.U. Abbasi *et al Ap. J. Lett.* **610** (2004) L73.

[70] R.U. Abbasi *et al Ap. J.* **623** (2005) 164.

[71] I. Allekotte *et al Nucl. Instr. Meth. A* **586** (2008) 409–420, [arXiv:0712.2832].

[72] J. Abraham *et al Nucl. Instr. Meth. A* **620** (2010) 227–251, [arXiv:0907.4282].

[73] B. Fick *et al J. Instr.* **1** (2006) P11003, [astro-ph/0507334].

[74] S.Y. BenZvi *et al Nucl. Instr. Meth. A* **574** (2007) 171–184, [astro-ph/0609063].

[75] R. Caruso *et al Proc. 22nd ECRS, Turku* (2010).

[76] J. Abraham *et al Nucl. Instr. Meth. A* **613** (2010) 29–39.

[77] K. Kotera and A. Olinto *Ann. Rev. Astron. Astrophys.* **49** (2011) [arXiv:1101.4256].

[78] C. Di Giulio *et al Proc. 31st ICRC, Lodz* (2009).

[79] R. Pesce *et al Proc. 32nd ICRC, Beijing* (2011) [arXiv:1107.4809].

[80] C. Bonifazi, A. Letessier-Selvon, E. Santos, *et. al. Astrop. Phys.* **28** (2008) 523–528, [arXiv:0705.1856].

[81] D. Allard *et al P. Auger GAP 2006* **026** (2006).

[82] B. Dawson, M. Debes and P.Sommers *P. Auger GAP 2001* **016** (2001).

[83] D.Gora *et al. Astrop. Phys.* **24** (2006) 6.

[84] M. Ungher *et al. Nucl. Instr. Meth. A* **588** (2008) 433–441.

[85] M. Giller *et al J. Phys. G* **30** (2004) 97–105.

[86] A.M. Hillas *et al J. Phys. G* **8** (1982) 1475–1492.

[87] F. Nerling *et al Astrop. Phys.* **24** (2006) 6.

[88] J. Abraham *et al Science* **318** (2007) 938, [arXiv:0711.2256].

[89] P. Abreu *et al Astrop. Phys.* **34** (2010) 314–326, [arXiv:1009.1855].

[90] P. Abreu *et al J. Cosm. Part. Phys.* **06** (2011) 022, [arXiv:1106.3048].

[91] P. Facal *et al Proc. 32nd ICRC, Beijing* (2011) [arXiv:1107.4804].

[92] J. Abraham *et al Phys. Rev. Lett.* **104** (2010) 91101, [arXiv:1002.0699].

[93] J. Abraham *et al Phys. Rev. Lett.* **101** (2008) 61101.

[94] R.U. Abbasi *et al Phys. Rev. Lett.* **100** (2008) 101101.

[95] F. Salamida *et al Proc. 32nd ICRC, Beijing* (2011) [arXiv:1107.4809].

[96] R. Ullrich *et al Proc. 32nd ICRC, Beijing* (2011) [arXiv:1107.4804].

[97] M. Settimo *et al Proc. 32nd ICRC, Beijing* (2011) [arXiv:1107.4805].

[98] J. Abraham *et al Astrop. Phys.* **31** (2009) 399–406, [arXiv:0903.1127].

[99] Y. Guardincerri *et al Proc. 32nd ICRC, Beijing* (2011) [arXiv:1107.4805].

[100] J. Alvarez-Muniz, R. Engel, and T. Stanev *Ap. J.* **572** (2002) 185, [astro-ph/0112227v1].

[101] A. Brandenburg and K. Subramanian *Phys. Rep.* **417** (2005) 1–209.

[102] L. Biermann *Zeits. Natur. Teil A* **5** (1950) 65.

[103] L. Mestel and I. Roxburgh *Ap. J.* **136** (1962) 615.

[104] K. Subramanian, D. Narasimha, and S. Chitre *Month. Not. R. Astron. Soc.* **271** (1994) L15.

[105] L. Widrow *Rev. Mod. Phys.* **74** (2002) 775, [astro-ph/0207240v1].

[106] V. Semikoz and D. Sokoloff *Phys. Rev. Lett.* **92** (2004) 131301, [astro-ph/0312567].

[107] D. Ryu, H. Kang, J. Cho, and S. Das *Science* **320** (2008) 909, [arXiv:0805.2466].

[108] R. Beck *Plasma Science, IEEE Trans. on* **14** (1986) 740–747.

[109] R. Beck *Phil. Trans. R. Soc. Lond. A* **358** (2000) 777.

[110] R. Beck *Astroph. Space Sci.* **289** (2004) 293–302, [astro-ph/0212288].

[111] D. Harari, S. Mollerach, E. Roulet, and F. Sanchez *J. High En. Phys.* **2002** (2002) 045, [astro-ph/0202362v2].

[112] J. Giacalone and J. Jokipii *Ap. J.* **430** (1994) L137–L140.

[113] J. Giacalone and J. Jokipii *Ap. J.* **520** (1999) 204.

[114] F. Casse, M. Lemoine, and G. Pelletier *Phys. Rev. D* **65** (2001) 023002, [astro-ph/0109223v1].

[115] G. Giacinti, M. Kachelriess, D. V. Semikoz, G. Sigl *Astrop. Phys.* **35** (2011) 192–200, [arXiv:1104.1141].

[116] A. Fraser and H. Swinney *Phys. Rev. A* **33** (1986) 1134.

[117] R. Kubo, M. Toda, and N. Hashitsume, *Statistical physics II. Nonequilibrium statistical mechanics.* Springer Series in Solid-State Sciences, 1991.

[118] J. Bieber and W. Matthaeus *Ap. J.* **485** (1997) 655.

[119] J. Candia and E. Roulet *J. Cosm. Astrop. Phys.* **2004** (2004) 007, [astro-ph/0408054v1].

[120] Serguei Vorobiov, Mustafa Hussain and Darko Veberic *P. Auger GAP 2008* **055** (2008).

[121] T. Stanev *Ap. J.* **479** (1997) 290, [astro-ph/9607086].

[122] D. Harari, S. Mollerach, and E. Roulet *J. High En. Phys.* **1999** (1999) 022, [astro-ph/9906309v2].

[123] P. Tinyakov and I. Tkachev *Astrop. Phys.* **18** (2002) 165–172, [astro-ph/0111305v1].

[124] M. Prouza and R. Šmída *Astron. Astroph.* **410** (2003) 1–10, [astro-ph/0307165v1].

[125] M. Kachelrieß, P. Serpico, and M. Teshima *Astrop. Phys.* **26** (2007) 378–386, [astro-ph/0510444v2].

[126] L. Page *et al Ap. J. Suppl. S.* **170** (2007) 335, [astro-ph/0603450].

[127] X. Sun, W. Reich, A. Waelkens, and T. Enblin *Astron. Astroph.* **477** (2008) 573–592, [arXiv:0711.1572].

[128] R. Jansson, G. Farrar, A. Waelkens, and T. Enßlin *J. Cosm. Astrop. Phys.* **2009** (2009) 021, [arXiv:0905.2228].

[129] G. Golup, D. Harari, S. Mollerach, and E. Roulet *Astrop. Phys.* **32** (2009) 269–277, [arXiv:0902.1742].

[130] E. Roulet, *Astroparticle theory: Some new insights into high energy cosmic rays*, in *Lepton and photon interactions at high energies: proceedings of the XXI International Symposium: Fermi National Accelerator Laboratory, USA, 11-16 August 2003*, p. 351, World Scientific Pub Co Inc, 2004. astro-ph/0310367v1.

[131] N. Globus, D. Allard, E. Parizot *Astron. Astroph.* **479** (2008) 97–110, [arXiv:0709.1541].

[132] V. Berezinsky and A. Gazizov *Ap. J.* **643** (2006) 8, [astro-ph/0512090v2].

[133] S.I. Syrovatskii *Sov. Astron. J.* **3** (1959) 22.

[134] V. Berezinsky and A. Gazizov *Ap. J.* **669** (2007) 684, [astro-ph/0702102v1].

[135] D. Hogg *arXiv* (1999) [astro-ph/9905116].

[136] A. Hopkins and J. Beacom *Ap. J.* **651** (2006) 142, [astro-ph/0601463].

[137] H. Yüksel and M. Kistler *Phys. Rev. D* **75** (2007) 083004, [astro-ph//0610481].

[138] T. Stanev, *Ultra-high-energy cosmic rays and neutrinos after auger*, in *Frontier Objects in Astrophysics and Particle Physics, Vulcano Workshop 2008, 26-31 May, Italy. Edited by F. Giovannelli and G. Mannocchi*, vol. 1, p. 449, 2009. arXiv:0808.1045.

[139] G. Hasinger, T. Miyaji, and M. Schmidt *Astron. Astroph.* **441** (2005) 417–434, [astro-ph/0506118].

[140] R. Engel, D. Seckel, and T. Stanev *Phys. Rev. D* **64** (2001) 93010, [astro-ph/0101216].

[141] D. Larson *et al Ap. J. Suppl. S.* **192** (2011) 16, [arXiv:1001.4635].

[142] M. Hauser and E. Dwek *Ann. Rev. Astron. Astrophys.* **39** (2001) 249–307, [astro-ph/0105539].

[143] X. Wu, T. Hamilton, D. Helfand, and Q. Wang *Ap. J.* **379** (1991) 564–575.

[144] A. Fabian and X. Barcons *Ann. Rev. Astron. Astrophys.* **30** (1992) 429–456.

[145] P. Sreekumar *et al Ap. J.* **494** (1998) 523, [astro-ph/9709257].

[146] M.A. Malkan and F.W. Stecker *Ap. J.* **555** (2001) 641, [astro-ph/0009500].

[147] G. Lagache, H. Dole, J.L. Puget *Month. Not. R. Astron. Soc.* **338** (2003) 555–571, [astro-ph/0209115].

[148] F.W. Stecker, M.A. Malkan, S.T. Scully *Ap. J.* **648** (2006) 774, [astro-ph/0510449].

[149] F.W. Stecker, S.T. Scully *Astron. Astroph.* **478** (2008) 1–3, [arXiv:0710.2252].

[150] A. Franceschini, G. Rodighiero, and M. Vaccari *Astron. Astroph.* **487** (2008) 837–852.

[151] J.D. Finke, S. Razzaque, C.D. Dermer *Ap. J.* **712** (2010) 238, [arXiv:0905.1115].

[152] P. Ponente, Y. Ascasibar, and J. Diego *arXiv* (2011) [arXiv:1104.3012].

[153] R. Protheroe and P. Biermann *Astrop. Phys.* **6** (1996) 45–54, [astro-ph/9605119].

[154] E. Dwek and M. Barker *Ap. J.* **575** (2002) 7, [astro-ph/0203472].

[155] J.L. Puget, F.W. Stecker, J.H. Bredekamp *Ap. J.* **205** (1976) 638–654.

[156] D. Fixsen and A. Kashlinsky *Ap. J.* **734** (2011) 61, [arXiv:1104.0901].

[157] L. Epele and E. Roulet *J. High Ener. Phys.* **1998** (1998) 009, [astro-ph/9808104v1].

[158] B. Funk, N. Magnussen, H. Meyer, W. Rhode, S. Westerhoff, and B. Wiebel-Sooth *Astrop. Phys.* **9** (1998) 97–103, [astro-ph/9802308v1].

[159] A. Uryson *Phys. Particles and Nuclei* **37** (2006) 347–367.

[160] M. Ave, A. Olinto, A. Watson, and T. Yamamoto *Astrop. Phys.* **23** (2005) 19–29, [astro-ph/0409316].

[161] T. Stanev, *High energy cosmic rays.* Springer Verlag, 2009.

[162] J. Rachen, *Interaction processes and statistical properties of the propagation of cosmic rays in photon backgrounds.* Universität zu Bonn, PhD thesis, 1996.

[163] T. Stanev *Proc. 30th ICRC, Merida* (2007) [arXiv:0711.2282].

[164] T. Stanev, R. Engel, A. Mücke, R. Protheroe, and J. Rachen *Phys. Rev. D* **62** (2000) 93005, [astro-ph/0003484].

[165] V. Berezinsky, A. Gazizov, and S. Grigorieva *Phys. Rev. D* **74** (2006) 043005, [hep-ph/0204357v3].

[166] G. Blumenthal *Phys. Rev. D* **1** (1970) 1596–1602.

[167] F.W. Stecker and M.H. Salamon *Ap. J.* **512** (1999) 521, [astro-ph/9808110].

[168] A. Cuoco, R. D'Abrusco, G. Longo, G. Miele, and P. Serpico *J. Cosm. Astrop. Phys.* **2006** (2006) 009, [astro-ph/0510765].

[169] M.J. Chodorowski, A.A. Zdziarski and M. Sikora *Ap. J.* **400** (1992) 181–185.

[170] L.A. Anchordoqui, M.T. Dova, L.N. Epele and J.D. Swain *Phys. Rev. D* **55** (1997) 7356–7360, [hep-ph/9704387].

[171] D. Harari, S. Mollerach, and E. Roulet *J. Cosm. Astrop. Phys.* **2006** (2006) 012, [astro-ph/0609294].

[172] T. Stanev *New J. Phys.* **11** (2009) 065013, [arXiv:0810.2501].

[173] A. Mucke, R. Engel, J. Rachen, R. Protheroe, and T. Stanev *Comp. Phys. Comm.* **124** (1999) 290, [astro-ph/9903478].

[174] M. Ahlers and A. Taylor *Phys. Rev. D* **82** (2010) 123005, [arXiv:1010.3019].

[175] D. Allard, E. Parizot, A. Olinto, E. Khan, and S. Goriely *Astron. Astroph.* **443** (2005) L29–L32, [astro-ph/0505566].

[176] E. Khan, S. Goriely, D. Allard, E. Parizot, T. Suomijarvi, A. Koning, S. Hilaire, and M. Duijvestijn *Astrop. Phys.* **23** (2005) 191–201, [astro-ph/0412109].

[177] K. Kampert, J. Kulbartz, N. Nierstenhoefer, M. Risse, and G. Sigl, *Propagation of ultra-high energy nuclei with crpropa*, in *Proc. 31st ICRC, Lodz*, 2009.

[178] M. Ahlers and J. Salvado *arXiv* (2011) [arXiv:1105.5113].

[179] A. Koning, S. Hilaire, and M. Duijvestijn, *Talys: Comprehensive nuclear reaction modeling*, in *Proc. Intern. Conf. Nucl. Data for Sci. Techn.* (R. Haight, M. Chadwick, T. Kawano, , and P. Talou, eds.), vol. 769, pp. 1154–1159, 2005.

[180] "Talys official website." http://www.talys.eu/home/, Aug., 2011.

[181] F. Stecker *Phys. Rev. Lett.* **21** (1968) 1016–1018.

[182] D. Hooper, S. Sarkar, and A. Taylor *Astrop. Phys.* **27** (2007) 199–212, [astro-ph/0608085].

[183] D. Allard, N. Busca, G. Decerprit, A. Olinto, and E. Parizot *J. Cosm. Astrop. Phys.* **2008** (2008) 033, [arXiv:0805.4779].

[184] D. Allard and R. Protheroe *Astron. Astroph.* **502** (2009) 803–815, [arXiv:0902.4538].

[185] D. Allard *arXiv* (2009) [arXiv:0906.3156].

[186] D. Allard, M. Ave, M. Malkan, A. Olinto, E. Parizot, F. Stecker, and T. Yamamoto *J. Cosm. Astrop. Phys.* **2006** (2006) 005, [astro-ph/0605327v2].

[187] R. Aloisio, V. Berezinsky, and S. Grigorieva *arXiv* (2008) [arXiv:0802.4452].

[188] S. Lee *Phys. Rev. D* **58** (1998) 043004, [astro-ph/9604098].

[189] V. Berezinsky *Nucl. Phys. B* **188** (2009) 227–232, [arXiv:0901.0254].

[190] V. Berezinsky, *Astroparticle physics: puzzles and discoveries*, in *J. Phys. Conf Ser.*, vol. 120, p. 012001, 2008. arXiv:0801.3028.

[191] V. Berezinsky *arXiv* (2005) [astro-ph/0509069].

[192] V. Berezinsky, A. Gazizov, and M. Kachelrieß *Phys. Rev. Lett.* **97** (2006) 231101, [astro-ph/0612247].

[193] R. Aloisio, V. Berezinsky, P. Blasi, A. Gazizov, S. Grigorieva, and B. Hnatyk *Astrop. Phys.* **27** (2007) 76–91, [astro-ph/0608219].

[194] G. Decerprit and D. Allard *arXiv* (2011) [arXiv:1107.3722].

[195] J. Cronin *Rev. Mod. Phys.* **71** (1999) 165–172.

[196] M. De Domenico and A. Insolia *arXiv* (2011) [arXiv:1104.5083].

[197] J. Aublin *et al Nucl. Phys. B - Proc. Supp.* **190** (2009) 94–98.

[198] J. Vallée *N. Astron. Rev.* **48** (2004) 763–841.

[199] T. Seshadri and K. Subramanian *Phys. Rev. Lett.* **103** (2009) 81303.

[200] D. Larson *et al Ap. J. Suppl. S.* **192** (2011) 16, [arXiv:1001.4635].

[201] P. Trivedi, K. Subramanian, and T. Seshadri *Phys. Rev. D* **82** (2010) 123006, [arXiv:1009.2724].

[202] S. Das, H. Kang, D. Ryu, and J. Cho *Ap. J.* **682** (2008) 29, [arXiv:0801.0371].

[203] J. Everett, E. Zweibel, R. Benjamin, D. McCammon, L. Rocks, and J. Gallagher III *Ap. J.* **674** (2008) 258, [arXiv:0710.3712].

[204] J. Everett, Q. Schiller, and E. Zweibel *Ap. J.* **711** (2010) 13, [arXiv:0904.1964].

[205] E. Armengaud, G. Sigl, T. Beau, and F. Miniati *Astrop. Phys.* **28** (2007) 463–471, [astro-ph/0603675].

[206] A. Neronov and I. Vovk *Science* **328** (2010) 73, [arXiv:1006.3504].

[207] J.P. Rachen and P.L Biermann *Astron. Astroph.* **272** (1993) 161.

[208] J. Abraham *et al Astrop. Phys.* **29** (2008) 188–204, [arXiv:0712.2843].

[209] R.U. Abbasi *et al Astrop. Phys.* **30** (2008) 175–179, [arXiv:0804.0382].

[210] K. Dolag, M. Erdmann, G. Muller, D. Walz, T. Winchen *Proc. 32nd ICRC, Beijing* (2011).

[211] R. Aloisio, D. Boncioli, A. F. Grillo, S. Petrera, F. Salamida *P. Auger GAP 2011* **117** (2011).

[212] G. Sigl *Proc. 32nd ICRC, Beijing* (2011).

[213] E.W. Weisstein, "Null Hypothesis.,
from MathWorld - A Wolfram Web Resource.
http://mathworld.wolfram.com/NullHypothesis.html."

[214] P.J. Bickel and K.A. Doksum, *Mathematical Statistics.* Prentice Hall, 2nd ed., 2006.

[215] R. A. Fisher, *Statistical Methods for Research Workers.* Oliver and Boyd, 1925.

[216] J. Neyman and E. Pearson *Phil. Trans. R. Soc. Lond. A* **231** (1933) 289.

[217] R. Hubbard and M. Bayarri *Institute of Statistics and Decision Sciences* (2003) 27708.

[218] R. Hubbard and M. Bayarri *Amer. Stat.* **57** (2003) 171–178.

[219] J. Robins, A. van der Vaart, and V. Ventura *J. Americ. Stat. Assoc.* **95** (2000) 1143–1156.

[220] D. Murdoch, Y. Tsai, and J. Adcock *Amer. Stat.* **62** (2008) 242–245.

[221] T. Sellke, M. Bayarri, and J. Berger *Am. Stat.* **55** (2001) 62–71.

[222] P. Sommers *Astrop. Phys.* **14** (2001) 271–286, [astro-ph/0004016].

[223] H. Goldberg and T.J. Weiler *Phys. Rev. D* **64** (2001) 056008,, [astro-ph/0009378].

[224] D. Harari, S. Mollerach, and E. Roulet *J. Cosm. Astrop. Phys.* **2004** (2004) 010, [astro-ph/0404304v1].

[225] E. Waxman, K. Fisher, and T. Piran *Ap. J.* **483** (1997) 1, [astro-ph/9604005].

[226] S. Dubovsky, P. Tinyakov, and I. Tkachev *Phys. Rev. Lett.* **85** (2000) 1154–1157, [astro-ph/0001317].

[227] M. Davis and P.J.E. Peebles *Ap. J.* **267** (1983) 465–482.

[228] S.D. Landy and A.S. Szalay *Ap. J.* **412** (1993) 64–71.

[229] A.J.S. Hamilton *Ap. J.* **417** (1993) 19–35.

[230] P.J.E. Peebles, *The Large Scale Structure of the Universe.* Princeton Univ. Press, 1980.

[231] M. Kachelrieß and D. Semikoz *Astrop. Phys.* **23** (2005) 486–492, [astro-ph/0405258].

[232] M. Kachelrieß and D. Semikoz *Astrop. Phys.* **26** (2006) 10–15, [astro-ph/0512498].

[233] A. Cuoco, G. Miele, and P. Serpico *Phys. Rev. D* **74** (2006) 123008, [astro-ph/0610374].

[234] A. Cuoco, S. Hannestad, T. Haugbølle, M. Kachelrieß, and P. Serpico *Ap. J.* **676** (2008) 807–815, [arXiv:0709.2712].

[235] A. Cuoco, S. Hannestad, T. Haugbølle, M. Kachelrieß, and P. Serpico *Ap. J.* **702** (2009) 825–832, [arXiv:0809.4003].

[236] M. Ave, L. Cazon, J. Cronin, J. de Mello Neto, A. Olinto, V. Pavlidou, P. Privitera, B. Siffert, F. Schmidt, and T. Venters *J. Cosm. Astrop. Phys.* **2009** (2009) 023, [arXiv:0905.2192].

[237] J. Hague, B. Becker, M. Gold, and J. Matthews *J. Phys. G* **36** (2009) 115203, [arXiv:0905.4488].

[238] M.P. Veron-Cetty and P. Veron *Astron. Astroph.* **455** (2006) 773.

[239] J. de Mello Neto *et al Proc. 31st ICRC, Lodz* (2009) [arXiv:0906.2347].

[240] B. Stokes, C. Jui, and J. Matthews *Astrop. Phys.* **21** (2004) 95–109, [astro-ph/0307491].

[241] S. Kullback and R. Leibler *The Annals of Mathematical Statistics* **22** (1951) 79–86.

[242] S. Kullback *Amer. Stat.* **41** (1987) 340–341.

[243] H. Akaike, *Information theory and an extension of the maximum likelihood principle*, in *2nd Int. Symp. on Inform. Th.*, pp. 267–281, 1973.

[244] H. Akaike *IEEE trans. on autom. control* **19** (1974) 716–723.

[245] D. Anderson, K. Burnham, and W. Thompson *J. Wil. manag.* **64** (2000) 912–923.

[246] A. Plastino, A. Plastino, and H. Miller *Phys. Lett. A* **235** (1997) 129–134.

[247] A. Plastino, H. Miller, and A. Plastino *Phys. Rev. E* **56** (1997) 3927–3934.

[248] M. Portesi, F. Pennini, and A. Plastino *Physica A* **373** (2007) 273–282.

[249] C. Fuchs *arxiv:quant-ph/9601020v1* (1995) [quant-ph/9601020v1].

[250] M. Reginatto *Phys. Rev. A* **58** (1998) 1775–1778.

[251] S. Abe and A. Rajagopal *Phys. Rev. A* **60** (1999) 3461–3466.

[252] S. Abe *Phys. Rev. A* **68** (2003) 32302.

[253] W. Gersch, F. Martinelli, J. Yonemoto, M. Low, and J. Mc Ewan *Science* **205** (1979) 193.

[254] K. Burnham and D. Anderson *Wil. Res.* **28** (2001) 111–120.

[255] E. Gumbel, *Statistical theory of extreme values and some practical applications: A series of lectures.* National Bureau of Standards Washington DC, 1954.

[256] E. Gumbel, *Statistics of extremes.* Dover Pub., 2004.

[257] M. Nagano and A. Watson *Rev. Mod. Phys.* **72** (2000) 689–732.

[258] C. Hill, D. Schramm, and T. Walker *Phys. Rev. D* **36** (1987) 1007–1016.

[259] V. Berezinsky and A. Vilenkin *Phys. Rev. Lett.* **79** (1997) 5202–5205, [astro-ph/9704257].

[260] V. Berezinsky, M. Kachelrieß, and A. Vilenkin *Phys. Rev. Lett.* **79** (1997) 4302–4305, [astro-ph/9708217].

[261] A. Venkatesan, M. Miller, and A. Olinto *Ap. J.* **484** (1997) 323, [astro-ph/9612210].

[262] G. Farrar and P. Biermann *Phys. Rev. Lett.* **81** (1998) 3579–3582, [astro-ph/9806242].

[263] D. Fargion, B. Mele, and A. Salis *Ap. J.* **517** (1999) 725, [astro-ph/9710029].

[264] J. Arons *Ap. J.* **589** (2003) 871, [astro-ph/0208444].

[265] G. Sigl, D. Torres, L. Anchordoqui, and G. Romero *Phys. Rev. D* **63** (2001) 81302, [astro-ph/0008363].

[266] Z. Fodor and S. Katz *Phys. Rev. D* **63** (2000) 23002, [hep-ph/0007158].

[267] P. Tinyakov and I. Tkachev *JETP Lett.* **74** (2001) 1–5, [astro-ph/0102101].

[268] G. Cusumano *et al Astron. Astroph.* **510** (2010) 47.

[269] M. Winn, J. Ulrichs, L. Peak, C. McCusker, and L. Horton *J. Phys. G* **12** (1986) 675.

[270] N. Hayashida *et al arxiv:astro-ph/0008102* (2000) [astro-ph/0008102].

[271] M. Ave, J. Hinton, R. Vazquez, A. Watson, and E. Zas *Phys. Rev. Lett.* **85** (2000) 2244–2247, [astro-ph/0007386].

[272] G. Medina-Tanco *Ap. J.* **549** (2001) 711, [astro-ph/0009336].

[273] M.I. Pravdin *et al Proc. 29th ICRC, Pune* **7** (2005) 243–246.

[274] R.U. Abbasi *et al Phys. Rev. Lett.* **100** (2008) 101101.

[275] T. Cover and J. Thomas, *Elements of information theory.* Wiley Series in Telecommunications, 1991.

[276] S. Eguchi and J. Copas *J. Multiv. Anal.* **97** (2006) 2034–2040.

[277] L. De Haan and A. Ferreira, *Extreme value theory: an introduction.* Springer Verlag, 2006.

[278] M. De Domenico, A. Insolia, H. Lyberis, M. Scuderi *J. Cosm. Astrop. Phys.* **2011** (2011) 008, [arXiv:1001.1666].

[279] P. Billoir and O. Deligny *J. Cosm. Astrop. Phys.* **2008** (2008) 009, [arXiv:0710.2290].

[280] P. Auger Collaboration, Anysotropy working group, "https://www.auger.unam.mx/AugerWiki/MockMaps."

[281] P. Auger Collaboration, Anysotropy working group, "https://www.auger.unam.mx/AugerWiki/power_table_fraction."

[282] W. H. Baumgartner, J. Tueller, C. Markwardt, and G. Skinner, *The Swift-BAT 58 Month Survey*, vol. 42 of *Bulletin of the American Astronomical Society*, p. 675, 2010.

[283] M. De Domenico and H. Lyberis *P. Auger GAP 2010* **081** (2010).

[284] M. De Domenico and H. Lyberis *P. Auger GAP 2010* **082** (2010).

[285] M. De Domenico, H. Lyberis, O. Deligny *P. Auger GAP 2011* **061** (2011).

[286] K. Mardia and P. Jupp, *Directional statistics.* John Wiley & Sons Inc., 2000.

[287] R. Fisher *Proceed. of the Royal Society of London. Series A* **217** (1953) 295.

[288] G. Watson and E. Williams *Biometrika* **43** (1956) 344–352.

[289] C. Macolino, O. Deligny *P. Auger GAP 2009* **050** (2009).

[290] R.M. de Almeida and J.R.T. de Mello Neto *P. Auger GAP 2010* **022** (2010).

[291] B.R. Becker, M. Gold and J.A.J. Matthews *P. Auger GAP 2010* **035** (2010).

[292] C. Bonifazi *et al Nucl. Phys. B* **190** (2009) 20–25.

[293] M. Ave, N. Busca, L. Cazon, F. Schmidt, T. Yamamoto *P. Auger GAP 2007* **052** (2007).

[294] C. Finley and S. Westerhoff *Astrop. Phys.* **21** (2004) 359–367, [astro-ph/0309159].

[295] M. Ave, P. Ghia, A. Grillo, S. Mollerach, D. Semikoz *P. Auger GAP 2008* **015** (2008).

[296] Pierre Auger Collaboration *arXiv* (2009) [arXiv:0906.2347].

[297] G. Cusumano, *The palermo swift-bat hard x-ray catalogue: Results after 39 months of sky survey*, in *AIP Conf. Ser.*, vol. 1126, pp. 104–107, 2009. arXiv:0906.4788.

[298] H. Yoshiguchi, S. Nagataki, S. Tsubaki, and K. Sato *Ap. J.* **586** (2003) 1211, [astro-ph/0210132].

[299] P. Blasi and D. De Marco *Astrop. Phys.* **20** (2004) 559–577, [astro-ph/0307067].

[300] M. Ave, A. Grillo, S. Mollerach *P. Auger GAP 2008* **051** (2008).

[301] J.P. Huchra *et al arXiv* (2011) [arXiv:1108.0669].

[302] T. Jarrett, T. Chester, R. Cutri, S. Schneider, M. Skrutskie, and J. Huchra *Astron. J.* **119** (2000) 2498.

[303] M. De Domenico *et al Proc. 32nd ICRC, Beijing* (2011) [arXiv:1107.4805].

[304] L. Caramete and P. Biermann *Astron. Astroph.* **521** (2010) A55+, [arXiv:0908.2764].

[305] L.I. Caramete, P. L. Biermann *arXiv:1107.2244* (2011) [arXiv:1107.2244].

[306] R. Blandford and R. Znajek *Mont. Not. R. Astron. Soc.* **179** (1977) 433–456.

[307] P. Biermann, V. De Souza, and P. Wiita *Ap. J. Lett.* **720** (2010) L155, [arXiv:1006.5022].

[308] M. De Domenico and H. Lyberis *P. Auger GAP 2011* **014** (2011).

[309] B. B. Siffert, L. Cazon, Joao R.T. de Mello Neto, A. Olinto, V. Pavlidou *P. Auger GAP 2007* **114** (2007).

[310] D. Harari, S. Mollerach and E. Roulet *P. Auger GAP 2009* **093** (2009).

[311] D. Boncioli, A. Grillo, S. Petrera, F. Salamida *P. Auger GAP 2011* **064** (2011).

[312] Aloisio, R. and Berezinsky, V. and Grigorieva, S. *arxiv:1006.2484* (2010) [arXiv:1006.2484].

[313] F. Schüssler *et al Proc. 31st ICRC, Lodz* (2009).

[314] O. Deligny, H. Lyberis *P. Auger GAP* **In preparation** (2011).

[315] F. Scheck, *Leptons, hadrons and nuclei.* North-Holland, 1983.

[316] W. Fetscher, H. Gerber, and K. Johnson *Phys. Lett. B* **173** (1986) 102–106.

[317] E.L. Lehmann and J.P. Romano, *Testing statistical hypotheses.* Springer, 3rd ed., 2005.

[318] V. Blobel, """The maximum likelihood method", Slides."

[319] R. Kass and A. Raftery *J. Americ. Stat. Assoc.* **90** (1995) 773–795.

[320] H. Jeffreys, *The Theory of Probability.* Oxford, 3rd ed., 1961.

[321] E.W. Weisstein, "Information Theory.,
from MathWorld - A Wolfram Web Resource.
http://mathworld.wolfram.com/InformationTheory.html."

[322] E. Jaynes and G. Bretthorst, *Probability theory: the logic of science.* Cambridge, 2003.

[323] E. Jaynes *Phys. Rev.* **106** (1957) 620–630.

[324] E. Jaynes *Phys. Rev.* **108** (1957) 171–190.

[325] E. Jaynes, "Information theory and statistical mechanics, in 1962 Brandeis Lectures."

[326] C. Shannon and W. Weaver *Urbana: University of Illinois Press* **97** (1949).

[327] A. Rényi, *On measures of information and entropy*, in *Proceedings of the 4th Berkeley Symposium on Mathematics, Statistics and Probability*, pp. 547–561, 1961.

[328] J. Lin *IEEE Trans. on Information theory* **37** (1991) 145–151.

[329] I. Csiszar *IEEE Trans. on Information Theory* **41** (1995) 26–34.