Search for cosmogenics photons in the ultra high energy cosmic ray flux

Julien Souchard

THÈSE

Pour obtenir le grade de

DOCTEUR DE L'UNIVERSITE GRENOBLE ALPES

Spécialité : **Physique Subatomique et Astroparticules**

Arrêté ministériel : 25 mai 2016

Présentée par

Julien SOUCHARD

Thèse dirigée par **Corinne BERAT**, Directrice de Recherche, Université Grenoble Alpes

préparée au sein du **Laboratoire de Physique Subatomique et Cosmologie**
dans **l'École Doctorale de Physique**

Recherche de photons cosmogéniques dans le flux des rayons cosmiques d'ultra-haute énergie

Search for cosmogenics photons in the ultra high energy cosmic ray flux

Thèse soutenue publiquement le **28 octobre 2020**,
devant le jury composé de :

Monsieur Laurent Derome
Professeur des universités, UGA, Président du jury
Madame Isabelle Lhenry-Yvon
Directrice de Recherche, IJCLAB, Rapporteur
Madame Cristina Carloganu
Directrice de Recherche, LPC, Rapporteur
Monsieur Frédéric Piron
Directeur de Recherche, LUPM, Examinateur
Madame Mariangela Settimo
Chargée de Recherche, SUBATECH, Examinateur

Contents

Chapter 1

Introduction

The Earth is constantly bombarded by ionized nuclei called cosmic rays (CR). Above 10^{18} electron-volt (eV) (or 0.16 Joules), the ultra-high energy cosmic rays (UHECR) are a probe of the extreme astrophysical events. Indeed in the universe, the conditions needed to accelerate UHECRs are hard to meet, and finding a sizeable flux of UHECRs being produced in an astrophysical object would constrain its properties. On the other hand, if such energetic particles are produced by the decay of unknown super heavy ones, it could shed light on these unknown species, and their existence in the universe. In recent years, the potential of UHECR astronomy has been highlighted by the multi-messenger campaigns conducted following the successful detections of gravitational waves emitted in merger events. In such events, it is crucial to observe several emission types, not only the usual wavelengths studied in astronomy but also charged cosmic rays, neutrinos, gravitational waves and ultra-high energy (UHE) photons. Multi-messenger astronomy has excellent prospects in the future, through the combined observations of multiple experiments.

The observation of cosmic neutral particles (neutrinos and photons) at ultra high energy would open a new window for the astronomy at the highest energies. Indeed, these neutral messengers are particularly interesting since they directly point to their emission sites as they are not deviated by galactic and extragalactic magnetic fields. UHE neutral particles may be produced by the hadronic interactions of UHECRs in the vicinity or inside the sources of the latter. As charged UHECR, in the so-called top-down models, they are expected to be produced in the decay of super heavy and still unknown particles. They may also come from the interactions of UHECRs with interstellar backgrounds, such as the cosmic microwave background (CMB) or other photon fields.

The determination of ultra high energy photon flux can provide powerful constraints for astrophysical models. In case of potential sources of high energy photons, it may help to set limits on the maximal emission energy achievable by them or in constraining their emission spectrum shape; it is also a crucial observable to determine the origin of the flux suppression in the CR spectrum observed at the highest energies above 5×10^{19} eV. Indeed, the two predominant scenarios to explain the suppression are an exhaustion of the sources, meaning that the sources of UHECRs are simply not capable of accelerating CRs further in energy, and the GZK cut-off scenario. In this scenario the UHECRs flux suppression is explained as a consequence of the resonant pion photoproduction of cosmic primaries with CMB photons. Indeed, above $\sim 6.9 \times 10^{19}$ eV, UHE protons can interact with the CMB photons resonantly, increasing the probability of the pion photoproduction considerably. Interestingly, this process should generate neutral pions which decay into photons. Thus, an increase in the photon flux below the threshold energy for the resonant process would be an indicator of the GZK process.

So far, no photon has been detected above a few TeV in energies. However, upper-limits on UHE photon flux have been set by cosmic rays observatories. These limits have disfavored some of the top-down models designed to explain the origin of UHECRs. At the current time, the most-stringent upper-limits on the UHE photon flux have been set by the Pierre Auger Observatory (PAO).

The PAO is the current flagship observatory studying UHECRs. It is a ground-based experiment, that measures the extensive air showers (EAS) induced by the CRs as they interact with the atmosphere, to reconstruct the characteristics of the primary cosmic particles. The PAO has collected the largest data set of EAS and has produced a wealth of results on UHECRs such as their energy spectrum, their mass composition and their arrival directions. Moreover, analyses are designed specifically to identify UHE photon primaries from the flux of UHECRs in the PAO data. Features of the EAS are used to differentiate between showers induced by photons, mostly electromagnetic, and the ones induced by nuclei primaries, with a larger hadronic component. Nevertheless, there is room to improve the UHE photon searches by using new observables in dedicated multivariate analyses. This was the main objective of my thesis work. I designed a new multi-variate analysis (MVA) for photon identification in the flux of UHECRs detected by the PAO. The aim of this analysis is to build discriminant observables from the water cherenkov detectors (WCD) signal to estimate the EAS muonic component. The observables are then combined into a multi-variate analysis to separate photons from hadrons in the UHECRs data.

Despite the progress made in UHECR physics thanks to the work performed by the Pierre Auger Collaboration, some questions have yet to be answered. An upgrade of the PAO called AugerPrime is thus underway, to improve the capabilities of the observatory. The most prominent feature of the upgrade is the addition of scintillation detectors (SSD) on each of the existing water cherenkov detectors making up the surface detector (SD). During my thesis, I contributed in the integration of part of the SSD. The assembly of the plastic scintillators was shared between 6 member institutes of the Pierre Auger Collaboration. The LPSC had the responsibility of building 90 SSDs. An assembly procedure was designed to guarantee an efficient construction process throughout the production. Tests were performed to guarantee that each detector functions correctly and to characterize each SSD response to atmospheric muons.

The present manuscript is divided in 6 chapters. The first one is dedicated to cosmic rays, with an emphasis on ultra high energy ones and among them the photons. Then in the second chapter, I describe the main features of extensive air showers, and I underline the differences between hadronic showers and the ones generated by photons, since these differences are used to discriminate between the two types of primaries in photon searches. In the third chapter, I present the Pierre Auger Observatory, with a focus on the surface detector, since the analysis I designed uses information delivered by it. After giving an overview of the current results from the Pierre Auger Observatory, the AugerPrime upgrade is introduced in the fourth chapter. The focus is put on the Scintillator Surface Detectors, since the LPSC constructed some of these detectors before shipping them to the Observatory. I also describe in detail the test and characterization procedures that were used at the LPSC. In the fifth chapter, the searches for UHE photons conducted by cosmic ray observatories are presented, along with the main observables used to discriminate between EAS induced by photons and those induced by hadronic primaries. This chapter is concluded by a description of the discriminant observables selected for my analysis and a presentation of the tools and method used to combine these observables into a multi-variate analysis. Finally, the results of my MVA for photon identification are presented in the last chapter.

Chapter 2

Ultra high energy cosmic rays

Contents

Compared to the time when Hess first noticed their existence, we now know much more about the cosmic rays, their natures, their energies and flux, thanks to more than a century of research since the discovery of radioactivity. But some questions are still left unanswered, in particular concerning the most energetic cosmic rays. In this chapter, after a short history on the discovery and the first studies on the cosmic rays, their spectrum is presented. Then the main acceleration mechanisms, some of the potential source classes and the propagation of cosmic rays are described. The third section is dedicated to the ultra high energy photons.

2.1 Cosmic rays, up to the highest energies

2.1.1 Discovery of the cosmic rays

The first hint towards the discovery of the later called cosmic rays came in 1785, when Coulomb noticed that his electrical balance spontaneously discharged over time [1]. Following that, many explanations were proposed to solve the mystery, the most credible in the early 1900s, being that it was due to charged particles produced by radioactive material. During ten years of research, experimentalists brought their instruments around the globe, down in salt mines, underwater and up in the air, to find out the origin of the charged particles.

In 1912, Hess brought electroscopes with him onto a hot air balloon and measured that the ionization decreased at a slower rate than expected at high altitudes. He concluded that it was proof of the extra-terrestrial origin of these radiations. At around the same time, Pacini also performed measurements of the ionization but instead of going higher in altitude, he brought his electroscopes to sea. He concluded, that the radiations did not come from the Earth's crust [2].

After a stop in the scientific investigations during WWI, in 1925-1926, Millikan and Cameron conducted new measurement campaigns and Millikan coined the term "*cosmic rays*" (CR). His publications however were widely criticized for not citing the previous works done by Pacini, Gockel and Hess, and for his "*marketing*" of the discovery in major U.S. scientific papers.

In 1927, Skobelzyn photographed the first CR, as a track in a cloud chamber. The same year, Clay discovered (by serendipity) that there was a latitude-dependance in the CR flux. At that time, the rays were thought to be gamma-rays, due to their high penetrating powers. In the 1930s, independent experiments by Rossi, Alvarez & Compton and Johnson found that CRs are mainly positively charged particles, using the latitude dependence of the CR flux.

In 1932, Anderson discovered antimatter, in the form of positron, using a cloud chamber placed in a magnetic field and in 1937, with Neddermeyer, Anderson discovered the muon in cosmic rays. In 1938, Pierre Auger measured cosmic rays in coincidence between independent detectors [3], confirming the results obtained by Rossi [4], Bothe [5] and Kolhörster [6] with improved accuracy. He concluded from his observations that these cosmic rays detected on the ground resulted from a single primary particle, at the top of the atmosphere and estimated the energy of the primary particle based on the number of particles detected at ground-level. The research on CRs was forced into a pause by WWII.

In 1949, Fermi proposed a mechanism able to accelerate CRs in shocks against magnetic clouds [7] (see more in section 2.2.1). In the 1960s, Greisen, Kuzmin and Zatsepin showed that above a given energy, CRs would interact resonantly with the photons of the cosmic microwave background (CMB), effectively reducing their energy and limiting their travel distance [8][9]. The expected cutoff in the UHECRs energy spectrum resulting from this effect is called the GZK cutoff (see more in section 2.3.1).

In 1963, Linsley reported the detection of an event with energy 1×10^{20} eV in the data from the Volcano ranch experiment [10]. In the 1990s, other cosmic ray observatories detected CRs with energies above 10^{20} eV [11][12]. As the observations pushed the boundaries to higher and higher energies, the lack of statistics due to the drop in the CR flux lead a group of physicists to start working on a giant array project, later named the Pierre Auger Observatory (see section 4.1).

Compared to the time when Hess first noticed their existence, we now know much more about the cosmic rays, their natures, their energies and flux, thanks to more than a century of research since the discovery of radioactivity. But some questions are still left unanswered.

2.1.2 Cosmic ray up to the highest flux

Cosmic rays are ionized atomic nuclei that travel through the Universe and can be detected on Earth. They span a range of energy from a few 10^8 eV to at least a few 10^{20} eV. Around 1000 CRs.m^{-2}.s^{-1} hit the top of atmosphere, of which around $\sim 90\%$ are protons. The CR flux intensity decreases with increasing energies, following an approximate power law on 12 orders of magnitude in energies and 30 in flux. At low energies (compared to UHECRs), CRs can be observed directly and are relatively well understood thanks to experiments launched on air balloons like BESS [13] and CREAM [14], in satellites like PAMELA [15] and on the International Space Station like AMS [16] (see [17] and the references therein for more information). At high energies however, CRs measurements become trickier due to the very low flux, and have to be detected indirectly, through the observation of the cascades that are produced as they go through the atmosphere (see chapter 3). The CRs with energies above 10^{18} eV are UHECRs.

The differential energy spectrum of CRs falls with increasing energies:

$$\frac{dN}{dE} \propto E^{-\gamma} \tag{2.1}$$

with $\gamma \sim 3$, meaning that for an increase of energy of one order of magnitude, the flux decreases by approximately 3 orders of magnitude. The value of γ indicates the steepness of the decrease in the spectrum. Features in the spectrum (see Figure 2.1), such as inflection points or sharp decrease, are usually interpreted in terms of astrophysical phenomena [18], as described below.
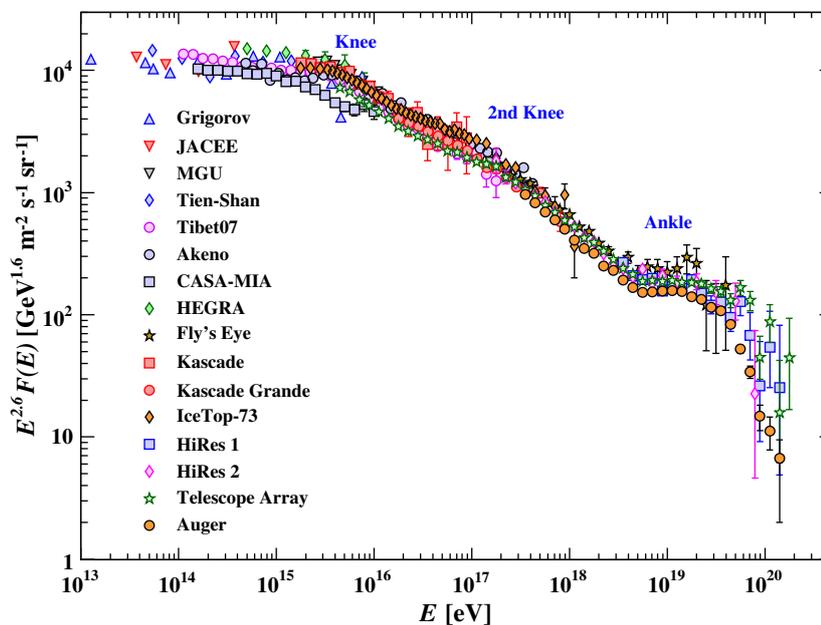
Figure 2.1: Charged cosmic ray spectrum measured from 10^{13} eV to few 10^{20} eV by cosmic ray observatories. The differential energy spectrum has been multiplied by $E^{2.6}$ in order to display the features of the steep spectrum that are otherwise difficult to discern. The spectrum features corresponding to inflection points around 4×10^{15} eV, 10^{17} eV and 4×10^{18} eV are highlighted with the "Knee", "2nd Knee" and the "Ankle". Taken from [17], see also [19] and references therein.

Knee - around 4×10^{15} eV

Up to $\sim 4 \times 10^{15}$ eV, the spectrum follows a power-law with $\gamma = -2.7$. Afterwards, there is a steepening at $\gamma = -3.1$. A popular explanation of this phenomenon is that this energy corresponds to the limit of magnetic confinement (hence a limit on the acceleration power) of proton by galactic supernovae [20]. A second knee is seen around 4×10^{17} eV, where a further steepening occurs and composition studies have hinted toward a transition from lighter to heavier nucleus [21]. The logical extension of the scenario presented for the first knee is that the acceleration limit of galactic supernovae is rigidity-dependent and thus the second knee would correspond to the acceleration limit for the heaviest CRs.

Ankle - around 4×10^{18} eV

Around 4×10^{18} eV, a flattening of the spectrum is observed: it is called the ankle. The ankle can be interpreted as an extragalactic component starting to take over the galactic one at sufficiently high energies [22]. This transition from galactic to extragalactic is located in the range 10^{16}-10^{18} eV. Galactic sources, would reach their maximum accelerating capabilities, leaving extra-Galactic sources to contribute more and more to the total flux. In the so-called "dip" model, assuming a proton extragalactic component, the ankle results from pair-production interaction of the UHE protons with CMB photons [23]. Another explanation is that the local environment of sources can act as a "high-pass" filter on the escaping UHECRs [24]. In this model, only the most energetic CRs can escape the surrounding medium without interacting. The discrimination between the different models relies mainly on mass composition studies.

Suppression - above 6×10^{19} eV

At the high end of the spectrum, above 6×10^{19} eV a suppression is observed in the CR flux. This suppression has two possible explanations: the theoretically predicted GZK cut-off (see subsection 2.3.1) and the sources reaching a sharp-limit to their acceleration capabilities (see section 2.2). So far there has been no conclusive evidence favoring one scenario over the other.

2.2 Acceleration and propagation processes

To understand and explain the spectrum of the UHECRs observed on Earth, it is necessary to also know the processes that are capable of accelerating particles up to these energies. Also, once these accelerated UHECRs are produced in the sources, they have to travel through the universe before reaching our planet. During this travel, the CRs will interact with the surrounding environment, especially photon fields. These interactions will modify the shape and composition of the spectrum seen on Earth from the one at the sources.

2.2.1 Main acceleration processes

In the bottom-up paradigm, to reach ultra high energies, the CRs have to be accelerated through magneto-hydro-dynamic (MHD) processes. The acceleration processes can mainly be divided in two types:

- the first type in which the CRs are accelerated through multiple stochastic encounters with moving magnetized plasma. The acceleration is slow and the UHECRs are hard to keep confined into the Fermi-acceleration zone.

- the second type ("one-shot") in which the CRs are directly accelerated to very high energies, relying on EM fields, jets or relativistic flows. These have the advantage of being fast acceleration processes, but the environment near the sources are usually high in density and thus prone to interactions for the CRs, which degrade their energy.

In the following paragraph, the Fermi acceleration mechanisms are described in more detail as they are usually considered as the main mechanism involved in the acceleration of UHECRs.

Fermi acceleration

In Fermi acceleration processes, the CRs collide with moving magnetized scattering centers. In his original paper [25], Fermi described the now called second-order Fermi acceleration, where charged particles are accelerated through elastic scattering with moving magnetized clouds.
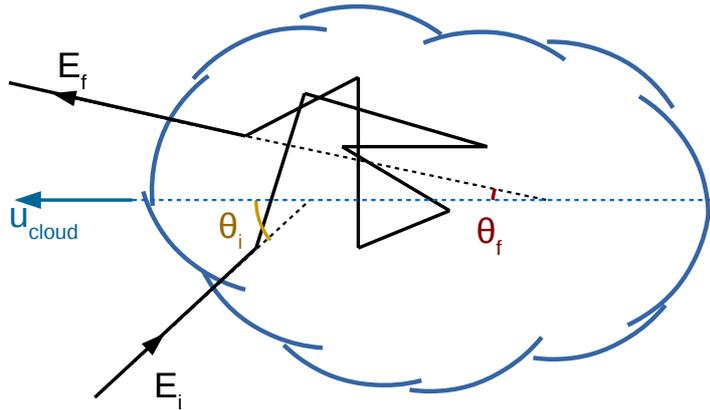
Figure 2.2: Illustration of the second-order Fermi acceleration process.

For a charged particle with initial energy E_i entering with an incident angle θ_i a magnetized cloud of Lorentz factor Γ_{cloud} moving with velocity $u_{cloud} = v_{cloud}/c$ its energy in the cloud rest-frame becomes E'_i:

$$E'_i = \gamma_{cloud}E_i(1 - u_{cloud}\cos(\theta_i)) \tag{2.2}$$

Inside the clouds, the energy of the CR is not expected to change on average, so $E'_i = E'_f$. And the final energy of the CR as it leaves the cloud with an angle θ_f, in the laboratory frame E_f is:

$$E_f = \gamma_{cloud}E'_f(1 - u_{cloud}\cos(\theta_f)) \tag{2.3}$$

On average and assuming that the particles are isotropized in the clouds ($\langle\cos(\theta_f)\rangle = 0$), the gain in energy is proportional to the squared velocity of the cloud:

$$\frac{\langle E\rangle}{E} \propto \frac{4}{3}u_{cloud}^2 \tag{2.4}$$

In the first-order Fermi acceleration, the particles are accelerated through consecutive scattering across a shock-front, between a shocked downstream region and an unshocked

upstream region. Let us consider a shock wave propagating at a velocity $u_{sh} = v_{sh}/c$ through a magnetized medium (like the interstellar medium). The medium can be divided into an upstream region, which has not been shocked yet and a downstream region, which has been shocked. Now, the situation can be seen from the three point-of-views (see Figure 2.3).
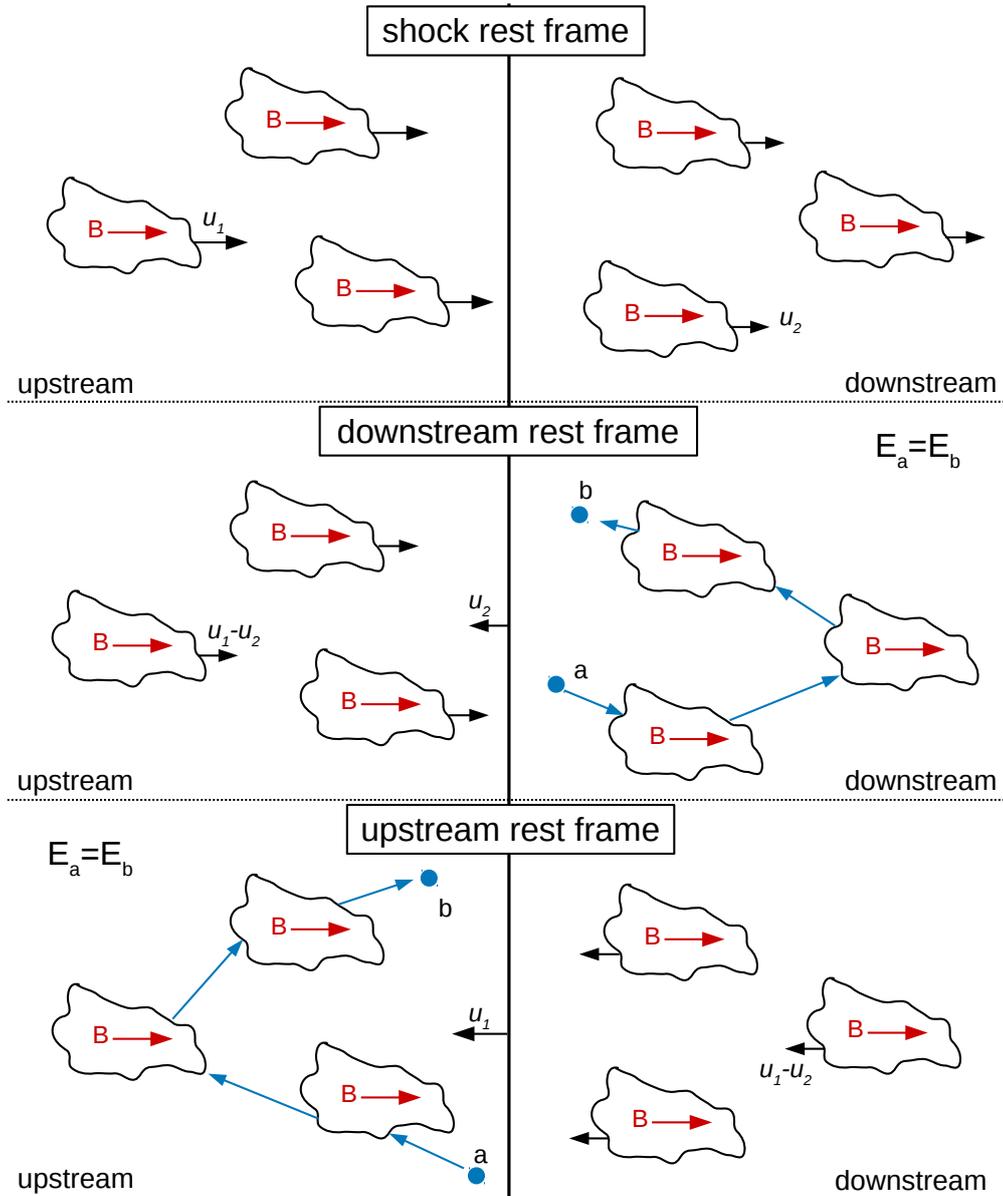
Figure 2.3: Illustration of the first-order Fermi acceleration mechanism.

In the shock rest frame, the upstream region is moving toward the shock wave with a velocity $u_1 = u_{sh}$ and the downstream region is moving away from the shock with a velocity $u_2 = u_{sh}/r$ On both sides of the shock, magnetic inhomogeneities isotropize the CRs, making them recross the shock. The CRs gain energy through these upstream→downstream→upstream cycles. On average, the gain in energy is proportional to the velocity of the shock:

$$\frac{\langle E \rangle}{E} \propto \frac{4}{3}(u_1 - u_2) \tag{2.5}$$

Since $u_{cloud}, u_{sh} \ll 1$, the first-order Fermi mechanism is more efficient than the second-order one to accelerate CRs.

Another advantage of the first-order mechanism is that it naturally produces a power-law spectrum [26] whereas the second-order mechanism produces a thermal spectrum [27] [28].

The first-order Fermi mechanism is usually postulated as the main acceleration mechanism for UHECRs. Other mechanisms have been proposed, like magnetic reconnection [29][30] and unipolar induction [31][32][33].

2.2.2 Non-acceleration processes

Other non-acceleration processes (top-down models) have been proposed to explain the energies of UHECRs, but most of them have been disfavored by experimental results. These models were mainly introduced to explain the spectrum published by the AGASA experiment in 1998, which did not observe the expected GZK cut-off (see subsection 2.3.1) at UHE [34]. Here the main types of top-down models are introduced.

Z-burst

In Z-burst models [35], a UHE-ν interact with a background relic neutrino into a Z-resonance which decays into a fermion/anti-fermion pair:

$$\nu_{UHE} + \nu_r \to Z \to f\bar{f} \tag{2.6}$$

To generate a Z-resonance with a sizeable cross-section, the energy available in the interaction must be:

$$E_\nu^{res} = \frac{M_Z^2}{2m_\nu} = 4.2 \times 10^{21} \, \text{eV} \, \frac{1 \, \text{eV}}{m_\nu} \tag{2.7}$$

where M_Z is the mass of the boson Z and m_ν is the mass of the neutrino. The fermion/anti-fermion pair originating from this interaction could thus have a large amount of energy available, in the order of the UHECRs.

Decay of heavy particle

Some models generate UHECRs from the decay of metastable heavy (mass$\gtrsim 10^{21} \, \text{eV}$) particles X. The super heavy dark matter (SHDM) models [36][37][38] assume X to be metastable heavy dark matter particles, formed during the inflation stage of the Universe, that are part of the cold dark matter component of DM. As DM density is expected to be higher in the halo of our galaxy compared to the extragalactic medium, the produced flux of UHECRs would originate from close enough to avoid the GZK cut-off on its way to Earth. Topological defects have also been proposed to generate these heavy particles X [39][40].

Most top-down models have been disfavored by UHE-photons and UHE-neutrinos searches performed on observational data (after the AGASA result) [41][42][43], usually because the observed flux are too low compared to the ones expected by these models.

2.2.3 Potential sources

In the Fermi acceleration paradigm, the minimal requirement for a potential source is that it must satisfy the Hillas criterion [44]. It states that to be accelerated up to UHE, a particle must be confined within the magnetic field of the source as it gains energy. The propagation of an UHECR of energy E and charge $Z.e$ in a magnetic field B is characterized by the length scale r_L, the Larmor radius:

$$r_L = 1.1 \frac{1}{Z} \left(\frac{E}{10^{18} \, \text{eV}} \right) \left(\frac{B}{\mu \, \text{G}} \right)^{-1} \text{kpc} \tag{2.8}$$

where α is the fine-structure constant. If the Larmor radius of a particle in the source is smaller than its size, the particle will be able to remain within the acceleration region.

The maximal acceleration energy accessible for a given source candidate and for a given particle of charge Ze can then be qualitatively estimated as:

$$E^{source} \lesssim Z \left(\frac{R}{\text{kpc}} \right) \left(\frac{B}{\mu \text{ G}} \right) \times 10^{18} \, \text{eV} \tag{2.9}$$

E_{max}^{source} is thus proportional to $B \times R$ and can be visualized with a "Hillas plot" as shown in Figure 2.4. An important caveat of relation 2.9 is that it does not take the energy loss happening in the source environment into account. To be a viable candidate, the source must be able to accelerate the CRs up to UHE and within a time and local environment that do not strip the CRs of their gained energy through interactions with the local radiation fields. The energy losses must not compete with the acceleration process. The density of the sources and the produced flux of accelerated CRs must also match the observed spectrum and composition on Earth. This also implies that the acceleration sites must be located within a distance that allows the propagation of the UHECRs to the Earth (see subsection 2.2.4).
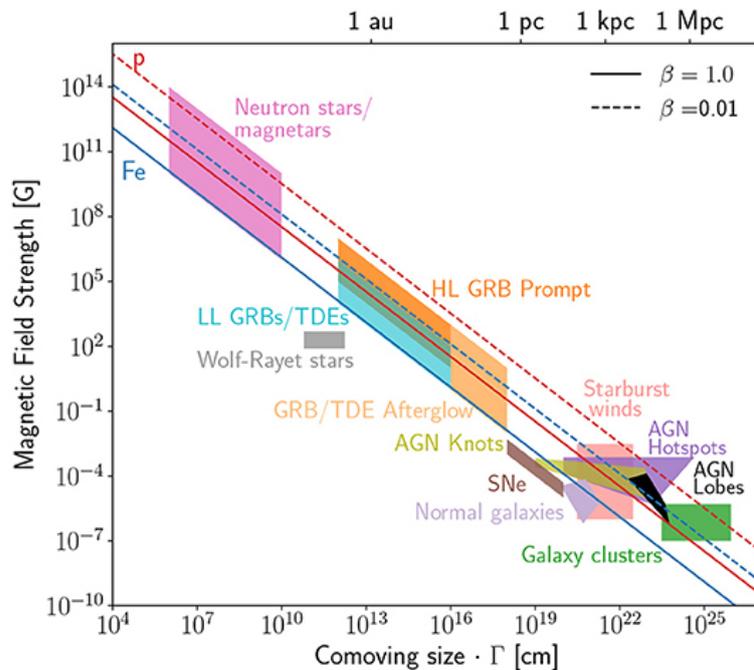
Figure 2.4: Hillas diagram, showing the source classes as a function of their characteristic size R, the product of the comoving size and the Lorentz factor Γ of the flow, and their magnetic field strength B. Following relation 2.9, the lines indicates the BR product beyond which the confinement of CRs with energy 10^{20} eV is possible. The continuous (dashed) lines indicate the values for protons, in red and for iron, in blue, for outflows with velocity $\beta_{sh} = 1$ ($\beta_{sh} = 0.01$). To be viable, a source must be in the upper-right region of the plot. The figure is taken from [45], more details on the sources can be found in it and in the references therein.

Although the mechanisms probably can be generalized (they should all stem from magneto hydro-dynamics phenomena), each type of source candidate has to be studied individually, because they do not have the same injection mechanisms and/or outflow flux and mostly because we lack observational data on the inner mechanisms of the sources. The agreement with the spectrum flux measured on Earth is an important criteria to constrain the source models. In the following, few potential sources types are briefly described.

StarBurst Galaxies

StarBurst Galaxies (SBG) are galaxies that are undergoing an intense phase of star formation. This activity generates magnetized wind which, coupled with the high matter density inside the SBG environment, is suitable for cosmic ray acceleration [46]. Some limitations have been shown recently however, on the maximum energy reachable through shock acceleration in the SBG winds [47]. However, since SBGs are regions of very high astrophysical activity, the acceleration of CRs up to the highest energies could take place in multiple steps [48] or in extreme explosions happening in the SBG (gamma-ray bursts [49][50], supernovae [51][52]).

Active Galactic Nuclei

Active Galactic Nuclei (AGN) are composed of an accretion disk, around a super-massive black hole. Some AGNs, classified as *radio-loud* produce relativistic collimated jets that end in radio-lobes. Gamma-ray emitting AGNs (γ-AGNs) are an interesting potential source class of UHECRs because they have been extensively studied, thus part of their population is known and because the jets+radio-lobe structure is prone to shock accelerations [53][54][55].

In AGNs, tidal disruptions events can occur when a star approaches a black hole and is disrupted by it. This generates flares that could accelerate particles up to UHECRs energies [56].

Gamma-Ray Bursts

Gamma-Ray Bursts (GRBs) have long been considered as potential sources of UHECRs [57][49]. They are brief bursts of high-energy radiation, often explained as extreme events happening in stellar-mass objects. Long-GRBs are often associated with the collapse of a massive star and short-GRBs with compact binary mergers. The main model used to describe the radiation of the GRBs is the fireball model [58][59]. According to this model, GRBs should be able to accelerate proton up to the highest energies [60].

"One-shot" sources

Other objects, relying on unipolar induction or magnetic reconnection have been proposed as potential UHECRs acceleration sites: white dwarfs [61][62][63], magnetars [64], pulsars [65][66] and black holes [67][68].

2.2.4 Propagation of the UHECRs

As they propagate from the acceleration sites to Earth, the UHECRs interact with radiation and magnetic fields. The main sources of energy losses for UHECRs are:

- pair-production: the CR interacts with a photon to create an electron/positron pair: $N\gamma \rightarrow Ne^+e^-$.

- photomeson production: the CR interacts with a photon to create a meson: $N\gamma \rightarrow N\pi$ [69].

- photodisintegration: a CR nucleus (not for protons) is excited by photons, triggering the emission of one or several nucleons.

- adiabatic expansion of the Universe: the adiabatic energy loss is given by $E = \frac{E_0}{1+z}$ [70].

Here, only an overview of the processes happening during UHECRs propagation is presented. A detailed description of the interaction processes encountered by the CRs as they propagate through the Universe can be found in [71] and the GZK process, source of cosmogenic photons is discussed in the next section. It is important to underline here that pair-production, pion photoproduction and photodisintegration processes all produce high energy photons. Taking the example of iron nuclei, with energies between 10^{18} eV and 10^{21} eV, the fraction of the energy transferred to the photon production channels range between a few $10^{-6}\%$ and a few $10^{-1}\%$ [72]. The fraction of the energy transferred to the EM channels in general increases with the nuclei energy.

The propagation processes are not the same for protons compared to heavier nuclei. For protons first (see Figure 2.5): above its energy threshold $E_{thr}^{pp} \sim 10^{18}$ eV pair-production starts taking over the energy loss. Around $E_{thr}^{\pi} \sim 7 \times 10^{19}$ eV, pion photoproduction becomes the most dominant energy loss process. Above 10^{20} eV the energy loss length falls to less than $\lesssim 100$ Mpc, the propagation of UHE protons gets very limited in range.
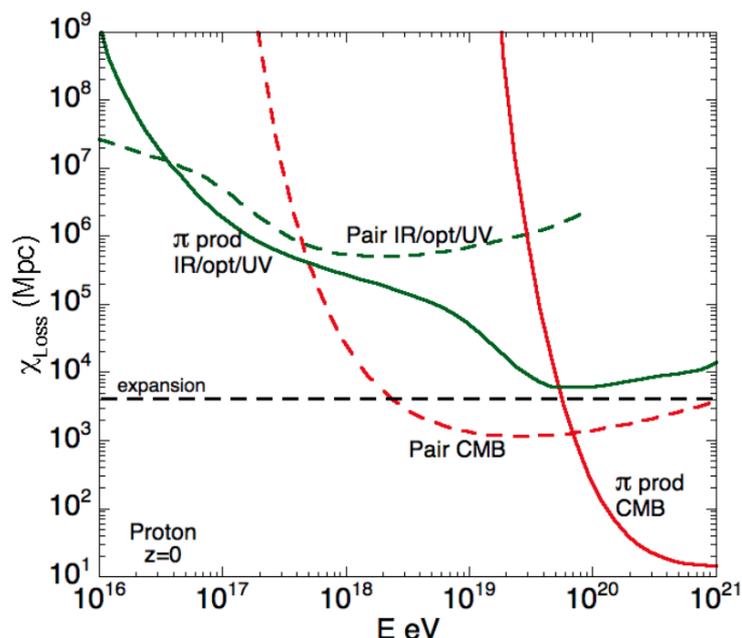
Figure 2.5: Evolution of the energy loss length for an UHE proton at redshift $z = 0$ as a function of its energy. Three processes are represented, the adiabatic expansion of the universe (black dashed horizontal line), the pair production mechanism (dashed curves) and the pion photo-production mechanism (continuous curves). The pair production and pion photo-production mechanism are drawn for interactions of the proton with CMB photons (in red) and for other sources of background photons (IR/opt/UV) (in green). From [73].

Heavier nuclei are affected by the same processes as protons but also by photo-disintegration mechanisms, in which the CR nucleus loses some of its nucleons. These processes include, giant dipole resonance and quasi-deuterium mechanisms. Photodisintegration processes have threshold energies which varies with the mass of the nucleus. The evolution of the contribution to the energy loss of photodisintegration and pair production is shown in Figure 2.6.

Another very important phenomenon that affects the UHECRs in their propagation to Earth is their deflections in magnetic fields. It is widely accepted (and supported by observation [74]) that UHECRs, at least at the highest energies, are of extragalactic origin. Thus, UHECRs will encounter different fields depending on whether they are accelerated in our Galaxy or if they have an extragalactic origin.
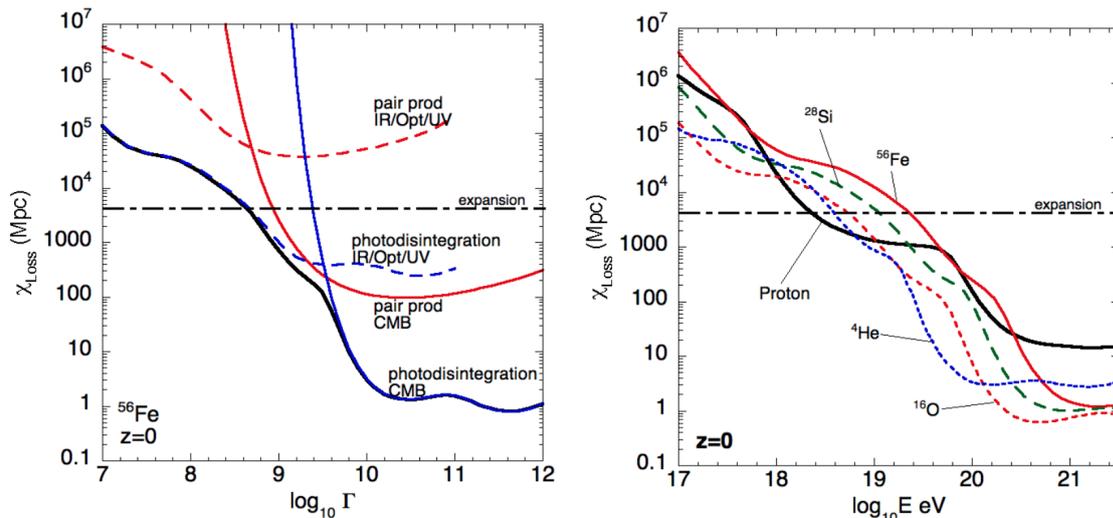
Figure 2.6: Left: Evolution of the energy loss length for an UHE iron at redshift $z = 0$ as a function of its Lorentz factor Γ. The photodisintegration (in blue) and the pair production mechanism (in red) are represented for interactions with the IR/opt/UV background photons (dashed curve) and for CMB photons (continuous curve). Right: Evolution of the energy loss length for UHECRs of various nuclear masses at redshift $z = 0$ as a function of their energy. The curves are drawn for z=1,4,16,28,56. From [73].

Extragalactic magnetic fields are still poorly known. They are expected to be rather low, since no convincing mechanism able to generate strong magnetic fields over a wide scale has been discovered. The processes that could generate extragalactic magnetic fields are still being discussed [75][76]. But the lack of observational data prevents a precise modeling of the EGMF and the uncertainties on the deflections of extragalactic UHECRs are high.

Galactic magnetic fields are comparatively well known and have been modeled against observational data [77][78]. For galactic magnetic fields, the deflections highly depend on the rigidity of the UHECR. In [79], using the model introduced in [77], the authors have found that the average deflection for UHECRs with rigidity[1] below $10\,\mathrm{EV}$ can be beyond $90°$ for almost all sources. However, the deflections quickly become less important with higher rigidity, with the deflections thought to behave as $\sim 3° \times Z(E/10^{20}\,\mathrm{eV})^{-1}$ [80]. If the deflections are weak enough, the arrival directions of UHECRs should mirror the distribution of the sources.

2.3   Ultra high energy photons

Like for CRs, the energy spectrum of photons extends over a very large range of energies. A continuous diffuse photon population is produced in the Galaxy by the interaction of high energy CRs with the interstellar medium. The ones that are particularly interesting for us are the ones at the highest energies, the UHE photons

Several processes can produce UHE photons from UHECRs. Aside from the propagation processes presented in the last section 2.2.4, the synchroton emission process has also been mentioned as a potential production channel of high energy photons [81][82]. However, the pion photoproduction channel is by far the most dominant process in producing photons with energies above $10^{18}\,\mathrm{eV}$ [72].

---

[1]The rigidity R of a CR represent the effect a magnetic field has on it. For ultra-relativistic particle $R \simeq E/Z$.

2.3.1 GZK effect and cosmogenic photons

Most models predict the existence of UHE photons mostly produced in a pion photoproduction process between an UHECR N and a photon γ from the cosmic microwave background (CMB). The cross-section $N\gamma$ is dominated by the resonance $\Delta(1232)$, which decays into a nucleon and a pion π. This resonant process in the case of UHE protons can be expressed as:

$$p + \gamma \rightarrow \Delta^+(1232) \rightarrow p + \pi^0 \tag{2.10}$$
$$\rightarrow n + \pi^+ \tag{2.11}$$

However, the pion photoproduction process has a threshold energy for the photon in the nuclear rest-frame of about $140\,\mathrm{MeV}$ (the mass of the pion) and about $340\,\mathrm{MeV}$ for the resonance $\Delta(1232)$.

For CMB photons this leads to a minimum energy for pion photoproduction with an UHE proton of $E_p^{thr} \sim 3\times10^{19}\,\mathrm{eV}$ while the resonance $\Delta(1232)$ appears around $E_p^{res} \sim 6.9\times10^{19}\,\mathrm{eV}$.

For heavier nuclei, the same processes can be observed, however, a nucleus composed of A nucleons can be approximated to a collection of A nucleons with energies E/A, thus the threshold energies are more important.

Any UHECR with an energy above the pion photoproduction energy threshold with a CMB photon will thus see its energy drained through this process. As a result, the attenuation length of UHECRs is rapidly reduced above the pion photoproduction threshold, and even more above the threshold for the resonant process (see Figures 2.5 and 2.6). For a proton with an energy $E_{pr} = 10^{21}\,\mathrm{eV}$, at redshift $z = 0$, the energy loss length is around $\chi_{loss} \sim 10\,\mathrm{Mpc}$. This value gets even lower for iron nuclei, where the χ_{loss} value is around $\chi_{loss} \sim 1\,\mathrm{Mpc}$ at $E_{Fe} = 10^{21}\,\mathrm{eV}$.

The expected drop in the UHECR spectrum at the highest energies due to the energy loss in pion photoproduction is called the "GZK cutoff". It was first theorized by Greisen and Zatsepin–Kuzmin [8] in 1966, shortly after the discovery of the CMB in 1964 [83][84]. Below the "GZK cutoff", a bump in the energy distribution of UHECRs is expected, corresponding to the primaries that have lost some of their energy through the resonant pion photoproduction process.

The decay of the pions resulting from pion photoproduction is one of the most efficient source of UHE neutral particles, photons and neutrinos, in bottom-up scenarios. Indeed, the dominant decay modes for neutral and positively charged pions are:

$$\pi^0 \rightarrow \gamma + \gamma \tag{2.12}$$
$$\pi^+ \rightarrow \mu^+ + \nu_\mu \tag{2.13}$$

The photons produced through this process have energies of about a tenth of the primary UHE protons energy, less for heavier nuclei (see Figure 2.7).

2.3.2 Propagation of UHE photons

The UHE photons are also subject to interaction processes during their propagation to Earth. However for UHE photons, the most relevant process is pair production with background photons:

$$\gamma + \gamma_{bkg} \rightarrow e^+ + e^- \tag{2.14}$$

The pair production process limits the propagation of UHE photons more strongly than for the UHE nuclei. The UHE photons can interact with photons from the CMB, the infrared background (IRB) and the radio background (URB). The interaction length of UHE photons as a function of their energies is shown in Figure 2.8
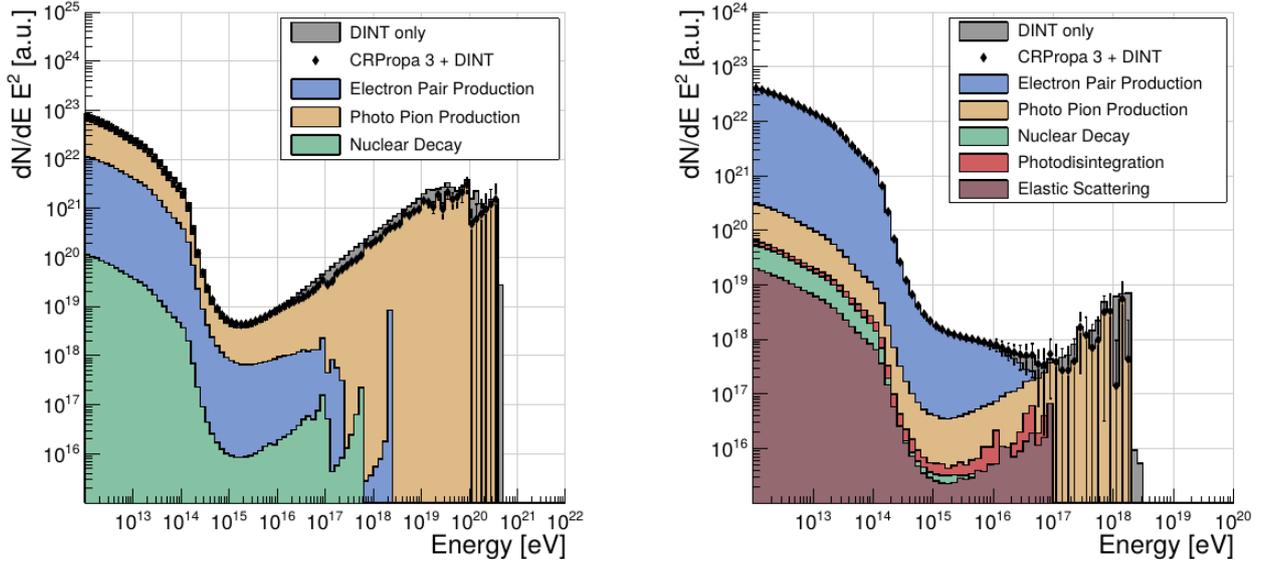
Figure 2.7: Distributions as a function of energy of the secondary photons produced by the interactions of UHECRs during their propagation (from CRPropa simulations). The photon flux is scaled by E^2, and it corresponds to the secondary photons produced by 10^4 primaries (left: proton, right: iron) drawn from a $dN/dE \propto E^{-1}$ distribution, between 10^{18} eV and 10^{21} eV. The primaries are produced from uniformly distributed sources located between 3 and 1000 Mpc from the observer. The black dots show the total photon flux while the individual-process contributions are color-coded. From [72].
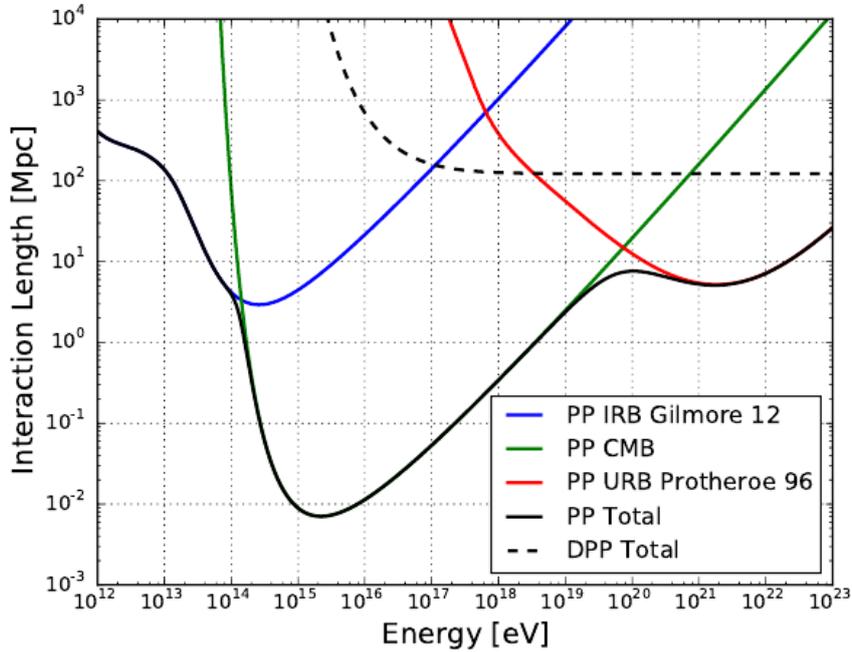
Figure 2.8: Interaction lengths λ for photons interacting with E.M. background as a function of photon energy. The relevant processes are represented individually, pair production (solid lines) and double pair production (dashed line) as a photon energy. The pair production process is further separated into the contributions of each type of photon fields, infrared background (IRB - in blue), cosmic microwave background (CMB - in green) and radio background (URB - in red). From [72].

At UHE, photons travel on average less distance than nuclei before interacting with the cosmic background, around 1 Mpc at $10^{18.5}$ eV. This restricts the observation of UHE photons to the local Universe. If the UHE photons are produced in sources, those must be located

relatively close to Earth, this limits the possibility of having a significant flux of photons directly accelerated to UHE (by contrast to cosmogenic photons). The spectrum of secondary UHE photons on Earth is expected to follow a power law E^{-2} (see Figure 2.7).

2.3.3 Expectations from GZK photons

The expectations on the GZK photon flux depend on the assumed spectrum of UHECRs, the location of the sources and the background environment. In [85], the flux of UHE photons are predicted for two astrophysical cases:

- the low-photon case, in which high infrared (IR) and radio background are considered, along with a strong intergalactic magnetic fields value and a steep source evolution[1] function, following the evolution for FRII type galaxies.

- the high-photon case, in which low IR and radio background are used, with weaker IGMF value and a source evolution following the star formation rate, higher than for FRII type galaxies.

In both cases, the sources are considered up to redshift $z = 1$, since further than that the contributions to the photon flux are negligible. The UHE photon fluxes are predicted for pure-proton and pure-iron composition at the sources, with the pure-proton case leading to higher fluxes of photons (about an order of magnitude more and the pure-iron photon spectrum cutoff at lower energies). The spectral index of the sources and their maximum energy[2] are also used as parameters and the ensuing photon fluxes are drawn.

In [86], GZK-photons flux predictions were also made for a high-photon case using the HiRes spectrum [87]. These predictions can be seen in Figure 6.13.

The detection of UHE photons would be a great help to answer some of the open questions remaining on UHECRs. In particular, the flux of UHE photons around the apparent cut-off at the end of the UHECRs spectrum is an important indicator to constrain the models of the suppression. Also, UHE photons astronomy could help identify and locate UHECRs sources, without inferences from the deflections induced by the GMF and the EGMF and constrain the source models.

---

[1] distribution of sources as a function of the redshift: dN/dz

[2] the energy at which there is an exponential cutoff in the sources energy spectra

Chapter 3

Extensive Air Showers

Contents

We have seen in the previous chapter that at low energy, the rate of cosmic rays is sufficient to have direct detection at the top of the atmosphere, using experiments on board (balloon, satellites, ...). At higher energy, the cosmic ray flux is weaker, and only indirect detection via air showers is possible. In this chapter, the extensive air showers are described in the first section. First, simple models describing the development of EM and hadronic showers are presented, to highlight their main characteristics and their behavior, which are confirmed by more sophisticated simulations. In the second section, the photon-initiated showers are considered, and the specificities of such showers are underlined.

3.1 Atmospheric showers

Along the discovery of the CRs, the air showers initiated by these primaries at the top of Earth's atmosphere were also being studied. In the 1930s, the advent of coincidence detection technique coupled with improved timing resolution in particle counters lead to the observation of secondary particles in coincidence between several independent detectors separated from one another. Around the same time, experiments were launched in the atmosphere to measure the evolution of the cosmic radiations/ionization rate with altitude. In 1935, Pfotzer and Regener sent three Geiger-Müller tubes on a balloon and measured the vertical intensity of the cosmic radiation up to $28\,\mathrm{km}$ in the atmosphere [88]. They found that the intensity increases as the balloon ascends but reach a maximum around $\sim 16\,\mathrm{km}$ before decreasing at higher altitudes.

These measurements were used to build the models explaining the key processes of cascade multiplication in the atmosphere. The fundamental rules based on QED were first laid out by Dirac [89], Fermi [90] and Bethe-Heitler [91][92]. Then Bhabha and Heitler expanded this theoretical framework up to extreme energies to explain the results obtained in cosmic-ray experiments [93]. Finally, Carlson and Oppenheimer completed this picture by using the equations of diffusion to calculate the energy losses by ionization and comparing their predictions to the Pfotzer-Regener measurements [94]. Thanks to those pioneering works, the processes generating extensive air showers are now understood.

When a CR hits the top of Earth's atmosphere, it starts a cascading process. This process generates secondary particles through nuclear and electromagnetic (EM) interactions with the atmospheric molecules. The atmosphere plays the role of calorimeter. In the case of UHECRs, the initiated showers are very large and are called extensive air showers (EAS). The development of an EAS follows the same laws as the cascades studied in particle physics experiments.

The simplified models used to describe EM and hadronic showers are able to derive the main behaviour of the shower development and to underline simple relations between main EAS parameters. Numerical simulations confirm the general behavior of the showers, but allow a more precise description of the showers using current models describing interactions and decays.

3.1.1 Electromagnetic showers

The development of an EM shower in a medium can be approximated and accurately predicted with two fundamental interactions, according to the Heitler model. This model [95] (see Figure 3.1) describes the cascade as a succession of two body splitting involving electrons e^-, positrons e^+ and photons γ through two interactions: bremsstrahlung emission and pair creation. Besides, cross-sections are assumed to be independent of energy and collision energy losses are ignored.

In this simplified model, every particle undergoes a splitting after traveling a fixed distance d related to the interaction length λ_r (assumed to be the same for all particles): $d = \lambda_r ln2$. The interaction length is usually measured in units of g.cm^{-2}, in air: $\lambda_r \sim 37\,\mathrm{g.cm}^{-2}$ [96]. After n splittings, the number of particles in the shower is 2^n. However, once the particles go below a critical energy ξ_c^e, the splitting is stopped. This critical energy corresponds to the energy at which radiative energy loss becomes less than collisional energy loss in air, this value is equal to $\xi_c^e = 85\,\mathrm{MeV}$.

The fact that after a given number of splits the number of particles stops increasing means that there is a *depth* at which the showers reach their maximum development, X_{max}.
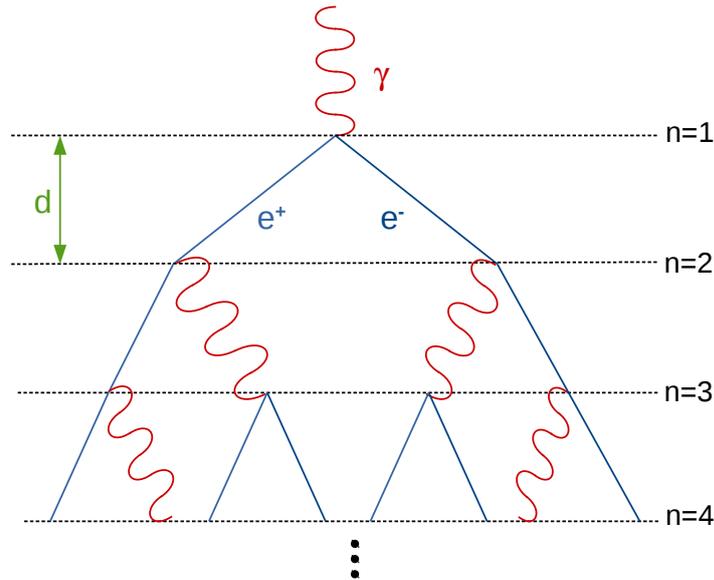
Figure 3.1: Scheme of the first four splits in an EM cascade. Redrawn following [95].

The slant depth X describes the amount of matter over a length dr in an environment of density ρ: $X = \int \rho dr$. By using X, one thus does not have to discuss the fluctuations and inhomogeneities of the atmosphere as they are already accounted for in the slant depth. In this simplified EM model, the maximum number of particles N_{max} is reached when all the particles have energies equal to the critical energy ξ_c^e, after a number of splits n_c ($N_{max} = 2^{n_c}$). X_{max} is then the slant depth required to reach this energy: $X_{max} = n_c \lambda_r ln(2)$.

For a shower initiated by a photon primary with energy E_0:

$$N_{max} = E_0/\xi_c^e \tag{3.1}$$

and

$$X_{max} = \lambda_r ln(E_0/\xi_c^e) \tag{3.2}$$

From this simplified model, one can extract three important aspects of the development of pure EM showers:

- there is a maximum number of particles in the shower N_{max} reached at the depth X_{max}.

- for a given medium, N_{max} is directly proportional to the energy E_0 of the primary.

- for a given medium, X_{max} is logarithmically proportional to the energy E_0 of the primary .

This model is a simplification, the ratio of photons to electron/positrons is underestimated as multiple photons can be emitted during bremsstrahlung. Also, the particles making up the shower do not all reach the critical energy at the same time, thus the population of electrons/positrons and photons do not peak at the same time. Finally, this model does not take into account the particles decoupling from the shower. However, this model is correct and it is a good approximation of the development of an electro-magnetic shower.

3.1.2 Hadronic showers

In [95] (see also [97] for a condensed description), an analogous description of the hadronic showers is presented (see Figure 3.2). The simplifying assumption is to consider that hadronic interactions create 2N charged pions and N neutral pions. The charged multiplicity of each hadronic interaction is N_{ch}. The neutral pions quickly decay into photons, which initiate secondary EM showers. The charged pions will continue interacting until their energy goes below a critical value ξ_c^π at which point they are more likely to decay into muons and neutrinos, and are thus decoupled from the shower.
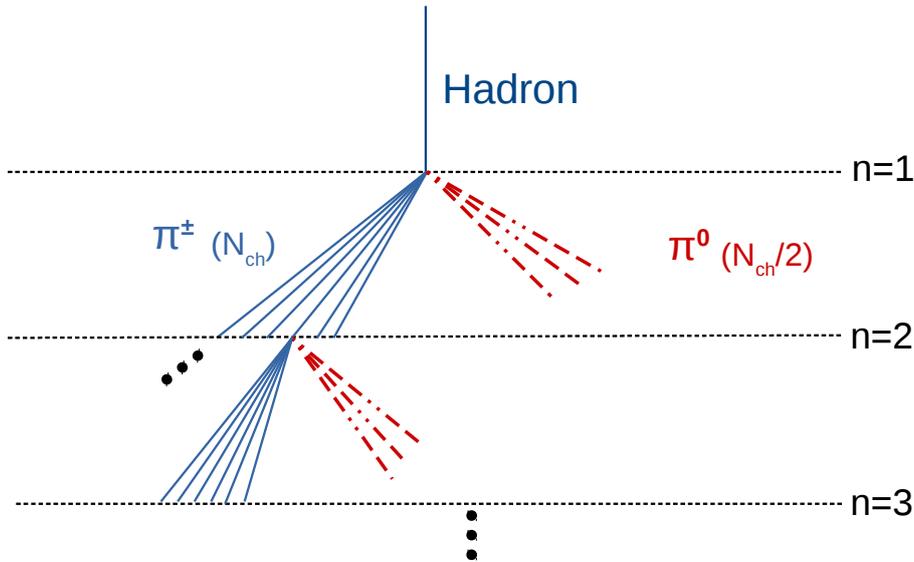
Figure 3.2: Scheme of the first three splits in a hadronic cascade. Redrawn following [95].

A good approximation of the energy passed onto the charged pions in each interaction i is around $E_i^{\pi,tot} = (2/3)E_{i-1}$. After n interactions, the energy carried by each individual charged pion is then, assuming equal energy distribution:

$$E_n^\pi = (\frac{2}{3}.\frac{1}{N_{ch}})^n E_0 \tag{3.3}$$

The critical energy for charged pion decay ξ_c^π is reached after $n_c = \frac{ln(E_0/\xi_c^\pi)}{ln((3/2)N_{ch})}$ interactions. The decay of a pion will produce a single muon, thus $N_\pi = N_\mu = (N_{ch})^{n_c}$. From this, the number of muons in the shower can be calculated:

$$N_\mu = \left(\frac{E_0}{\xi_c^\pi}\right)^\beta \text{ with } \beta = \frac{ln(N_{ch})}{ln((3/2)N_{ch})} \tag{3.4}$$

Due to the conservation of energy, the total energy of the shower can be written as the sum of the EM energy and the hadronic energy: $E_0 = E^{EM} + E^h$. At the maximal development of the shower, the total energy can thus be expressed as:

$$E_0 = \xi_c^e N_{EM}^{max} + \xi_c^\pi N_\mu \tag{3.5}$$

where N_{EM}^{max} is the number of EM particles in the subshowers at the maximal development of the shower and N_μ is the total number of muons produced in the shower.

At each hadronic interaction, about 1/3 of the charged pions energy is transferred from the hadronic to the EM component of the shower through the decay of neutral pions. As the energy of the primary increases, the number of times the charged pions are able to interact before decaying n_c increases and thus the fraction of energy converted into the EM component increases.

$$E^{EM} = E_0 \left(1 - \frac{N_\mu \xi_c^\pi}{E_0} \right) = E_0 \left(1 - \frac{1}{E_0} \left(\frac{E_0}{\xi_c^\pi} \right)^{\beta-1} \right) \simeq E_0 \left(1 - \left(\frac{2}{3} \right)^{n_c} \right) \qquad (3.6)$$

The depth at shower maximum X_{max} can be estimated using this model by taking only the EM component produced in the first interaction into account. This approximation is made because the maximal development of the EAS is dominated by the EM component. The first interaction happens at a depth X_0 which depends on the cross-section of the primary particle with the medium. Given the interaction length of a proton in the atmosphere λ_I, the depth of first interaction is $X_0 = \lambda_I ln(2)$. The X_{max}^p in a proton shower can then be estimated as the X_{max} of an EM shower starting at X_0:

$$X_{max}^p = X_0 + \lambda_r ln \left(\frac{E_0}{3 N_{ch} \xi_c^e} \right) = X_{max}^\gamma + X_0 - \lambda_r ln(3 N_{ch}) \qquad (3.7)$$

This approximation underestimates the real value of X_{max}^p since more subshowers are produced in subsequent neutral pion decays, however it correctly estimates the correlation between X_{max}^p and E_0: $X_{max}^p \propto ln(E_0)$.

For heavier nuclei, a good approximation is the superposition model, where the shower initiated by a nucleus of atomic number A is treated as the superposition of A showers initiated by protons with energies E_0/A. The shower observables can thus be calculated using the relation derived from the proton case:

$$N_\mu^A = N_\mu^p A^{1-\beta} \qquad (3.8)$$

and

$$X_{max}^A = X_{max}^p - \lambda_r ln(A) \qquad (3.9)$$

Although many approximations are made, this simple model for hadronic showers development make several correct predictions:

- the primary energy is proportional to a combination of the number of electrons and the number of muons: $E_0 \sim (N_{e^-} + \alpha . N_\mu)$ with $\alpha = 24$ considering $N_{ch} = 10$, $\xi_c^e = 85\,MeV$, $\xi_c^\pi = 20\,GeV$ $\lambda^I = 120\,g.cm^{-2}$ and $\lambda^r = 37\,g.cm^{-2}$ according to the approximations made in [95].

- the number of muons increases with the primary energy according to a power law with $N_\mu \propto E_0^\beta$

- the maximum development depth of a hadronic shower is shallower than for an EM shower, the effect is accentuated for heavier nuclei primary (from 3.9).

- the heavier the nucleus, the larger is the muon fraction

Obviously, these simplified analytical models cannot fully describe the complex processes happening as an EAS develops. A more accurate description of the content of a shower in terms of number and type of particles at a given slant depth and core distance gives the results which are shown in Figure 3.3. For a vertical shower initiated by a proton with an energy of 10^{20} eV, there is around $\sim 5 \times 10$ electrons and $\sim 2.5 \times 8$ muons at ground level (see Figure 3.4).
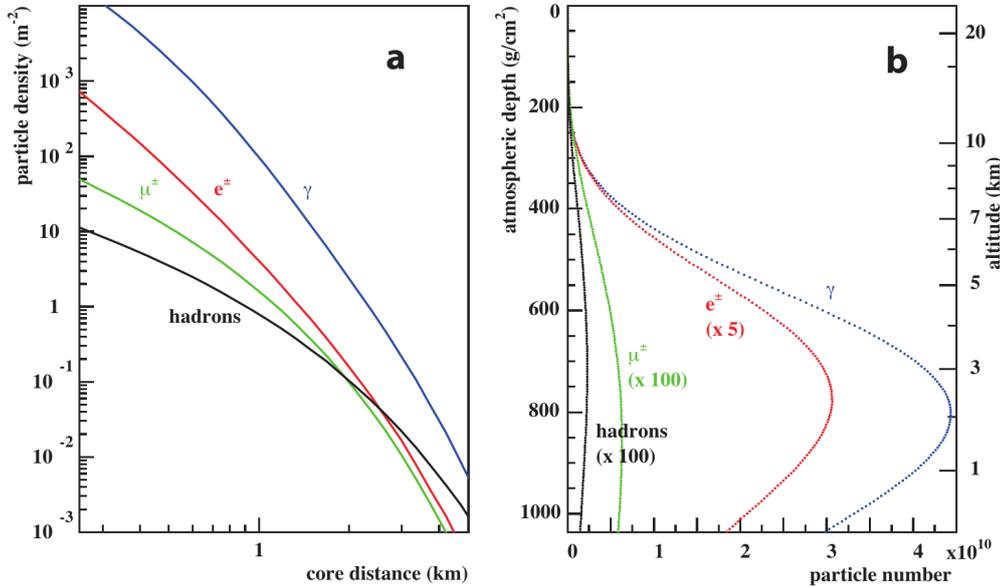
Figure 3.3: Average lateral (a) and longitudinal (b) profiles of vertical showers induced by a proton of energy $E_0 = 10^{19}$ eV simulated with CORSIKA [98]. The number of particles at ground (a) is calculated at depth $X = 870$ g.cm^{-2} which corresponds to the average ground-level for the Pierre Auger Observatory. From [99].

Also the amount of fluctuations from shower-to-shower is a very important information for some studies (like mass-composition studies) that is not at all described within these models. Full Monte-Carlo simulations, following each individual particle and its interactions are needed to model precisely an EAS development.

3.1.3 Lateral distribution of charged particles

The spatial and temporal lateral distributions of the secondary particles are correlated to the longitudinal development of the showers in the atmosphere and are thus also sensitive to the characteristics of the primary particle.

Starting from the first interaction point, the secondary EM particles from an EAS start spreading around the shower axis in multiple scatterings and travel to the ground in a curved shower front. A shower with a first interaction happening at lower depth (higher in the atmosphere) thus has its secondary EM particles spread wider at ground-level. Hadrons are scattered with a larger transverse momentum and travel further away from the shower core.

3.1.4 Electromagnetic and hadronic interaction models

To perform Monte-Carlo simulations (MC-simulations) of EAS, the electromagnetic (EM) and hadronic interactions must be evaluated. The electromagnetic interactions can be calculated very precisely from Quantum Electrodynamics, most EAS simulations software use standard EM interaction modules like the EGS4 code [100] or the Geant4 code [101]. Alternatively, analytical NKG treatment of the EM component can be performed [98]. Although the

hadronic interactions have been studied extensively in particle colliders, these experiments do not yet reach the energies encountered in UHECR collisions with the atmosphere. Also, the colliders are mostly studying collisions with high transverse momentum which accounts for only a fraction ($\ll 10^{-6}$) of the total number of particles produced and are described by Quantum Chromodynamics. However at UHE some mechanisms appear or become non-negligible that need to be added in the simulations specifically for the ultra-high energies, like the production of muon pairs or the LPM effect (see subsection 3.2.3).

These three main UHE hadronic interactions models are: EPOS [102][103], QGSJet [104] and Sibyll [105]. The models use different frameworks for hadronic interactions and thus when comparing the showers simulated using these different interaction models, one can find some discrepancies (see Figure 3.4). The three models have been compared in several studies [106][107][108] and the models are regularly updated.
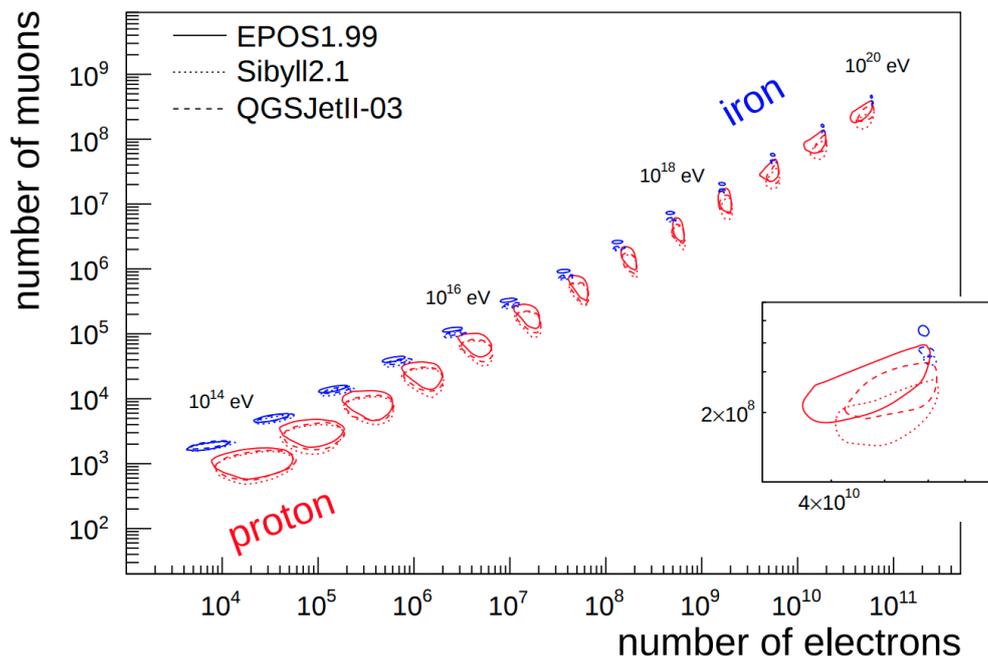
Figure 3.4: Number of muons and electrons in proton (red) and iron (blue) initiated showers, simulated with hadronic models: EPOS-LHC, QGSJet and Sibyll. From [106].

Comparatively, the low energy hadronic interactions models have less of an impact on the simulations since there are more data available at these energies. The codes commonly used for these low energy hadronic interactions are FLUKA [109], GHEISHA [110] and UrQMD [111]. The simulations of the showers in air shower experiments are usually performed with full Monte-Carlo codes (CORSIKA [112][98], AIRES [113][114]). These softwares track each secondary particle above a threshold energy as the shower develops. At UHE, the number of secondary particles becomes too important and a thinning has to be applied. The thinning keeps a representative fraction of the secondary particles and add weights to them while discarding the rest to reduce the calculation time [115]. Other simulation frameworks use hybrid codes, where the MC technique is combined with numerical solutions of cascade equations (CONEX [116], SENECA [117]). The hybrid codes save a lot of CPU-time but open the possibility of missing some rare or finer processes.

Muon deficit

It is important to note that all the current hadronic interactions models lead to a deficit in the number of muons in the simulated showers compared to data [118]. Indeed, if the muon content of simulated showers was correct, the much larger muon numbers measured with the PAO would correspond to a very heavy composition above 10^{19} eV, with elements heavier than iron. This interpretation is in disagreement with other mass-composition studies, based on the depth of shower maximum. For example, the discrepancy in the number of muons between data and simulations is estimated to be between 30% and 80% at 10^{19} eV [119].

Through the use of these models and codes, the simulation softwares track the primary through the atmosphere, calculate its interactions with the air nuclei or its decay and repeat this process for each secondary particle created (except those discarded by the thinning). This way, the full EAS is simulated. These simulations codes have been used by many air shower experiments to simulate EAS initiated by CRs ranging from proton to iron but also photons.

3.2 Photon-initiated showers

The phenomenology of photon-induced air showers is different from the hadron-induced showers and this can be used to differentiate between photon and hadronic showers. The main differences come from the fact that photon-initiated showers are EM showers. Also, some other photon-specific effects appear at the highest energies as described in the following sections.

3.2.1 Specificities of photon-initiated showers

A shower initiated by a photon will be a (almost) pure-EM shower. Thus, the structure and longitudinal development of the photon-induced showers (γ-showers) will differ from the structure of a proton-induced[1] showers (p-shower). Since the multiplicity of hadronic interactions is on average higher than in EM interactions, the primary energy is split through less particles in each interaction for γ-showers, thus the development of EM showers is slower. Thus the maximum number of particles in the shower is reached at lower altitudes on average for γ-showers. The hadronic component in γ-showers is very small as it is produced by photonuclear interactions, whose cross-section is small compared to the pair production cross-section (e.g. at 10 EeV, the photonuclear cross-section is estimated around 10 mb while the pair-production cross-section is estimated around 500 mb [120]). Additionally, the multiplicity and cross-section of hadronic interactions increases faster with energies than EM interactions, leading to further distinction between the average shower maximum $\langle X_{max} \rangle$ between γ-showers and p-showers (e.g. at 10^{20} eV, the average depth of maximal development $\langle X_{max} \rangle$ for p-showers is around 200 g.cm^{-2} less than for γ-showers). Fluctuations of the position of the X_{max} are also primary-sensitive. These fluctuations are affected by the depth of the first interaction of a shower and by the secondary interactions along the shower development. More fluctuations on the X_{max} are expected for γ-showers compared to p-showers. Also, the reduced hadronic component in γ-showers results in a lower number of muons on ground (around an order of magnitude less).

To summarize, the main differences are:

---

[1]The γ-showers are compared to p-showers since the differences are more pronounced for nucleon primaries with higher masses.

- the depth of maximum development of a γ-shower is on average deeper than for a p-shower (see Figure 3.5).

- the fluctuations on the X_{max} of γ-showers are more important than for p-showers (see Figure 3.5).

- the muonic component fraction is larger for p-showers than for γ-showers (see Figure 3.6).

- the lateral distribution function of γ-showers is steeper than for p-showers, since their first interaction point is on average located deeper in the atmosphere and they have less muons, which travel further away from the core compared to EM particles (see subsection 3.1.3).
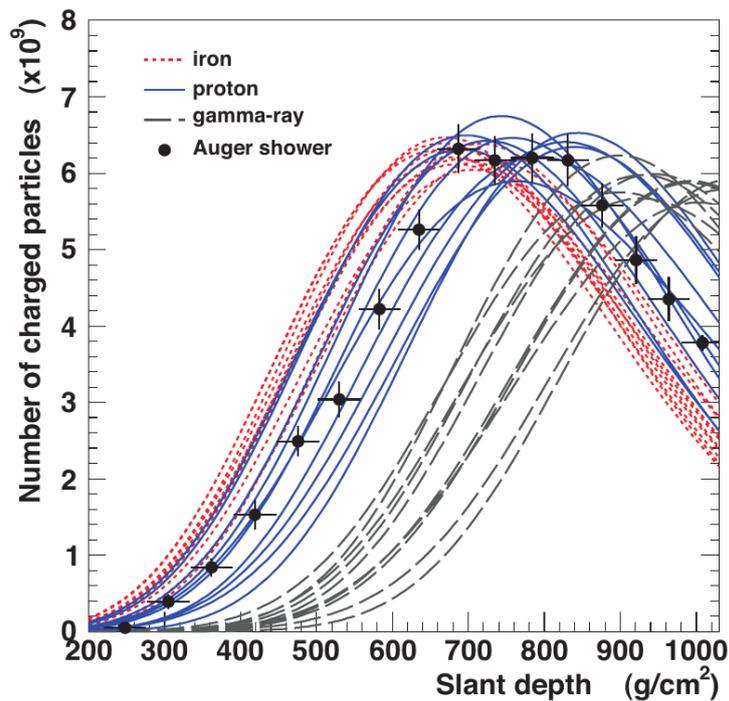
Figure 3.5: Longitudinal profiles of 10 proton-, iron- and photon-induced showers of energy $E_0 = 10^{19}$ eV simulated with Sibyll. The data points in black show the longitudinal development of a shower observed with the Pierre Auger Observatory at approximately the same energy. From [99].

The energy reconstruction performed by UHECRs observatories is tailored for hadronic showers, leading to the energy of photon-initiated showers being underestimated. Thus, a recalibration of the shower energy, making the assumption that the primary is a photon, is usually perfomed in photon analyses to properly estimate the photon energy E_γ.

3.2.2 Preshowering

At energies above $E_0 > 10^{19}$ eV a photon can convert into the geomagnetic field into an electron/positron (e^-/e^+) pair, before entering the atmosphere [121][122]. The probability of an UHE photon converting increases with the energy of the photon and with the strength of the transverse (respective to the propagation axis of the particle) magnetic field[1] $B_\perp$. The

---

[1]Thus it depends on the arrival direction of the primary
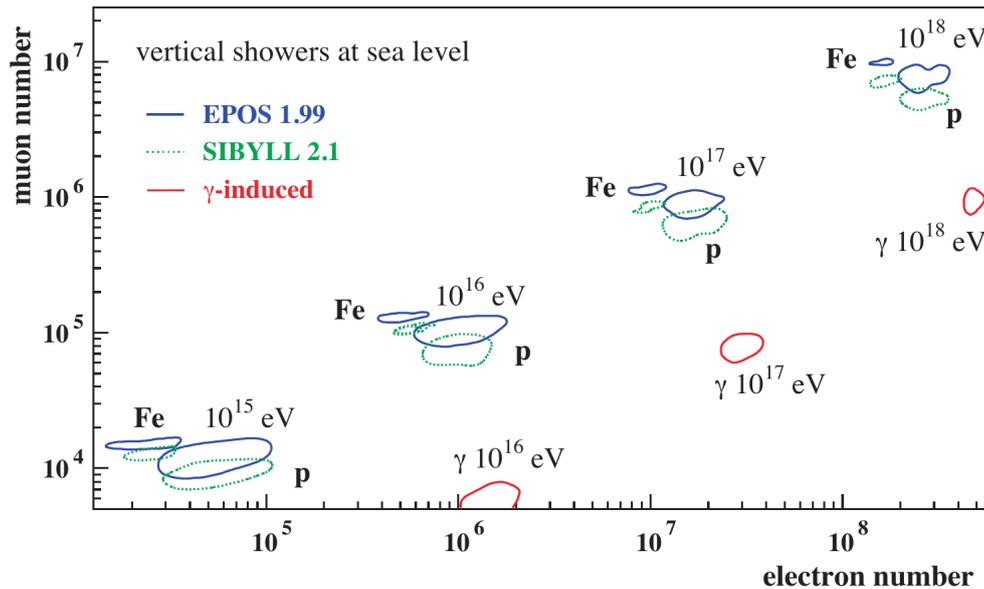
Figure 3.6: Number of electrons and muons on ground (sea-level) predicted for vertical showers initiated by proton, iron and photon primaries with energies $E_0 = 10^{15}$ eV, 10^{16} eV, 10^{17} eV, 10^{18} eV using CORSIKA. From [99].

resulting electrons can then lose energy through synchrotron radiation. The synchrotron radiated photons can then, if they are energetic enough, produce another e^-/e^+ pair.

Due to this process, several EM subshowers, initiated by less energetic photons and electrons, will be created in the atmosphere, instead of the initial more energetic one induced by the primary photon with energy E_0. Effectively, a preshowered γ-shower has a shallower X_{max} than a non-preshowered one.

3.2.3 Landau-Pomeranchuk-Migdal effect

At the highest energies, the electromagnetic interactions are reduced due to the Landau-Pomeranchuk-Migdal (LPM) effect [123][124]. The LPM effect appears when ultra-relativistic particles travel through matter and emit bremsstrahlung radiation or interact through pair-production. In this situation, less radiation is emitted and the pair-production cross-section is reduced compared to the case of an isolated particle [125].

Due to the LPM effect, the development of UHE γ-showers is slower in the atmosphere (deeper $\langle X_{max}\rangle$). Another important consequence is that since the LPM effect is energy dependent[1] (the threshold energy is about $\sim 10^{17}$ eV) the height of the first few interactions will have a greater impact on the longitudinal development of the showers. A shower that has a first interaction high in the atmosphere will be able to exit the LPM effect energy zone quicker, which will speed up the rest of its development. This leads to an increase in the shower-to-shower X_{max} fluctuations.

Furthermore, the LPM and preshower effect can both be present and have an impact on each other [126]: a preshowered event will be less affected by the LPM effect since the energies of each subshowers are lower. These two effects have an important impact on the development of γ-showers through the atmosphere (see Figure 3.7) and must thus be taken into account in the simulations.

In photon searches, the differences between γ-shower and p-showers presented in this section

---

[1]The LPM effect also depends on the matter density of the medium
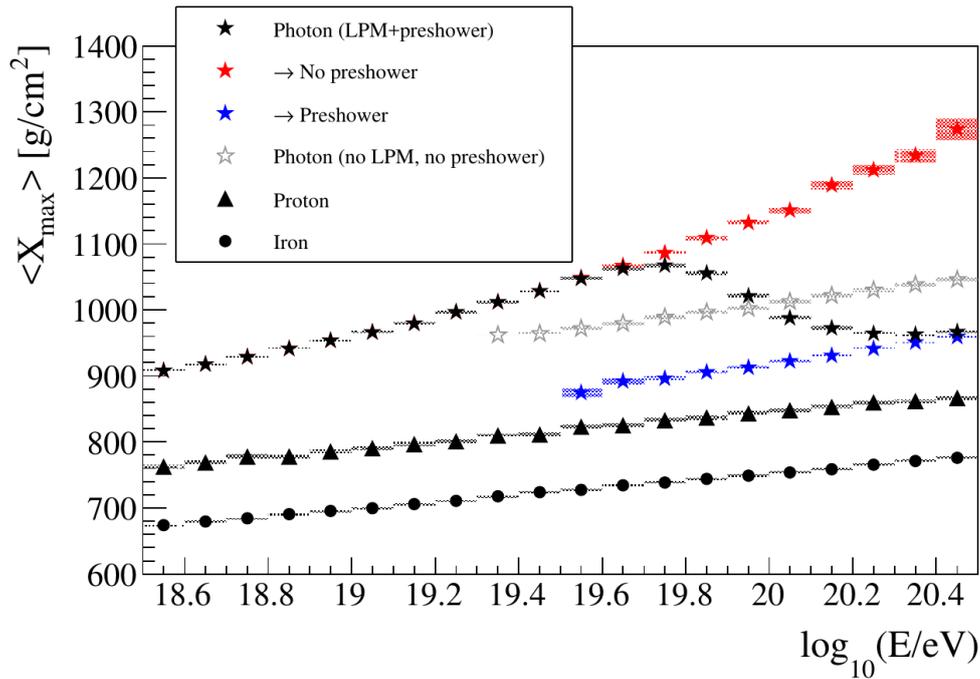
Figure 3.7: Mean X_{max} values of proton-, iron- and photon-induced vertical showers simulated with CONEX. The preshowered (red stars) and non-preshowered (blue stars) photons are drawn along the total photon population (black stars). A specific population of photon-showers, simulated without both LPM and preshowering effect is also drawn (hollow stars).

are used to select only the signal (γ-showers) by opposition to the background (p-showers) in the showers reconstructed with the data of the Pierre Auger Observatory (PAO). The LPM and preshower effects must also be taken into account and added to the simulations of the EAS to correctly reproduce the expected signal. Nonetheless, the development of proton- and photon-initiated showers can be very similar, in particular in the case of p-showers with a leading neutral pion. When the most energetic secondary in the first interaction is a neutral pion, the ensuing shower development resembles the development of an EM shower [127][128]. As the elasticity of the interaction increases, i.e. the leading particle can take a larger fraction of the primary energy, a leading π^0 p-shower can exhibit a development very similar to a γ-shower. The leading π^0, pre-showering and LPM effects are simulated and thus taken into account for the analysis presented in chapter 7.

Chapter 4

The Pierre Auger Observatory

Contents

In the previous chapters, we have shown that at ultra high energy, the cosmic rays have to be studied through the observation the extensive air showers they induce as they interact with the atmosphere. Ground-based observatories have been built to detect the secondary particles in order to reconstructed the air showers. In this chapter, the largest such observatory, the Pierre Auger Observatory, is presented. The history behind its construction is presented, before detailing its instruments. Then the air shower reconstruction procedure is presented, focusing on the surface detector reconstruction since it the one used in this work analysis. Finally, some of the main physics results produced by the Pierre Auger collaboration are presented.

4.1   The Pierre Auger Observatory and collaboration

In 1991, Jim Cronin (University of Chicago) and Alan Watson (University of Leeds) began to conceive a project to face the problem of the detection of cosmic rays at energies above 10^{19} eV. They concluded that the only way to address the questions raised by these astroparticles was to build an observatory designed as an array of detectors, spread over a large area to be able to detect enough particles above 10^{19} eV. During the early 1990s, the idea started to get traction and in 1994, the *Giant Array Project* workshop took place in Fermilab. For about a year, scientists from all over the world were invited to come to the Fermilab to participate on the design of the project, helping to refine it. Several designs for the surface detectors were proposed, using plastic scintillators, resistive plate chambers,... finally the water Cherenkov detectors, with a design similar to the ones already used in the Havera Park experiment [129] were selected. This choice was motivated in part because this kind of detectors was well known, cost-effective and because of its low impact on the environment. Later, the idea arose of building a hybrid observatory, adding another type of detectors having already been experimentally validated. Thus, air fluorescence detectors were added to the project, following the technique pioneered by the University of Utah in the Fly's Eyes experiment [11]. During this workshop in Fermilab, it was decided to name the Observatory after Pierre Auger.

The Observatory was initially supposed to be installed on two sites, one in the southern hemisphere, near Malargüe, Mendoza Province, in Argentina and the other in the northern hemisphere, in the state of Utah. An international collaboration was officially formed around the project during a meeting hosted by UNESCO in its Paris headquarter in november 1995. The collaboration consisted of 69 institutions, located in 17 countries from Oceania, Europe, North- and South-America.

In 2000, the first prototypes arrived on the Malargüe site while the real construction began in 2001, with the installation of the first water Cherenkov detector in the field. In 2003, the Pierre Auger Observatory became the largest cosmic-ray observatory in the world and in 2004, the Pierre Auger collaboration reported on the performances of a prototype instrument [130]. The data-taking officially started in 2004 and the construction finished in 2008. Along the years, complementary detection systems were deployed to enhance the capabilities of the Observatory : the low energy extensions with a denser array and specific fluorescence detectors, the Auger Engineering Radio Array (AERA), using antenna to exploit the coherent radiation in EAS, the Auger Muon and Infill for the Ground Array (AMIGA) to directly detect the muon content of EAS with buried scintillators. Being the largest and most precise observatory in the world, the Pierre Auger Observatory is the present flagship experiment studying ultra-high energy cosmic rays.

4.2   Detection systems of the observatory

The surface detector samples the particles in a slice of the showers at ground-level, using an array of water Cherenkov detectors (WCD), which are sensitive to both muonic and electromagnetic particles. The individual stations (WCD+local electronics) are self-sufficient detectors and their signals are centralized and combined by the central data acquisition system (CDAS) to select shower candidates. The fluorescence detector (FD) measures the longitudinal development of the showers by looking at the fluorescence light emitted as they developp in the atmosphere. The fluorescence telescopes overlook the surface array and record the profile of the showers, providing new and complementary information on the shower detected by the surface detector.

A hierarchical system of triggers is implemented starting from the individual detector stations up to the offline events reconstruction. At the station level, the trigger are designed to remove most of the atmospheric muons background and to record, at the CDAS level, only signals compatible with a shower event. Then the rest of the triggers are applied online, on the combined array signals.

4.2.1 Water Cherenkov Detectors and Surface Detector

The first building block of the Pierre Auger Observatory (PAO) is the Water Cherenkov Detector (WCD). The outer-shell of the WCDs is made of polyethylene, with an average wall thickness of 1.3 cm and it is light-tight. Inside the plastic shell, a sealed liner with a reflective inner surface is installed. The liner is a cylinder with a diameter 3.6 m and a height of 1.2 m. It is filled with 12000 liters of ultra-pure water. Three photomultiplier tubes (PMTs) are symmetrically distributed above the liner, looking down into the water. The read-out electronic is located on top of the detector, above one of the three PMTs. A GPS receiver and an antenna are installed on the WCDs to time-stamp and transmit the data collected. Each WCD unit is self-sufficient in power thanks to a solar power system with solar panels mounted on top of the outer-shell and batteries.
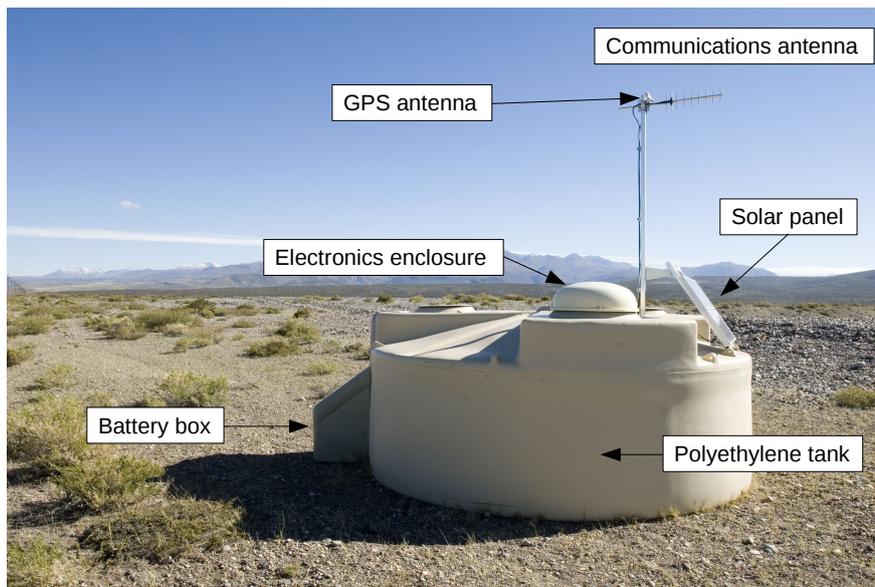
Figure 4.1: Picture of a Water Cherenkov Detector in the pampa, with annotations indicating the main elements.

When a relativistic charged particle goes through the water, the PMTs collect the cherenkov light produced. The WCDs are also sensitive to high energy photons converting into electron/-positron pairs in the water. The ultra-pure water and the reflective liner increase and unify the light collection of the detector. The PMTs signals are digitized by flash analog-to-digital converters into FADC traces. Each PMT has two outputs, an anode signal and a last dynode signal amplified and inverted to provide a signal with 32 times the charge gain of the anode. The anode signal is used if the dynode signal is saturated.

The FADC traces are locally calibrated into vertical equivalent muon (VEM) units, the signal deposited by a vertical through-going muon into the detector. To perform the local calibration, 60 s of background signals are collected to build the histogram of the integrated charge Q in ADC units (see Figure 4.2). Then the peak, corresponding to the charge deposited by atmospheric muons is fitted and its position is set as the ADC to VEM reference after
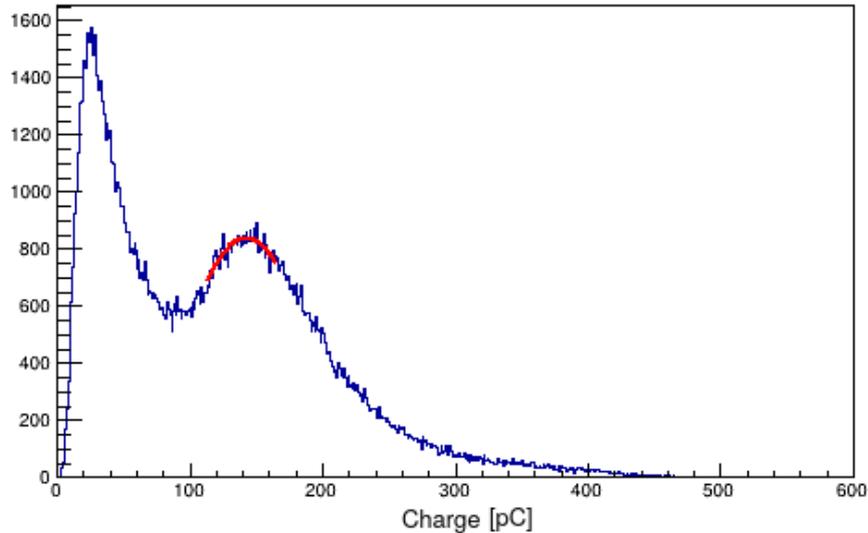
Figure 4.2: Example of a fitted charge histogram. The bin in which the maximum of the peak is reached, around $150\,\mathrm{pC}$, from the fitted part of the histogram, corresponds to the VEM reference.

applying a correction factor to account for the fact that the throughgoing muons are omni-directional. In all recorded events, the charge histograms are stored to perform a more precise offline calibration. This calibration allows the comparison between the individual WCDs through a common reference level. The offline calibration of the charge histograms allows a precision of the VEM unit calibration of $\sim 3\%$.

The total signal of a station is the sum of the VEM trace bin values between a start-time and a stop-time. The FADC traces are scanned through to locate signal fragments, consisting of consecutive bins above 3 ADC counts. Once these signal fragments are located, they are merged[1] into a single trace and the start-time and stop-time are assigned accordingly (see Figure 4.3). Finally, the signal traces are averaged on the three PMTs.
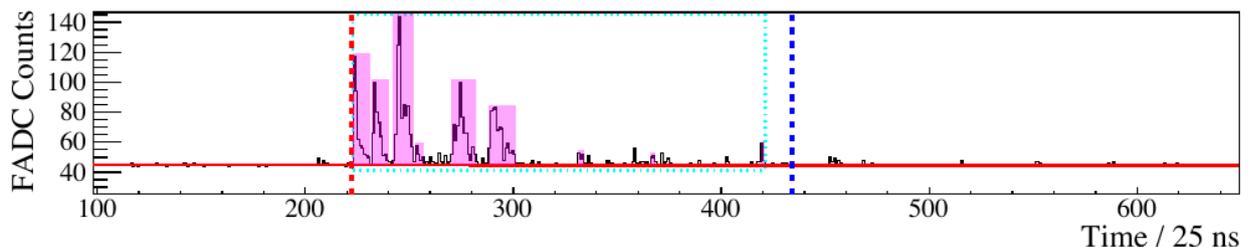
Figure 4.3: Example of a high gain FADC trace from a water cherenkov detector PMT. Signal fragments are highlighted in the purple boxes and the signal merging window is shown as the indigo dashed box. The total signal is the sum of the FADC counts in the bins between the start and stop time represented by the vertical dashed red and blue lines respectively. Taken from [131].

For non-saturated stations, the uncertainty on the signal S varies with the zenith angle θ and according to a poisson-like parameterization:

$$\sigma_s^2(\theta) = f_S^2(\theta)S \text{ where } f_S^2(\theta) = 0.34 + 0.46.\sec(\theta) \tag{4.1}$$

For saturated stations (stations with saturated PMTs) however, the uncertainty goes from

---

[1]the fragments are scrutinized before merging to ensure that no background signal is added to the signal trace.

$\sim 5\%$ up to $\sim 60\%$ for signals around $10000\,\mathrm{VEM}$.

The individual water cherenkov detectors are combined into an array to form the surface detector (SD) (see map in Figure 4.4). The main array of the SD is composed of 1660 WCD stations, spread over an area of $\sim 3000\,\mathrm{km^2}$ in a triangular pattern, where two neighboring stations are separated by a distance of $1500\,\mathrm{m}$. Another denser array in the SD, where the stations are separated by $750\,\mathrm{m}$, covers $28\,\mathrm{km^2}$ and studies air showers of lower energies compared to the SD-1500 array.
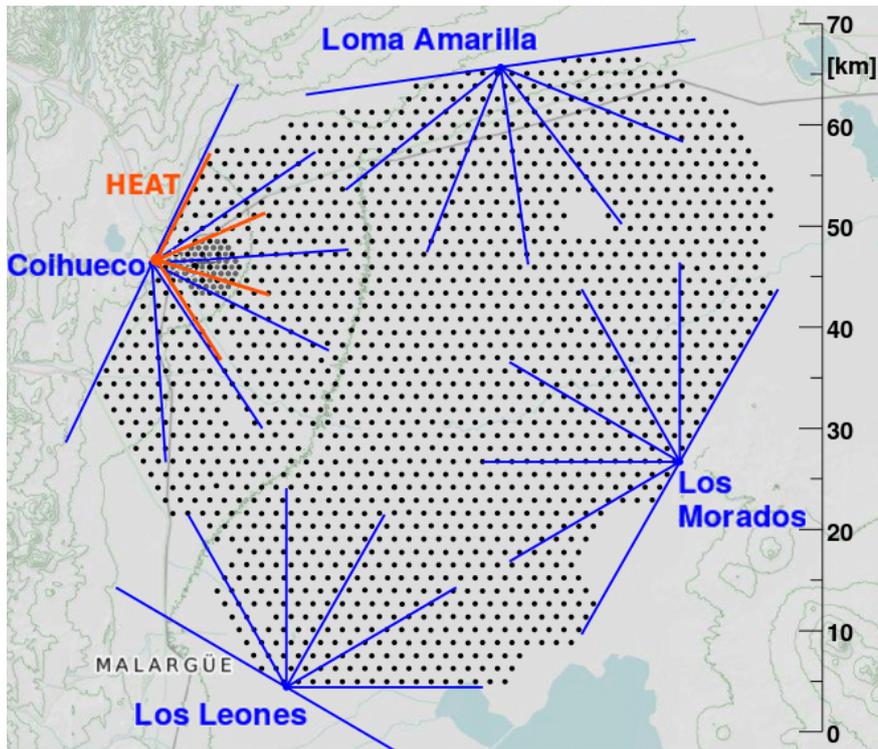
Figure 4.4: Map of the Pierre Auger Observatory. The stations of the surface detector are represented with the black points. The four telescope sites of the fluorescence detector are marked in blue.

The area over which the stations are spread is generally flat and located at an altitude between $1310\,\mathrm{m}$ and $1610\,\mathrm{m}$. The mean altitude at which the stations are located is $\sim 1410\,\mathrm{m}$. This high altitude is suited to have the surface detector slightly below the maximum of development for the vertical showers.

4.2.2 Surface Detector triggers

These traces are then kept in a local memory and local trigger algorithms are applied to them. The local triggers[1] decide whether the FADC traces are recorded or not. In the individual stations, the rate of signal detection is around 3 kHz [132][133], too high to record every single one. The first level of selection for the trigger, T1, has two modes:

- the simple threshold trigger Th. It requires that all three PMTs simultaneously record a signal over a $1.75\,\mathrm{I}_{VEM}$ threshold, where I_{VEM} is the pulse height value of a VEM. It selects strong signals that are not necessarily spread in time.

---

[1] other types of local triggers exists, the scaler triggers and the calibration triggers. The scaler triggers have a very low signal threshold and provide information on space weather. The calibration triggers have a low signal threshold and are used to control the general CR activity at the observatory as well as to record the stations calibration histograms.

- the time-over-threshold trigger ToT. It requires that at least two out of the three PMTs record a signal above $0.2\,\mathrm{I}_{VEM}$ for 13 bins and within a $3\,\mu s$ window. This trigger selects series of low signals spread in time.

A second level of selection, T2, is implemented in a way so that the triggering rate is down to $\sim 20 - 25\,\mathrm{Hz}$, in order to keep 50% of the bandwidth free for data transmission. A T1-ToT is automatically upgraded to a T2-ToT but a T1-Th is required to be above $3.2\,\mathrm{I}_{VEM}$ to be promoted to T2-Th. All the T2s are sent to the CDAS each second with their time stamps to determine the global trigger conditions (T3).

The CDAS collects all the T2 sent by the individual stations and construct the level 3 triggers T3. The main trigger condition of the T3 corresponds to the shower candidates. It is based on spatial and temporal conditions of the T2s.

The main condition of the T3 has two modes. To describe these modes, the notion of hexagon cells must first be introduced. For an active station, the 6 closest stations surrounding it form the first crown and together the 7 stations form an hexagon. Outside of this hexagon, the 12 stations surrounding it form the second crown. The third crown is formed by the 18 stations surrounding it, and so on.
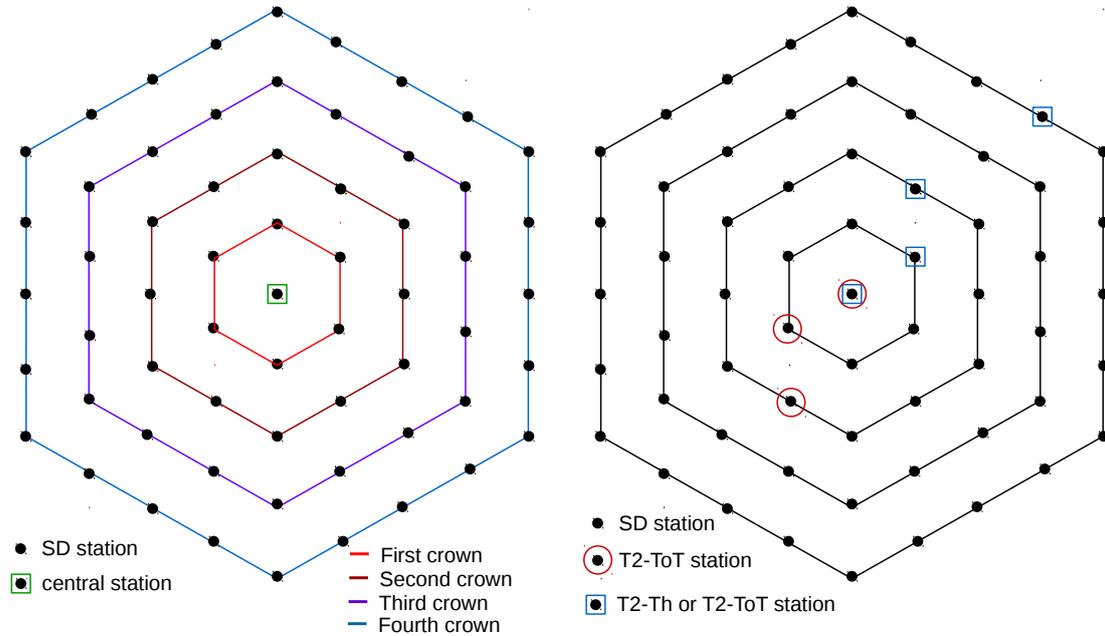
Figure 4.5: Scheme of the SD stations crowns. On the left, the crowns from first to fourth, centered on a central SD station, are represented. On the right, examples of the T3 main trigger geometrical conditions are shown.

The first T3 mode requires to have at least 3 stations passing the T2-ToT trigger in coincidence, with a triggered station on the first crown and another no further than the second crown. The second one requires to have at least 4 stations passing either the T2-Th or the T2-ToT trigger in coincidence, with one station as far as the fourth crown if a station is on the first one, and another station is on the second crown. A scheme of the crowns, as well as examples of the T3 main trigger conditions are shown in Figure 4.5.

When a T3 main trigger is identified, the CDAS records all the traces from the T2 and T1 triggered stations within six crowns of the central station. The rate of T3 triggers is around $0.1\,\mathrm{Hz}$.

4.2.3 Fluorescence Detector

The fluorescence detector (FD) is composed of 24 telescopes, located in 4 sites (Cohueico, Loma Amarilla, Los Leones and Los Morados). Each site has 6 independent telescopes, located in separated, clean buildings, overlooking the SD. A fluorescence telescope is composed of an aperture system, a UV filter, a segmented spherical mirror and a camera, located in the focal plane and composed of 440 photomultipliers.

As an EAS develops through the atmosphere, the secondary particles excite N_2 molecules, which in turn emit fluorescence light isotropically. Cherenkov light is also induced by the showers and picked up by the FD. This Cherenkov light is reconstructed based on the atmospheric conditions at the time of the detection. When the light produced by the showers arrives at the telescope building, it first passes through a diaphragm and a UV filter. The light is then focused by the segmented spherical mirror before reaching the camera. The fluorescence detector operation is limited to clear moonless night, which leads to a duty-cycle of $\sim 15\%$.

To calibrate the FD, laser shots are fired regularly during the FD operations. Also, the atmospheric conditions above the Observatory are constantly monitored and atmospheric databases are regularly filled and updated.

The observation of a shower with the FD provides two important measurements:

- the fluorescence light emission is correlated to the amount of energy lost by the shower into particle collisions. The atmosphere above the array thus acts as a calorimeter and it provides an independent estimation of the calorimetric energy of the showers.

- by observing the profile of the fluorescence light emitted by the showers through the atmosphere, the atmospheric depth at which the showers reach their maximal development X_{max} can be measured directly (see Figure 4.6).
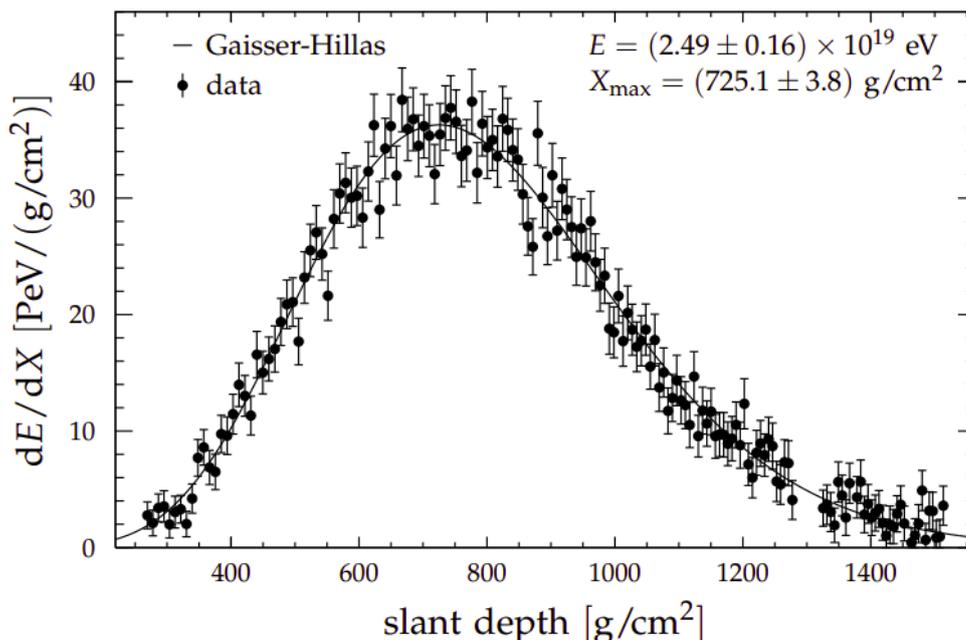
Figure 4.6: Example of a longitudinal profile reconstructed by the FD (dots) as a function of the slant depth in the atmosphere. A Gaisser-Hillas function is fitted to this longitudinal profile to estimate the X_{max} and the energy of the shower. Taken from [134].

Aside from the calorimetric energy measured by the FD, there are other contribution to a shower total energy. The "invisible energy" held by the muons and the neutrinos that do not interact before reaching the ground. The total energy of a shower is estimated by applying a correction on the calorimetric energy to account for the invisible energy. This correction is estimated through a data-driven analysis at about $\sim 14\%$ at 2.5×10^{18} eV and goes down with increasing energy to about $\sim 12\%$ at 10^{20} eV [135]. A dataset of hybrid events, simultaneously triggering the SD and the FD, and passing strict quality selection criteria are used to guarantee a precise and unbiased E_{FD} measurement. Additionally, the primary energy estimation of the FD is used to calibrate the SD energy estimation. This is the main advantage of the hybrid design of the observatory.

Three additional fluorescence telescopes were added on the Cohueico site, near the building housing the regular fluorescence telescopes. They are called the high elevation auger telescopes (HEAT) and they have two distinctive features compared to the standard Auger FD:

- they can be tilted 29° upward.

- their electronic kit allows a higher sampling (smaller time bins). This improves the measurement of showers developing closeby to the HEAT telescopes.

The HEAT telescopes work independently from the other fluorescence telescopes and their aim is to lower the energy threshold of the hybrid dataset. Thus they overlook the SD 750 m array.

4.3 Shower reconstruction

From the information collected by the CDAS from the detection systems, shower parameters can be reconstructed, ultimately leading to estimations of the primary particles characteristics. In this section, the steps of the SD event reconstruction are detailed. Then, the performances of the reconstruction are presented.

The hybrid reconstruction collects shower signals in the camera pixels and reconstructs the shower geometry from the SD+FD information. Then, the shower profile is fitted using a Gaiser-Hillas function, allowing the estimation of the maximal development of the shower X_{max} and the primary energy E_{FD}. The hybrid reconstruction is not explained further in this work as it has not been used in the analyses. The interested reader can learn more on the hybrid reconstruction in [136].

4.3.1 SD Reconstruction

It is noted here that there exists two reconstruction frameworks in the Auger collaboration: the *Observer* and the *Herald*. For this work, only the *Observer* reconstruction has been used, the *Herald* reconstruction will thus not be described. Moreover, only the reconstruction of "vertical" showers (zenith angle lower than 60°) is considered

Event-triggers

Before reconstructing the shower based on the T3 data, two additional triggers are applied. The T4 trigger is the physics trigger. It ensures that the data selected only contains real showers by requiring that the selected stations have signals coherent temporally and topographically with the propagation front of a shower. To estimate the shower front, a *seed* reconstruction is performed. This first reconstruction approximates the core of the shower

on the ground as the signal-weighted center-of-mass of the stations in the event. Then the direction of the shower axis is estimated by comparing the time at which the shower front passes through the stations.
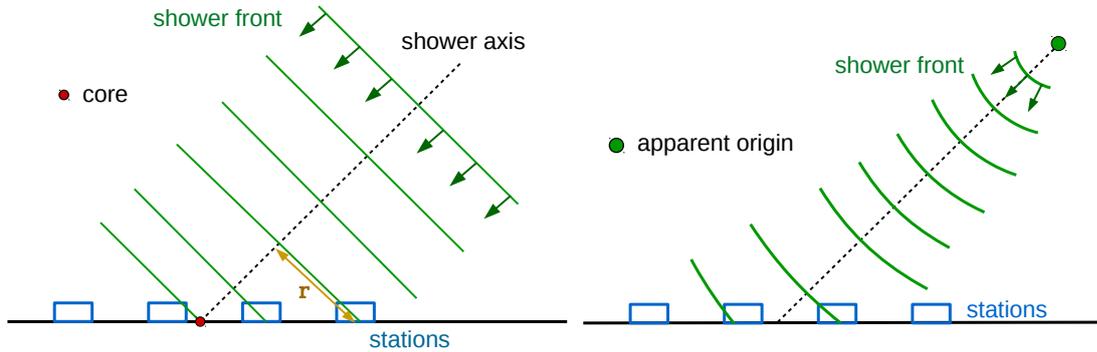
Figure 4.7: Scheme of the shower front estimation. Left: the first seed reconstruction performs a planar shower front fit on the individual stations. The distance of a station to the shower axis (r) is represented. Right: representation of a spheric shower front.

Stations with signals that are not compatible with the expected shower front are set as *accidental* stations and are not used in the reconstruction, such as signals produced by random muons. The limit for time-delay between recorded time and expected time of the signals in a station is set so that more than 99% of the stations receiving physical signals from the showers remains in the reconstructed events.

A fiducial cut is then applied, the 6T5 trigger, to ensure an accurate reconstruction of the showers. It requires that all 6 stations forming the first crown around the station with the highest signal were all active at the time of the signal detection (the shower core is within the hexagon).

Shower geometry

Using the stations selected according to the method presented in the previous paragraph, the geometry of the showers is reconstructed. A spherical shower front is used to estimate the arrival time of the shower front particles at a given position. This arrival time is compared to the signal start of each triggered station. The shower axis, the time at which the core hits the ground and the radius of curvature[1] are fitted using a MINUIT minimization [137] on the arrival time of the signal in the triggered stations.

Shower size and Lateral Distribution Function

The next step of the shower reconstruction is to estimate the *size*[2] of the shower. The lateral distribution function (LDF) is an empirical function, which describes the evolution of the signal with the distance r from the shower core. The distance r is the perpendicular distance between a station and the shower axis (see Figure 4.7).

The size of a shower is directly correlated to the energy and the mass of the primary. The LDF also changes depending on the arrival direction of the primary. Fitting the LDF thus brings information on these three primary CR characteristics. However fluctuations occur,

---

[1]only if more than 4 triggered stations are in the event, otherwise there are more free parameters than there are stations available for the fit.

[2]the *size* of a shower is the number of particles in the shower.
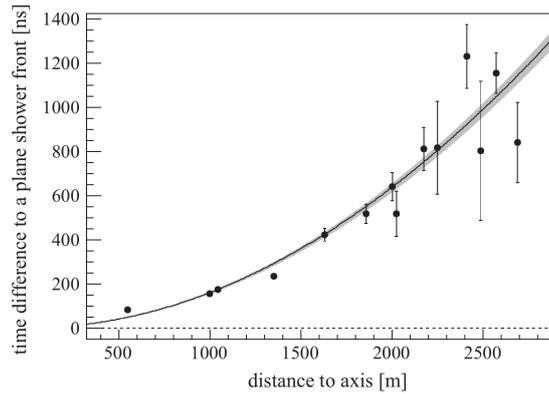
Figure 4.8: Geometrical reconstruction of a shower. As the stations get further away from the core, the time-difference between the expected start time from a planar shower front and the measured one gets larger. From [133].

due to the stochastic behavior of the first interaction (position, particle production). Also, the reconstruction of the shower can be deteriorated by the changing atmospheric conditions and the non 100% uniformity of the SD. The optimal distance r_{opt} is defined as the distance at which the shower-to-shower fluctuations are the weakest. For the 1500 m array, $r_{opt} \simeq 1000$ m. The expected signal at distance r_{opt} of a shower $S(r_{opt})$ is a shower-size estimator. As such, $S(r_{opt})$ is also correlated to the energy of the shower.

Most events do not have enough triggered stations to fit the shape of the LDF, hence a scaled data-averaged of the LDF is calculated and the stations signals are fitted on it. The LDF uses a modified NKG function [138][139][140]:

$$ f_{\text{LDF}}(r) = \left(\frac{r}{r_{opt}} \right)^{\beta} \left(\frac{r + r_s}{r_{opt} + r_s} \right)^{\beta + \gamma} \tag{4.2} $$

where $r_s = 700$ m. The averaged slope parameters β and γ are fitted on selected data sample and then parameterized as a function of the zenith angle θ and the expected signal at 1000 m, S_{1000}.

Finally, a log-maximization is performed to fit the S_{1000} and the core location of the events. It is important to note that the (functioning) non-triggered stations are also used in the fit, to constrain the size of the showers.

The relative uncertainty on the reconstructed S_{1000} varies with the shower size. Larger showers have a more precise reconstruction of S_{1000} down to about 6% for showers with $\log(S_{1000}/\text{VEM}) > 2 \, \text{VEM}$[1]. The average uncertainty is $\sim 14\%$ and it is relatively constant across the values of θ.

Arrival direction

The arrival direction of a shower is obtained from the geometry reconstruction and the LDF fit, which respectively estimate the point of origin and the position of the shower core at ground level.

The angular resolution depends on the size and zenith angle of the showers, but it is always better than 1.4° and improves with increasing number of triggered stations reaching less than one degree for $N_{stations} > 5$

---

[1]until $S_{1000} > 200 \, \text{VEM}$ where the presence of saturated stations reduce the quality of the reconstruction.

Energy reconstruction

The aim of the energy reconstruction procedure is to build an estimator that is independant of the zenith angle θ. The shower-size estimator S_{1000} depends on the shower zenith angle θ, as a higher θ means that the shower has traversed more atmospheric depth, and is thus more attenuated than a vertical one, having the same energy. To estimate the energy of a primary, this attenuation effect must thus be corrected. This is done by following the constant intensity cut (CIC) method [141].

Since the flux is isotropic, any significant fluctuation in the flux is due to the attenuation and is θ-dependent. The attenuation curves is characterized by the CIC method as a function of θ and a θ-independant shower-size estimator, S_{38}, is built. S_{38} is the S_{1000} value a shower would have if it had a $\theta = 38°$.

The conversion from S_{38} to the SD-reconstructed energy E_{SD} is obtained through a calibration with the FD (see subsection 4.2.3). The systematic error of the E_{SD} reconstruction is $\simeq 14\%$ and comes mainly from the uncertainties on E_{FD}. The bias on E_{SD} is larger than 20% at energies around 10^{18} eV but tends to 0% in the full-efficiency regime (see subsection 4.3.2). The resolution on E_{SD} is around 20% at energies around 10^{18} eV and tends to 7% above 2×10^{19} eV.
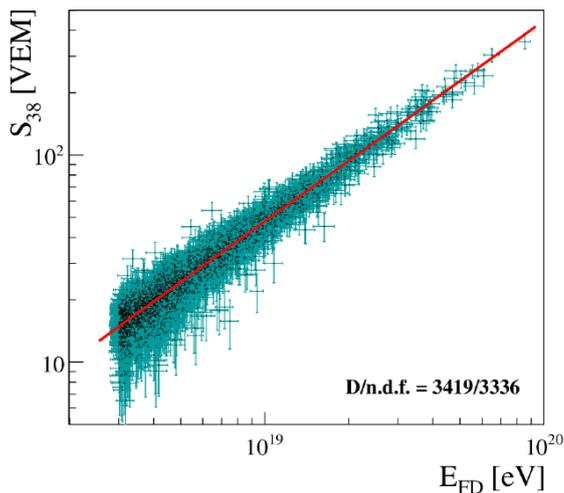
Figure 4.9: Calibration of the SD shower-size estimator S_{38} using the energy reconstructed by the FD E_{FD}. The events used to perform the calibration passed very strict quality selection criteria to ensure that the FD energy is reconstructed properly, and that no biases are introduced in the energy of the showers observed by the FD. From [142].

4.3.2 Efficiency and exposure

The trigger efficiency of the SD characterizes the probability a shower has of triggering the array. For a given primary energy, the trigger efficiency varies depending on the mass of the primary and on the trigger conditions required. For events with at least 3 stations triggered the triggering probability reaches 100% above $\sim 3 \times 10^{18}$. Using simulations where the trigger conditions are reproduced, the trigger probabilities are calculated for three types of primary: proton, iron and photons (see Figure 4.10). The showers initiated by iron primaries reach full-trigger efficiency at lower energies compared to ones due to proton primaries. The photon primaries generate showers deeper into the atmosphere (smaller footprint) and with less muons, hence they reach full trigger efficiency at higher energies compared to proton, around $10^{19.5}$ eV.
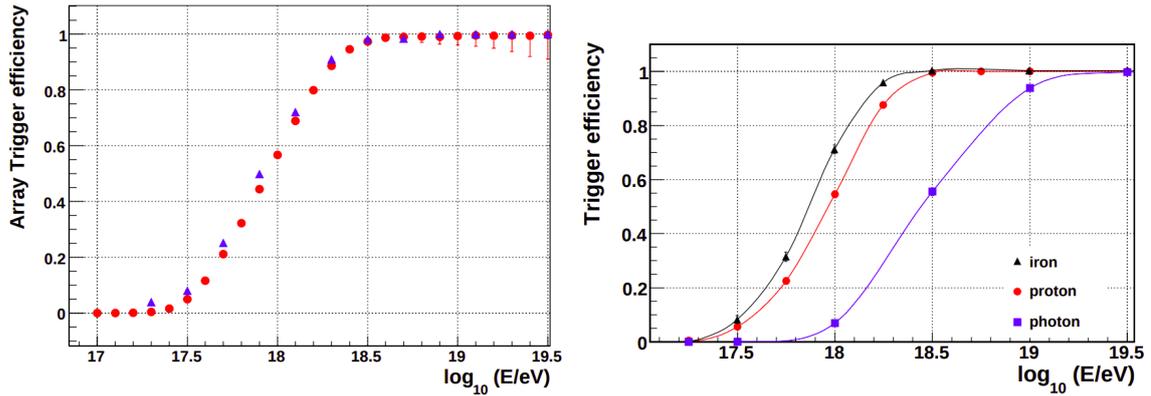
Figure 4.10: Trigger efficiency of the SD 1500 m array as a function of the primary energy. Left: trigger efficiency estimated from SD data (red dot) and from hybrid data (blue triangles). Right: trigger efficiency obtained from Monte Carlo simulations of showers initiated by proton (red dots), iron (black triangles) and photon (blue squares) primaries. The full trigger-efficiency is reached at higher energies for photons than for hadronic showers, since they tend to produce deeper showers with a smaller muonic component. From [143]

The aperture of the SD array quantifies its observation capabilities, at a given time, for a given shower energy and for a specific trigger condition. It is defined as the integrated probability to have the specific trigger conditions met, for a shower of energy E, over the effective area of the SD at an instant t.

In most cases, the aperture is calculated for 6T5 events, above full-trigger efficiency. In this case, the aperture of the surface array can be calculated as the sum of the aperture of each active hexagon cell in the surface array. A hexagon cell is the combination of a central station and its 6 closest neighbors (see Figure 4.11). Each individual cell contributes $\simeq 4.59\,\mathrm{km^2.sr}$ to the total aperture of the SD.
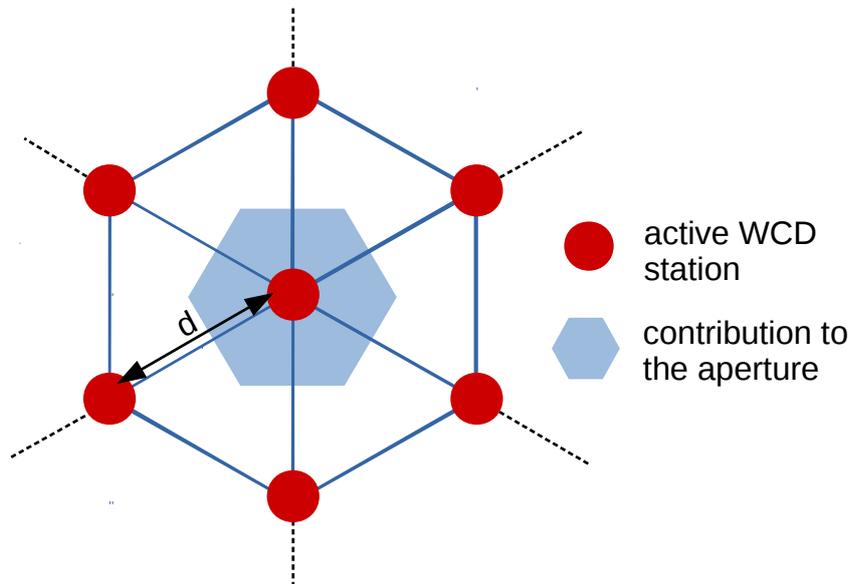
Figure 4.11: Scheme of an hexagon cell. The aperture of each active hexagon cell of the SD is represented as the light blue area.

The exposure of the array is the time integrated aperture. The total exposure collected by the Pierre Auger Observatory between 01/01/2004 and 31/05/2020[1] for 6T5 events is

---

[1] last date available for exposure calculation at the time of writing.

$69870 \, \mathrm{km^2}$.sr.year. It is used to make CR flux estimations.

4.4 Main results of the Pierre Auger Observatory

For more than 15 years, the Pierre Auger Observatory has investigated the Ultra High Energy Cosmic Rays. The combination of a large array of surface detectors covering $3000 \, \mathrm{km^2}$ and fluorescence telescopes enhances the performances of the extensive air shower detection and measurements, resulting in both high statistics and unprecedented data quality. Moreover, the operation of a denser sub-array has extended the sensitivity to lower energies. Altogether, these well performing detectors contribute to provide a wealth of important and valuable information on key questions on cosmic rays. The characteristics of the primaries initiating the showers detected by the observatory have been studied, the energy spectrum measured, and the composition of the flux estimated. The study of the origin of UHECRs also requires the careful study of their arrival direction, to find indications of anisotropies and try to locate potential acceleration sites. Also, recent years have shown the importance of multi-messenger studies in astrophysics and the Pierre Auger Observatory is participating to these efforts at the ultra high energies. Besides the UHECR fields, the Pierre Auger Observatory published results in other physics fields, in particular concerning the atmosphere.

In the following, the most important and recent results are presented, concerning the UHECR energy spectrum, the mass composition, and the arrival directions studies. More detailed information can be found in [144]. The searches for the neutral particles are more detailed since it concerns this thesis work.

4.4.1 Energy spectrum

The energy spectrum of UHECRs is one of the main result of the Pierre Auger Observatory. In fact the PAO produces an energy spectrum for five independent datasets, corresponding to different detection set-ups and their dedicated reconstruction procedures [142] :

- the SD 1500 m array for vertical showers ($\theta < 60°$), above $2.5 \times 10^{18} \, \mathrm{eV}$.

- the SD 1500 m array for horizontal showers ($\theta > 60°$), above $4 \times 10^{18} \, \mathrm{eV}$.

- the hybrid reconstruction, above $10^{18} \, \mathrm{eV}$.

- the SD 750 m array, above $10^{17} \, \mathrm{eV}$.

- the reconstruction of showers detected by the HEAT telescope through their Cherenkov light, between $10^{16.5} \, \mathrm{eV}$ to $10^{18} \, \mathrm{eV}$.

Using a maximum likelihood method to combine all five spectra, allowing a shift in the normalization of each of them, a combined spectrum is established over more than three decades in energy, as shown in Figure 4.12. The overall systematic uncertainty on the energy scale is 14% [145]. Some characteristic features are clearly visible, and a fit is performed with an empirical function, the form of which is driven by the structures and inflection points observed, leading to the determination of the spectral indices and the energy of the inflection points. The identified features are:

- an inflection point commonly known as the second knee around $10^{17} \, \mathrm{eV}$

- a dip corresponding to the ankle around $6 \times 10^{18} \, \mathrm{eV}$
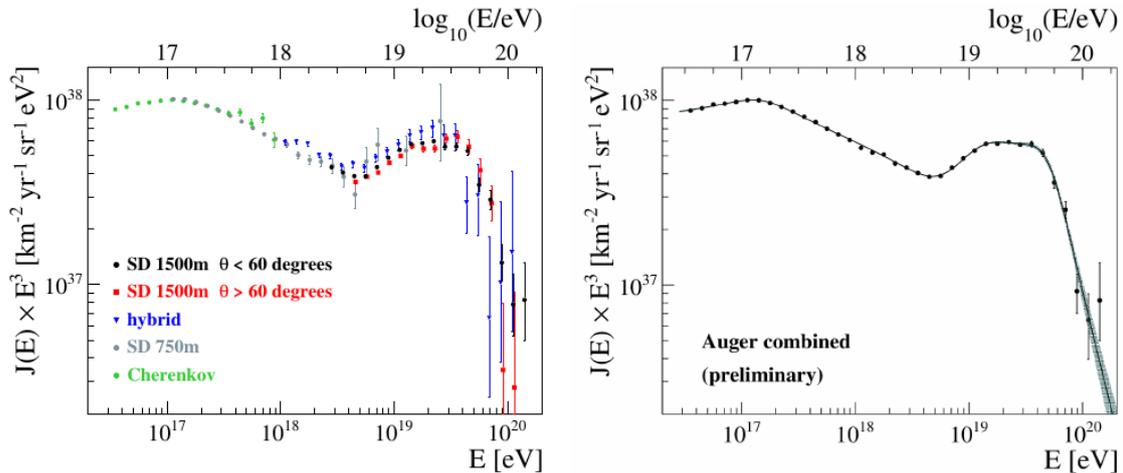
Figure 4.12: Energy spectra measured by the Pierre Auger Observatory (left) and the corresponding combined spectrum fit (right). From [142]

- a flattening is visible around 10^{19} eV

- a flux supression above 5×10^{19} eV

The careful and precise reconstruction of the spectrum using the vertical showers detected with the SD 1500 m array has been described recently[1] in a detailed publication [146]. The features observed above 2.5×10^{18} eV have been clearly highlighted, in particular the new one observed in the region above the ankle. Moreover no significant dependence of the spectral features on the declination has been found in the studied range.

These features identified in the energy spectrum can only be interpreted by looking at the results from other analyses, in particular the mass composition studies.

4.4.2 Primary composition

The average mass composition of shower primaries is an important observable to understand the acceleration and propagation processes of UHECRs. The primary composition of the showers detected by the Observatory is best estimated by the FD, using the average depth of maximal development of the showers $\langle X_{max} \rangle$, as well as $\sigma(X_{max})$ which is correlated to the shower-to-shower fluctuations of the primaries mass and to the mass-spread/the purity of the primaries (see Figure 4.13) [146]. Complementary studies on SD data have been conducted, using the correlation between the SD risetime observable (see subsection 6.1.3) and the X_{max} of a shower. This analysis using the SD benefits from a much larger statistics at the cost of larger uncertainties.

The measurements by the Pierre Auger Observatory suggest a rather light and mixed composition at lower energies and a primaries mass increase at the highest energies. By simultaneously fitting the spectrum and the X_{max} evolution above $\sim 5 \times 10^{18}$ eV, the Auger results can be interpreted assuming a simple astrophysical scenario, with sources assumed to be stationary and uniform in a co-moving volume, in which several nuclear components are injected at the sources with a power-law spectrum and with the maximal energy of the sources modeled with an exponential cut-off. The best fit supports the hypothesis of a flux suppression partly due to the reach of the source maximum energy (see Figure 4.14).

---

[1]based on 215030 events detected between the 1st of January 2004 and the 31st of August 2018 for an exposure of $60400 \pm 1810 \text{km}^2.\text{sr.yr}$
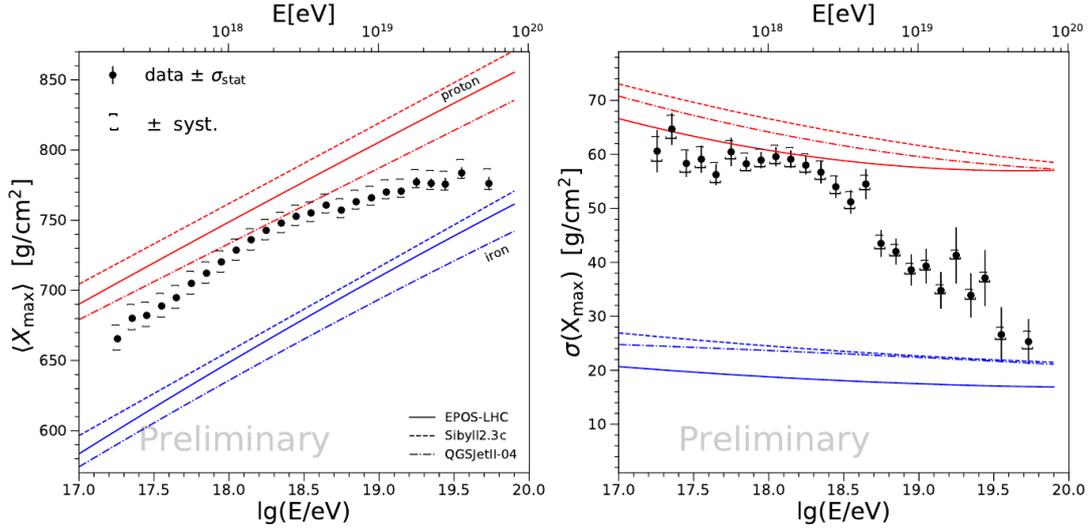
Figure 4.13: Evolution as a function of energy of the primaries mass sensitive observables, $\langle X_{max} \rangle$ (left) and $\sigma(X_{max})$ (right). The predictions from proton and iron primaries are represented by red and blue lines respectively, for hadronic models EPOS-LHC, QGSJetII-04 and Sybill 2.3c. From [147]
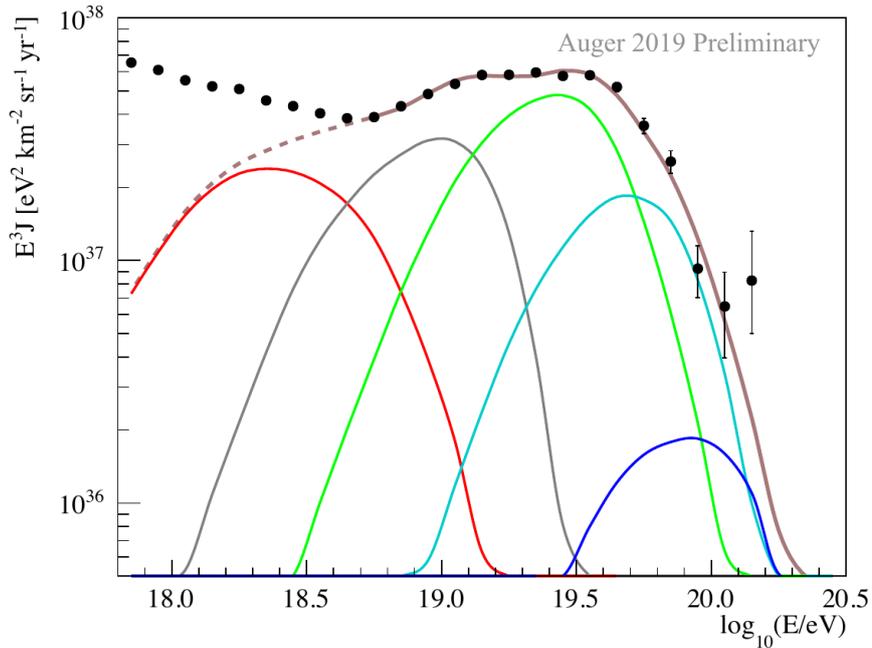
Figure 4.14: The Auger energy spectrum (multiplied by E^3) at the top of Earth's atmosphere, fitted with a simulated energy spectrum following the model used in [148]. The simulated energy spectra (colored line) is obtained by performing a combined fit of a simple astrophysical model to the energy spectrum obtained from SD data (black dot) and the values of X_{max} (which are composition-sensitive). The all-particle simulated spectrum (brown line) is the sum of the mass-restricted energy spectra: in red A=1, in grey $2 \leq A \leq 4$, in green $5 \leq A \leq 22$, in cyan $23 \leq A \leq 38$ and in blue $A \geq 39$. From [149].

4.4.3 Arrival direction

The main way to understand the origin of UHECRs is to study the arrival direction of the primaries. The distribution of UHECR arrival directions is scrutinized to get information on their origin and the possible sources. Charged cosmic rays interact with both the magnetic fields in our Galaxy and the intergalactic medium that lies between the sources and Earth, and can be strongly deflected. Therefore, it is difficult to locate their sources. However,

anisotropies in the distribution of their arrival directions may be detectable on large and intermediate angular scales, depending on the UHECR energies. The angular resolution of Auger and its coverage of about 85% of the sky, with decreasing exposure up to a declination of 42° , enable the probe of both large- and intermediate-scale anisotropies. The observation of a large-scale anisotropy in the arrival directions of UHECRs above 8 EeV was reported by the Auger Collaboration in [74] in 2017, and further studies were published in 2018 [150]. The distribution of UHECRs on the celestial sphere (see Figure 4.15) is consistent with a 7%-amplitude dipole pointing 120° away from the Galactic Center, and appears to be consistent with the large-scale distribution of galaxies and deflections within Galactic magnetic fields (GMF). This provides, for the first time, observational evidence for an extragalactic origin of UHECRs beyond the ankle.
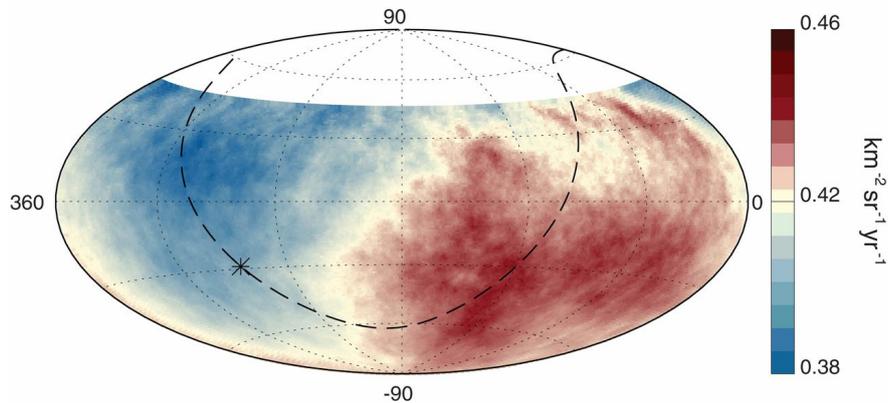
Figure 4.15: Sky map in equatorial coordinates of the UHECRs flux above 8×10^{18} eV in a Hammer projection, smoothed with a 45° top-hat function. The colormap indicates the intensity of the flux. The dashed line indicate the galactic plane and the asterisk indicates the galactic center. From [74].

Several analyses are performed to search for anisotropies at small to intermediate angular-scales, such as a model-independent blind search for overdensities, or the search for correlation with Centaurus A, the nearest radio-loud active galaxy. Correlation searches compare the location of a type of source with the arrival direction of UHECRs in bins of energy, corresponding to the expected emission spectrum of the source type. Only a few source classes could be capable of accelerating charged nuclei up to ultra-high energy (see subsection 2.2.3) and inhomogeneities in their spatial distribution may imprint anisotropy on a smaller angular scale, which can be revealed by comparing measured UHECR arrival directions with position of astrophysical objects. Given a model of candidate sources, a probability map of the arrival distribution of cosmic rays can be obtained in the form of a smoothed density map and compared to the observed distribution of events using an appropriate test statistic. With new data of UHECRs above 3.2×10^{19} eV up to the end of August 2018, the excess around Centaurus A has increased as well as the significance of rejecting the isotropic hypothesis from a comparison with a starburst galaxies model, reaching $3.9\,\sigma$ and $4.5\,\sigma$ respectively [150][151].

4.4.4 Neutral particles

The Pierre Auger Observatory has proven that aside from its main target, the nucleus primaries, it also has good sensitivity to photon and neutrino primaries. From the limits set by the Pierre Auger Observatory on UHE photon and neutrino fluxes, most top-down UHECR generation processes have already been rejected, and other astrophysical scenarios

can be explored, such as the GZK cut off (see subsection 2.3.1). Dedicated analyses have been designed to look specifically for these kinds of primaries, using the development of the showers to distinguish them from hadronic primaries or exploring channels where the hadronic background is almost non-existent.

Photon searches

The differences between photon-induced and hadron-induced showers are presented in subsection 3.2.1 and the various searches for UHE photons are described in section 6.2. The study of UHE photons for multi-messenger observations has been constrained so far by the limited distances they are able to travel without interacting (~ 4.5 Mpc at $1O^{18}$ eV). So far no observation of UHE photons have been reported by the Auger collaboration. Nonetheless, the PAO has demonstrated its background rejection power in photon search analyses (see subsection 6.2.2) and has set the most stringent limits on the photon flux at the highest energies.

Neutrino searches

Neutrinos are promising astrophysical messengers, as illustrated by the several observatories designed specifically to look for high energy neutrinos, such as IceCube [152] and Antares/KM3NET [153][154]. The astrophysical neutrinos observed by these observatories can be combined with data on UHECRs detected by the Pierre Auger Observatory. The complementarity between the observatories has been highlighted in the recent multi-messenger follow-up searches (discussed further in section 4.4.5). Since neutrinos interact only weakly with matter, neutrino-induced showers may occur very deeply in the atmosphere. One can thus search for neutrino primaries in very inclined showers since the EAS from hadronic primaries are reduced to mostly high energy muons well before reaching the SD at these high zenith angles (low background). A very inclined "young" shower (i.e. still developing when reaching the array) can only be induced by a deeply interacting neutrino. Signatures such as the time spread and the signal shapes are used to sign them with negligible hadronic background.
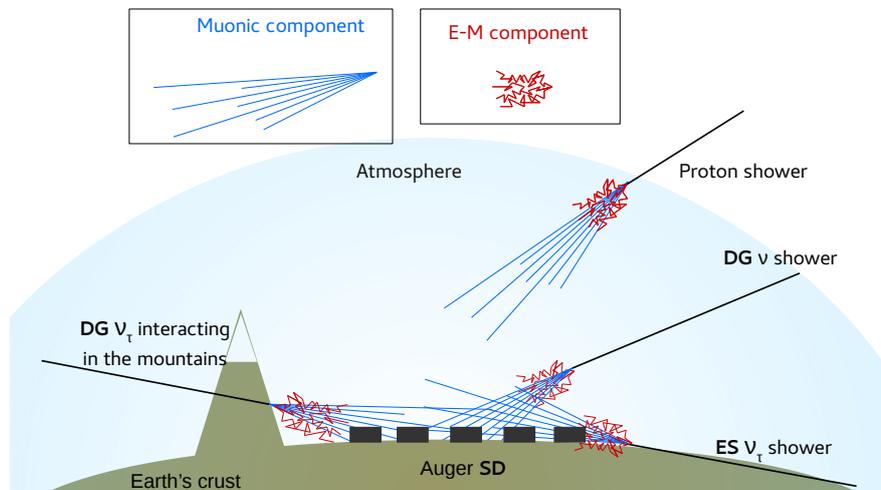
Figure 4.16:   Scheme of the neutrino channels

In Figure 4.16, the two channels[1] that are used for neutrino searches with the Pierre Auger Observatory are illustrated: the Downward Going (DG) and the Earth-Skimming (ES) channels [155]. The Downward Going channel includes showers that enter the atmosphere at very high zenith angles ($60° <$ Zenith $< 90°$), induced by any neutrino flavors. The Earth-Skimming channel is used to look for showers initiated by tau neutrinos (ν_τ) going through the Earth's crust. Some ν_τ can interact close to the surface after skimming through the ground, producing a tau lepton that will start an upward-going shower above the surface array. The ES channel is looking for showers with zenith angles between $90°$ and $95°$. No UHE neutrinos have been identified by the Pierre Auger Observatory, although it has a very good sensitivity in the two channels [155][156]. This has allowed the collaboration to derive limits on the diffuse flux of UHE neutrinos, comparable to the one achieved by the IceCube experiment (see Figure 4.17). Also, the Pierre Auger collaboration has conducted searches for point-like sources of UHE neutrinos [150].
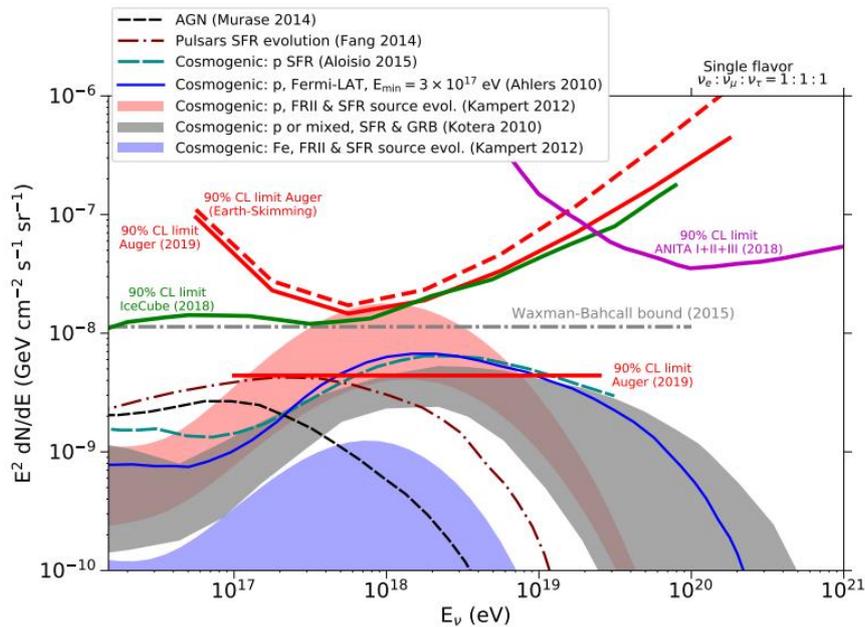
Figure 4.17: Upper limit to the normalization of the diffuse flux of UHE neutrino. The upper limits obtained by the Pierre Auger Observatory are compared to those obtained by IceCube and ANITA, as well as predictions from cosmogenic and astrophysical neutrino production models. From [156].

4.4.5 Multi-Messenger

Due to its good sensitivities to UHE photons and neutrinos, the Pierre Auger Observatory can play a major role in the multi-messenger astrophysics field. By combining its data with other observatories, the Pierre Auger Observatory can help pinpoint the location of potential sources, put constraints on the highest energies counterparts and broaden the energy spectrum of the observations. Prime examples of that are the large multi-messenger observations campaigns, following-up gravitational wave events. Given the expected distance of these events from Earth (>10 Mpc), the search for follow-up signals in the Pierre Auger Observatory data was focused on neutrino searches. More specifically, when the observatory receives an alert from another observatory about an ongoing event, a search is conducted in physically motivated time windows, around the time of the event and after it.

---

[1]a channel is a range of zenith angle,between the shower axis and the ground

The PAO has participated to the follow-up searches around the binary black holes (BBHs) mergers GW150914 and GW151226 as well as around the binary neutron star (BNS) merger GW170817 (see Figure 4.18) detections by the LIGO/Virgo collaboration.

No UHE neutrinos were found in coincidence with these events. Nonetheless, the Auger collaboration was able to set limits on the UHE neutrinos flux as well as energy dissipated as neutrinos of these extreme events.
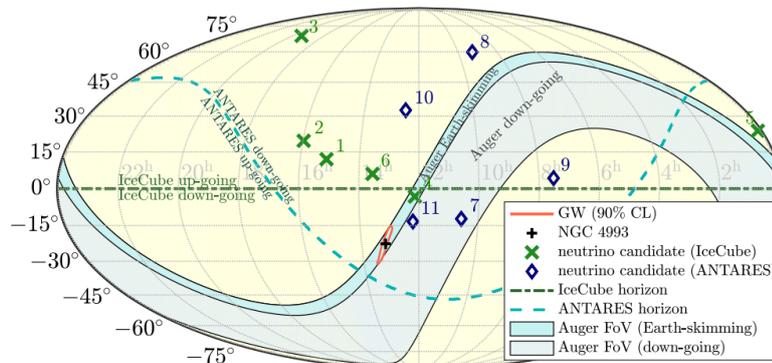
Figure 4.18: Skymap in equatorial coordinates during the GW170817 event. The expected localization of the BNS event and its host galaxy are shown, as well as the neutrino candidates within $\pm 500\,\text{s}$ of the merger and the horizons of the ANTARES and IceCube experiments. The Pierre Auger Observatory fields of view in the ES and DG channels are also depicted. [157] and references therein.

AugerPrime

The Pierre Auger Observatory has greatly improved our knowledge of the ultra-energetic cosmic rays, but leading to a complex and unexpected picture of UHECR, challenging the original expectation of few sources and light primaries. The existence of a flux suppression above $5 \times 10^{19}\,\text{eV}$ is incontestable, being due to propagation effect or to the exhaustion of the sources. With no UHE photon or neutrino found in data, limits on their fluxes are derived, clearly disfavoring "top-down" scenarios and leading to an astrophysical origin of UHECR, but no clear clue on any existing UHECR source has been revealed so far, even if evidence of anisotropies in the UHECR arrival directions emerge from several studies. The image of a light and mixed composition around the ankle and a heavier composition at the highest energies is strengthened by different composition studies. Unfortunately, the X_{max} measurements above $4 \times 10^{19}\,\text{eV}$ with the FD are limited by its low duty cycle. The detailed interpretation of the data is currently limited by uncertainties in the hadronic interaction models, for which hints of deficiencies have been observed. The astrophysical scenario resulting from Auger measurements is very complex and cannot at present be understood in terms of a unique interpretation for the sources, propagation and nature of the UHECRs. Information on the mass composition is crucial to go further in the interpretation of the observations, in particular at the highest energies. These questions are going to be addressed by the AugerPrime project [158], the Observatory upgrade program, which is described in chapter 5.1.

Chapter 5

Contribution to the AugerPrime upgrade

Contents

In this chapter, we will first describe the AugerPrime upgrade, the physics motivations of the upgrade, how these can be achieved and the implementations selected. Then, a focus will be made on one aspect of the upgrade, the Scintillator Surface Detectors (SSD). We will dive in the physics of the detectors, how the signal is produced and collected, before describing the detectors design that has been selected for the upgrade and how to go from SSD signal to muon/EM disentanglement. Finally, the assembly process and tests of the SSDs that occurred at the LPSC are detailed.

5.1 The AugerPrime upgrade

In 2015[1], the Pierre Auger collaboration proposed an ambitious upgrade of the Pierre Auger Observatory, that is underway since then. The aim of this upgrade is to push further the boundaries of our knowledge on UHECRs (see 2). To achieve these scientific goals, the EAS reconstruction has to become even more precise, by adding new composition-sensitive observables, to constrain the shower development models. New scintillation detectors are being installed on top of the existing WCD stations to study the EAS with new instruments. The WCD will remain mostly unchanged through the upgrade (apart from the new electronics). It will serve as a continuous reference to assess the performances of the upgraded observatory.

5.1.1 Motivations of the upgrade

To find answers to the open scientific questions described previously, the Pierre Auger observatory is undergoing a significant upgrade: the *AugerPrime* upgrade. The *AugerPrime* project has multiple physics motivations:

- elucidate the mass composition and the origin of the flux suppression at the highest energies, by finding out if it is due to a phenomenon during the CRs propagation like the GZK cut-off [8][9] or if it corresponds to a hard limit on the sources acceleration power [159]. Each effect (or a combination of both) could explain the strong suppression in the spectrum above $E>10^{19.5}$eV . Understanding the phenomenon behind the cut-off, would be very useful to put constraints on the potential sources and to have indications on the expected fluxes of UHE neutrinos and photons.

- an extension of the mass composition topic is: What is the proportion of light and heavy nuclei in the CR flux and especially, are there protons at the highest energies? Knowing the ratio of protons in the cosmic rays flux at the highest energies is crucial to estimate the science potential of proton astronomy [160] and also to predict the fluxes of secondary photons and neutrinos.

- gain new insight into hadronic interactions at the ultra high energies. So far, the models used to simulate the development of EAS initiated by ultra high energy cosmic rays rely on extrapolation [107] from the LHC results (EAS are created in very forward processes and the center of mass energies for UHECRs are around hundreds TeV compared to the 14 TeV reached for protons collisions at the LHC). Probing these extreme energies directly is crucial for our field of physics.

The key to reach all these scientific goals is to improve the event-by-event identification of the primaries. These scientific goals are the directions that are followed to fulfill the broader mission undertaken by the Pierre Auger observatory, to understand the nature and origin of the UHECRs.

5.1.2 Means to achieve the scientific goals

The main focus of the upgrade, to achieve better performances out of the observatory in terms of composition-sensitive observables, is to measure the muonic and electromagnetic components on a shower-by-shower basis. The muonic component of the shower is a key observable for primaries discrimination [161]. The number of muons given a primary's energy

---

[1]The preliminary design report for the AugerPrime upgrade was released in 2015

is different depending on the mass of the CR (see 3.1.2). The way a shower *ages* also strongly depends on the nature of the interacting primary.

Presently, the muonic component of the showers can be estimated statistically on horizontal showers[1] or on averaged groups of showers through indirect measurements that lack precision and require direct detection for validation. Muons going through the WCDs create characteristic peaks in the signal distribution that can be identified when they aren't drowned in the flux of other secondary particles. Very similar peaks can also be induced by secondary high energy ($> 100\,\mathrm{MeV}$) photons, which prevent an unambiguous identification of muons.

But these methods cannot describe the full picture of the muonic component of the showers on an event-by-event basis. Additionally, all the UHECRs observatories find a discrepancy between the expected muonic component from the hadronic UHE models, and what can be inferred from data [118]. Getting a measurement of the muonic component would provide a model-assumption free observable and a way to challenge the existing UHE hadronic models (and potentially help select the best one).

Getting more composition sensitive observables in general, especially in the flux suppression region, would not only improve the primaries identification capabilities of the observatory, but also the global reconstruction performances and lower the statistical uncertainties due to hadronic models and the reconstruction algorithms. The main observable used for mass composition analysis, X_{max}, is not available for all the data collected by the observatory, since the FD has a $\sim 15\%$ duty-cycle (and even further reduced by strict cuts to ensure data quality), as mentioned in 4.2.3. Adding new mass-composition sensitive observables to the SD, that has an almost 100% duty-cycle, is thus most efficient.

5.1.3 Description of the upgrade

The new systems brought by the AugerPrime upgrade are:

- a Scintillator Surface Detector (SSD) made of modules installed on the Water Cherenkov Detectors. The scintillators provide a different response to the electro-magnetic and muonic components of the showers compared to the WCDs. The two detections can thus be combined to extract the showers electromagnetic and muonic components.

- the Auger Muons and Infill for the Ground Array (AMIGA) project, where Undergound Muon Detectors (UMD), made of plastic scintillators, are buried near the WCDs in a small part of the array of about 23.5 km² [162]. The aim of the AMIGA project is to study showers at lower energies (down to $\sim 10^{16.5}$ eV) by creating two denser array, with WCDs deployed in a 750m (SD-750) and a 433m (SD-433) triangular grids. The SD-750 WCDs are already fully deployed, and an AMIGA Engineering Array, consisting of 7 WCDs and 9 UMD [2] in an hexagonal cell (Unitary Cell), has been operated by the collaboration as proof of concept from March 2015 to November 2017. Using data from the Unitary Cell, the UMD design has been optimized, they have been calibrated and the reconstruction of muons by a single UMD as well as an event reconstruction by the array have all been demonstrated [163]. The AMIGA project is now in production phase. When completed, it will give a direct measurement of the number of muons inside showers at ground level. Although only the denser arrays will be equipped with these buried scintillators, this fraction of showers with a UMD measurement will be used to validate/fine-tune, the muonic component extracted with the WCD/SSD combination.

---

[1]highly inclined showers have to go through more atmosphere before hitting the array, their electromagnetic component thinned out in the process

[2]two WCDs are equipped with two UMDs each

- new electronics [164] and small PMTs [165] (sPMT) inside the WCDs. The new electronics will support the use of the new detectors installed on the WCDs, as well as support a new sPMT installed inside the WCD and increase the FADC sampling rate [166]. The signal of the WCDs and the SSDs will be sampled synchronously at a 120 MHz rate, up from 40 MHz, while the timing accuracy will also be improved down to 5 ns. The combination of the sPMTs with the new electronics will increase the dynamic range of the WCDs by a factor of 32, collecting the complete signals closer to the shower core, where regular PMTs saturate. The increased capacities of the electronics could also potentially lead to new triggers and the increased sampling rate will slightly improve the angular resolution of the surface detector by improving the temporal resolution.

- radio antennas[1]. These antennas will be mounted on top of the WCDs and will detect the radio signals emitted by EAS as they propagate through the atmosphere. The design selected for the radio antennas allows a detection of the radio signals in the frequency range 30 to 80 MHz [167]. Horizontal showers in particular can produce very large radio emission footprints, up the 100 km^2, which means that the multiplicity (the number of stations recording a signal) for each event in radio will be high and allows for radio detection using sparse arrays [168]. The radio detection has been studied using the AERA detector, to optimize the antennas design and start working toward a radio reconstruction of inclined showers [169]. The radio antennas will be complementary to the SSDs, since they measure the inclined showers while the SSDs measure the vertical showers.

5.2 Physics of the Scintillator Surface Detectors

To disentangle the several types of particles making up the showers at ground-level, a signal from a detector with a different response to these components compared to the WCDs is required. But these signals must be taken at the same spatial and temporal position. These detectors must also be reliable, easy to install in the field and require low (hopefully none) maintenance.

5.2.1 Signal collection

Plastic scintillators have been used to detect air showers by many experiments [170][171][172] since they provide a cheap and reliable solution to detect secondary particles as well as being very effective and having a linear response. As charged particles go through the plastic scintillators, they excite the molecules of the medium which then emit photons to go back to their fundamental state.

The photons produced by the through-going particles must then be transported, as efficiently as possible, to a collector while preventing them from interacting with the medium. A scintillation module consists of scintillator elements, where the emitted photons are centralized into a single collector. Once the light produced by the scintillators is collected into a PMT, the module's signal can be directly compared to the WCD signal.

For WCDs, the aim of the calibration is to set the relation between the signal recorded by the PMTs and a VEM (see 4.2.1) The SSDs will be calibrated in a similar way, but by measuring the signal recorded by a detector when a minimum ionizing particle (MIP) goes through it. A scintillation detector is thin so a MIP signal could be lost in the low energy

---

[1]the radio part was not inside the initial Design Report of the upgrade but was added after.

background, but since the SSDs will be located right above the WCDs, the tanks can be used to cross-trigger and remove this background. While a VEM, the calibration unit for the WCD, is around 94 photon-electron (PE) [133], the baseline design selected for the SSD is around 12 pe per MIP, with an estimated degradation to around 8 PE after 10 years of operations in the field [158]. This means that the signal is much closer to the background in the SSD readout, but through the crosscheck with the WCD, the signal can be tested against very low background.

5.2.2 Technical description

The solution selected for the AugerPrime upgrade is to install a scintillation detector on top of the existing WCDs. A scintillation detector is composed of three *active* elements:

- two scintillator modules. Each of them produces light when secondary particles pass through its sensitive area.

- wavelength shifting optical fibers. They guide the light produced by the scintillators to the photomultiplier, and shift its wavelength into one that is more efficient for PMT collection.

- a photomultiplier. It converts the photons into an electric signal.

The above active elements are enclosed in an aluminium protective and insulating casing.

The two scintillator modules making up each SSD are made of 24 scintillator bars each, aligned symmetrically. They have dimensions of 160cm x 5cm x 1cm and are produced on the FNAL/NICADD extrusion line [173]. The FNAL scintillators are co-extruded with an outer-layer of polystyrene mixed with a TiO_2 pigment. This outer-layer provides some protection against external factors like mechanical stress on the active inner part of the bars and also increases the amount of photons reaching the fibers thanks to its diffusive reflective behaviour. Each bar has two extruded channels for the wavelength shifting optical fibers.

The plastic fibers are Kuraray Y11(300)M S-type fibers [174], of 1 mm diameter. This model was specifically chosen to match the spectrum of the light emitted by the scintillator bars in absorption (absorption peak at 430nm) and to shift it from blue to green (emission peak at 476nm). The emission spectrum of the scintillators, as well as the absorption and emission spectrum of the WLS fibers are shown in Figure 5.1. The wavelength shifting avoids most of the light absorption along the fibers. The fibers are of S-type because these types allow more bending at the cost of a shorter attenuation length (> 3.5m according to the constructor specifications).

Each fiber traverses along two scintillator bars through channels separated by 10cm and they are cut to all have the same length. The two ends of a fiber are then guided to the central part of the module by styrofoam routers. Other routers are also used to guide the fibers going in and out of the scintillators in a "U"-shape, with an appropriate curvature radius, so as not to put too much strain on the fibers and maximize light-yield. Reading the fibers on both ends provides a better light collection uniformity.

At the center of the module, all fibers ends are bundled together and glued inside a *cookie* (see Figure 5.2). The *cookie* is a polymethyl methacrylate (PMMA) cylinder, that is used to ensure the optical connection with the PMT. It is fixed to the frame in the middle of the module long side inside an aluminum tube.

The PMTs selected are Hamamatsu R9420 PMTs (1.5" diameter) [175]. This model has a quantum efficiency [1] around 10% (see Figure 5.3) in the wavelength range of interest (between

---

[1]The quantum efficiency is an estimator of the cathode photoemission, it is defined as the ratio between
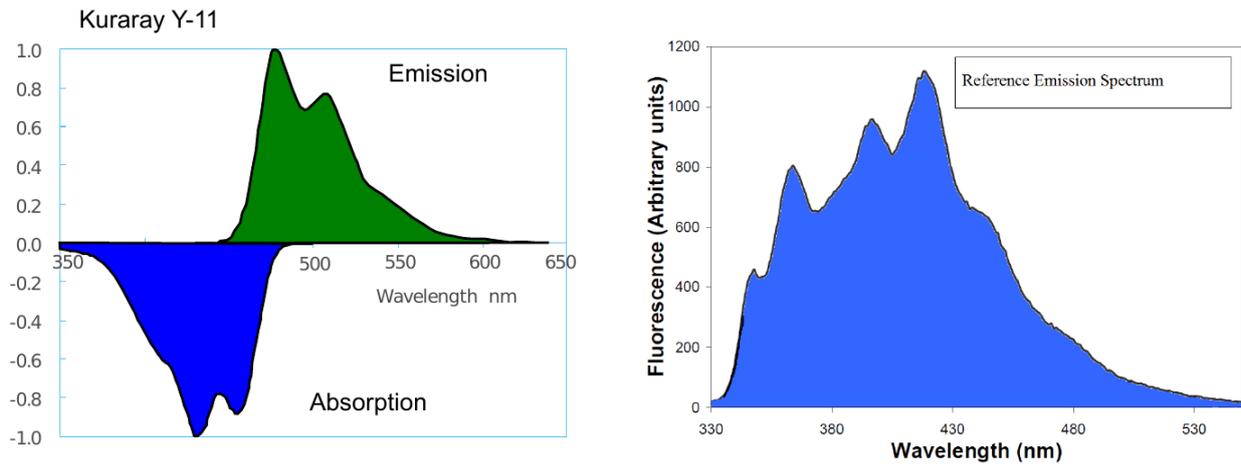
Figure 5.1: Left: Absorption and emission spectrum of the Kurasay Y11(300)M S-type plastic fibers. Right: Reference emission spectrum of the FNAL-NICADD extruded scintillators.

Figure 5.2: Left: Picture of a finished cookie, with the plastic fibers (in green) in place. Right: Picture of the same cookie, taken from above, with a mirror placed below to control the quality of the fibers alignment and gluing.

400 and 600 nm, green region) and excellent linearity range of up to 200 mA of peak anode current for an operating gain of 7.10^5. They are installed after the SSD is mounted on a WCD to prevent them from being damaged during shipping.

All these active elements are encased in an aluminum box of dimensions (3.8 m x 1.3 m). A composite panel is glued to the aluminum frame to serve as a bottom, strengthen the structure and support the inner elements. This composite panel is made of extruded polystyrene foam (XPS), between two sheets of aluminum. The XPS was selected because it is waterproof, sufficiently resistant to mechanical pressure and durable to last 10 years in the field. The aluminum sheets are added to guarantee the robustness of the panels. Once all the active elements are correctly assembled, an aluminum sheet is glued and riveted on top of the casing to close it. An air-permeable light-tight venting hole through the frame allows pressure equilibrium. The detector is designed to be light-tight and weather-proof.

Additionally, a roof of corrugated metal sheet will be installed in the pampa to shield the module from the sun while preserving air circulation.

5.2.3 From SSD signal to EM/muon disentanglement

A WCD and a SSD will produce a different signal for the same particles going through them. By combining the two detections at the same location, one has access to two independent

---

the number of photoelectrons emitted by the cathode to the number of photons reaching the PMT's window [176].
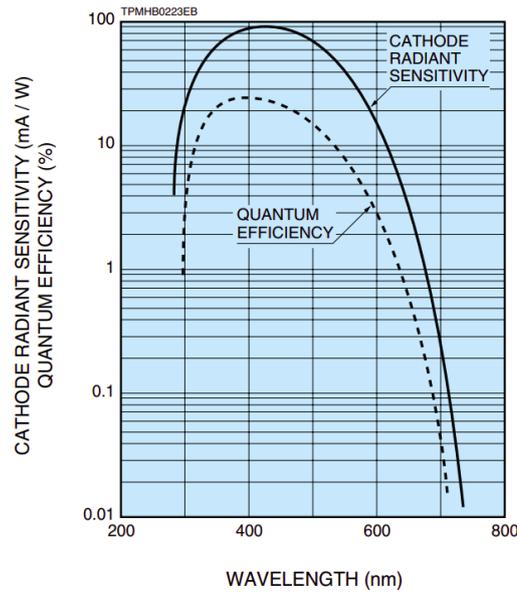
Figure 5.3: Quantum efficiency of the Hamamatsu R9420 PMTs as a function of the wavelength according to the specifications provided by the constructor.
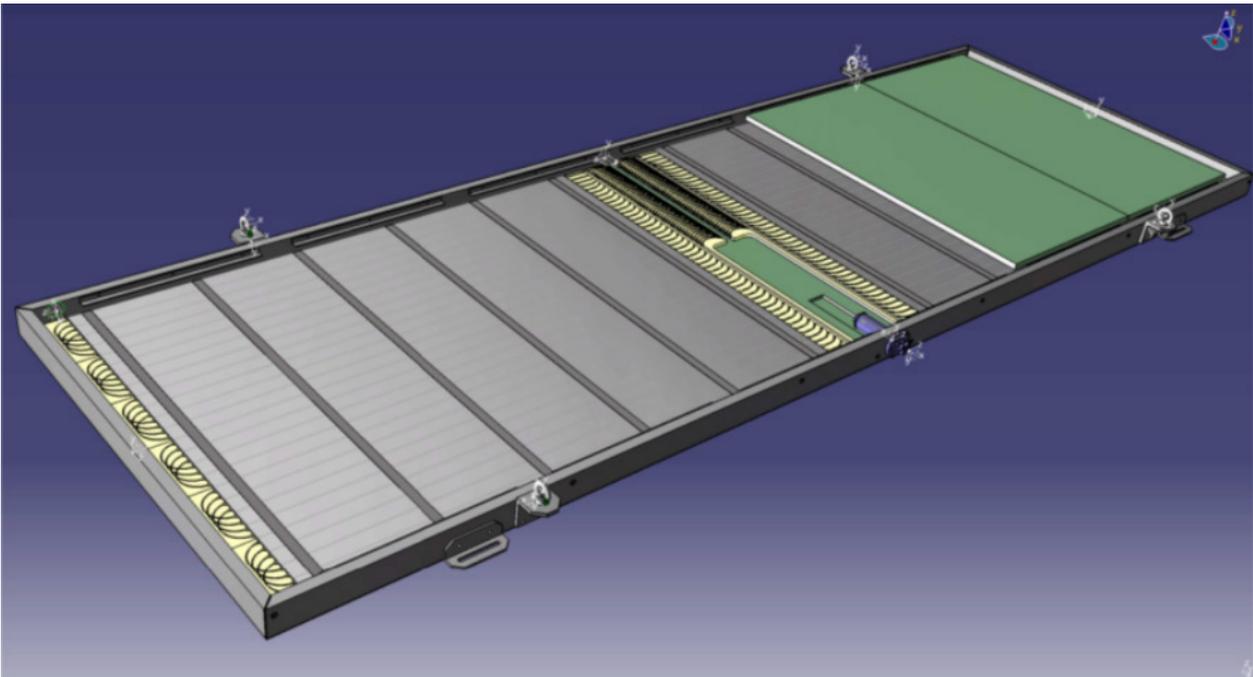
Figure 5.4: Scheme of an "open" scintillator surface detector. Two symmetric collections of 24 scintillator bars each make up the active part of a module. Two WLS fibers go through each bar of plastic scintillator. Outside of the bars, the fibers are guided by polystyrene routers (in yellow). One fiber goes through two bars and all fibers ends are guided back to a PMT in the center. The PMTs are installed inside an aluminum tube specifically meant to receive them inside each PMT. Finally slabs of extruded polystyrene (in green) fill-up the empty space left in the aluminum casing.

measurements of the secondary particles from an EAS. With these two measurements, one should thus be able to separate the parts of the signals due to the electro-magnetic component of the shower and the parts due to the muonic component of the shower.

The two main approaches to disentangle the shower components are the universality approach and the matrix inversion approach.

The matrix inversion approach uses equation 5.1 to relate the signals from the two detectors directly to the fractions of electromagnetic and muonic particles using information on the detectors geometry and efficiency compared to one-another [158].

$$\begin{pmatrix} S_{SSD} \\ S_{WCD} \end{pmatrix} = \begin{pmatrix} \lambda.A_{SSD} & A_{SSD} \\ \beta.A_{WCD} & A_{WCD} \end{pmatrix} \begin{pmatrix} F_{em} \\ F_{\mu} \end{pmatrix} \tag{5.1}$$

where S_{SSD} and S_{WCD} are the signals of the two detectors in MIP and VEM respectively, A_{SSD} and A_{WCD} their horizontal areas in m^2 and, F_{em} and F_{μ}, are the electromagnetic and muonic fluxes in VEM.m^{-2}. λ is the average energy deposited in the scintillator per VEM of electromagnetic flux in $MIP.VEM^{-1}.\cos(\theta)^{-1}$. β is the ratio between the WCD surface perpendicular to the zenith angle and its horizontal surface ($\theta=0$).

With this relation, and using the signals from both the WCD and the SSD of a station, it becomes possible to calculate the EM and muonic fluences:

$$F_{em} = \frac{1}{\lambda - \beta} \left(\frac{S_{SSD}}{A_{SSD}} - \frac{S_{WCD}}{A_{WCD}} \right) \tag{5.2}$$

$$F_{\mu} = \frac{1}{\lambda - \beta} \left(\lambda \frac{S_{SSD}}{A_{SSD}} - \beta \frac{S_{WCD}}{A_{WCD}} \right) \tag{5.3}$$

Then from these fluences, one can get the muonic part of the WCD signal:

$$S_{WCD,\mu} = A_{WCD}.F_{\mu} \tag{5.4}$$

The universality approach uses the fact that the secondary particles making up extensive air showers behave similarly for air showers at the same stage of their development. EAS can thus be described using only a few parameters[1]: E and X_{max} for electro-magnetic showers and it can be extended to hadronic showers by adding the relative muon fraction of the shower R_{μ} The universality **method** is the application of the universality **approach** to a specific detector. It aims at building a model that predicts the signal expected in the WCD from a given shower [177].

By using the shower parameters, the individual stations characteristics (distance, total signal,...) and the traces shape, the universality method is able to provide an estimation of each component contribution to the signal. The four components used by the universality method are: 1. muons (μ), 2. e$^{\pm}$ and γ from high energy pions (eγ) 3. e.m. particles from the muon decays (e$\gamma(\mu)$) 4. e.m. particles from low energy hadrons (eγ(had)) and their contribution are calculated as:

$$S_i = S_i(r,\Delta X,E,\theta,\psi) \qquad \text{with i one of } \mu,\ e\gamma,\ e\gamma(\mu),\ e\gamma(had) \tag{5.5}$$

where: r is the distance of the station from the shower core, ΔX is the distance to X_{max} in station interval unit, E is the primary energy, θ is the zenith angle and ψ the azimuth.

These individual component contribution are then combined into a signal S_{tot}:

$$S_{tot} = S_{e\gamma} + R_{\mu}(S_{\mu} + S_{e\gamma(\mu)}) + R_{\mu}^{\gamma}(S_{e\gamma(had)}) \tag{5.6}$$

where γ is a factor that depends on the distance to the core.

Then once this model for the signal is built, it can be applied to data, fit the measured signals and reconstruct macroscopic parameters of the showers. The same method can be

---

[1]the main point is that secondary particles in an EAS show *Universality*, meaning that their behaviour are almost independent of the primaries properties or the hadronic models used
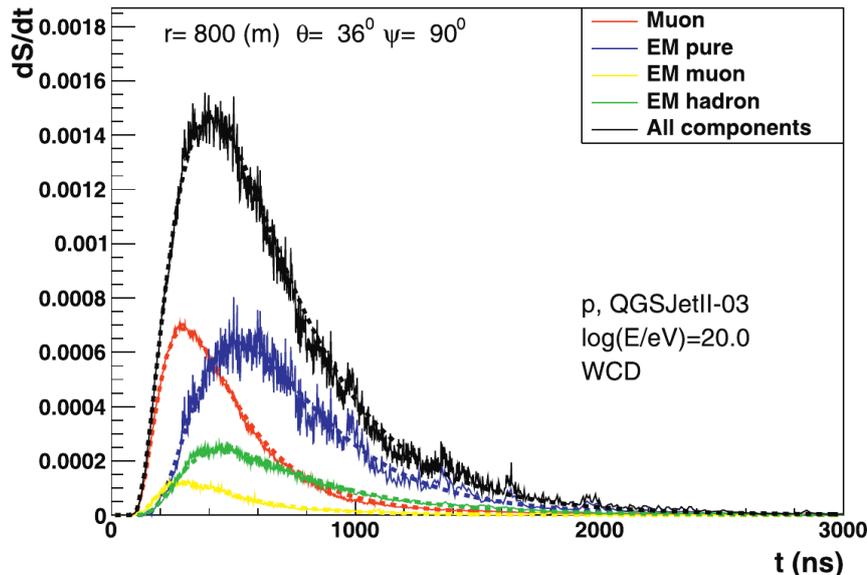
Figure 5.5: Averaged time dependent signals from proton showers, simulated with CORSIKA using QGSJetII-03, at an energy E=10^{20}eV with a zenith angle θ=36° in a WCD station located at 800m from the shower core at an azimuth angle of ϕ=90°. The signals from the, muonic component, electo-magnetic total, muonic and hadronic components and the total signal are represented in red, blue, yellow, green and black solid lines respectively. In dashed lines are the predictions for the components using the Universality method. Taken from [177]

applied to signals from the SSDs and this new information will be added to the data/model fit.

The two methods are complementary. Close to the shower core, the inversion matrix method allows for an unbiased disentanglement without relying on hadronic model simulations. However, the resolution on this method evolves as $1/\sqrt{s}$ where s is the total signal. So as the signals get weaker (stations farther from the core), the universality method becomes a more appropriate choice. It is important to note however that for the universality method to work properly, it needs a high enough density of particles, so it has to be applied on stations with signals above $5\,VEM$ and not too far from the core (good accuracy, around 5 to 10 %, from 100 to 2000 m from the shower core).

The underground muons detectors installed in a sub-portion of the array will allow a cross-check of these muonic component estimations with a direct measurement.

5.3 Construction and characterisation of the Scintillator Surface Detectors

The construction of the SSDs has been divided between 6 institutes members of the Pierre Auger collaboration: RWTH Aachen University (Aachen), LPSC (Grenoble), Karlsruhe Institute of Technology (Karlsruhe), Institute of Nuclear Physics PAN (Krakow), INFN (Lecce) and NIKHEF (Nijmegen). The PMTs are tested, controlled and prepared in two more institutions: Bergische Universität Wuppertal and INFN Napoli.

Each institution had to set up its own assembly area and some specific tools were built in the institutes. In the following section, we will only present the assembly process followed by the LPSC team but all institutions followed similar procedures [178].

Aside from the assembly, each institute was also responsible for the test and characterization

of the modules produced. It is very important to make sure that each SSD is light-tight, as otherwise it would make the detector unreliable in the field, crippled by false signals. The response of the detectors to cosmic muons has also been studied by all the institutes to control the SSDs assembly through time and to keep track of their individual performance.

5.3.1 Scintillator Surface Detectors Assembly

Preparation towards the assembly process: The LPSC had the responsibility of assembling and testing 90 SSDs[1]. The necessary components were bought by different institutes and dispatched to the assembly sites. The LPSC bought ∼100 composite panels; all other components were received. Aside from the detectors components, the consumables used during the production (glue, propanol, gloves, ...) were needed and specific tools were designed. The general organization of the assembly chain was defined before hand: the assembly steps, the number of people necessary for each step, where each step can be taken, where to store the components and the assembled modules.

Assembly hall: The assembly process was performed in a hall, protected from natural sunlight to prevent the WLS fibers from any damages. This hall was chosen first because of the space available, it allowed us to set up three assembly tables (one being a tilting table) as well as a testing area next to each other, and second because of an existing mechanical bridge that allowed us to move the (partly-)assembled modules between the working tables.

Two other areas were used during the construction process: a tent and pallet boards outside the hall to store the bulky materials (frame, composite panels) and assembled modules, and another hall, to prepare the composite panels.
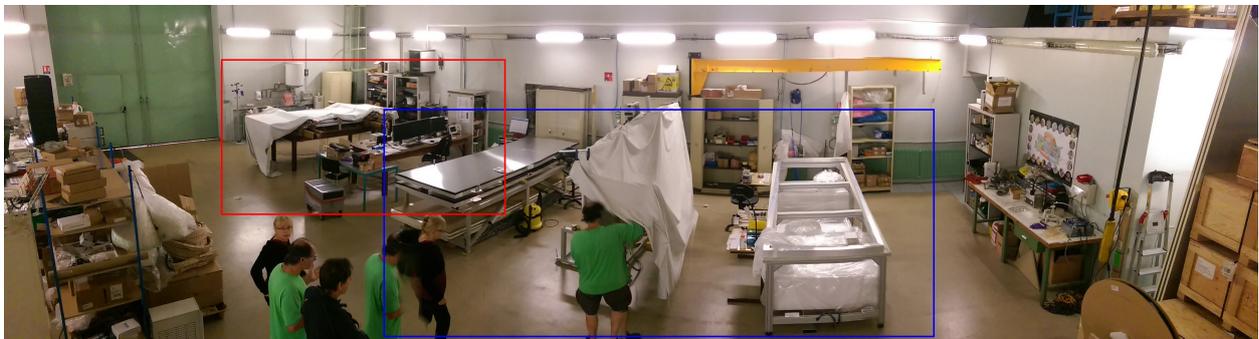
Figure 5.6: Layout of the assembly hall. In the blue rectangle, are the three tables on which the aluminum frames can be put to assemble the detectors. Starting from the right, the second table was used to insert the fibers and make the cookie since it could be tilted via a pressurized air system. In the red rectangle is the testing area, with the test bench at the leftmost table and the table with the computer, DAQ and HV supply beside it.

Preparation of the mechanical structure: Before the assembly the composite panels were prepared by digging a ridge in the center of each one to accommodate the PM tube. The aluminum frames are the skeletons of the SSDs. First they are placed on an *elcom* table, a supporting structure made of aluminum profiles. The frames are then drilled to later rivet the attachment brackets used to attach the SSDs to the supports on the WCDs. Three additionnal holes were drilled for the sintered metal, a slice of porous metal, used to allow pressure equilibrium inside the SSDs while keeping them light- and water-tight. In Grenoble,

---

[1]the number of SSDs each institute was in charge of wasn't the same

the frames received were welded by a company before reception, this wasn't the case in Aachen, Karlsruhe, Lecce and Nijmegen, where they had to assemble the four side bars together (with rivets, corner piece and glue for tightness).

Once the holes are drilled, the frame is vacuumed and cleaned with isopropanol and the protective plastic film on the undug face of the composite panel is removed. The panel is then glued to the bottom of the frame using Ottocoll S610 sealant, a special glue that ensures light-tightness. At every assembly step involving gluing components, said components are always thoroughly cleaned with isopropanol.

Active components Once the frame is prepared, the components are installed inside it: sintered metal and PM Tube are glued and riveted while the scintillator bars and polystyrene routers are taped to the composite panel symmetrically on both sides of module. The WLS fibers are inserted inside the channels drilled into the bars.

Figure 5.7: Picture of the installation of a scintillator bar inside the aluminum frame. Beforehand, double-sided adhesive stripes are laid out on top of the composite panels two on each side of the module. The protection on the other side of the adhesive stripes are removed as the scintillator bars are installed in the frames.

Inserting the fibers into the scintillator bars was done on the second table of the hall, the tilting table. All the fibers ends are routed to the middle of the module, in line with the PM Tube, where they are bundled together. Once it is drawn through the scintillator bars, each fiber is cut at roughly the same length: $\sim 5.8\,\mathrm{m}$. The length selected for each fiber had to be a balance between cost-efficiency and practicality, to not run-out of fibers by using too much, while also being able to redo the surface smoothening of the tips in case a mistake was made. Once the fibers are bundled together, the table is tilted, and the fiber bundle is brought near a custom-made "melting station". When the fibers are cut it creates inhomogeneities on the surface, leading to an uneven light transmission out of the fibers. To prevent that, the fibers ends are cut once again, just a few millimeters, with a more precise tool to have flatter fibers tips. The cross-section is then slightly melted, just enough to get an even surface without spreading too much molten matter outward.

The *cookie* is the optical interface between the fibers and the SSD's PMT. It is a special connector made to ensure that the PMT receives the integral signal from all the optical fibers. The fibers ends are bundled into the cookie and an optically-clear two component epoxy cement is poured into the vessel to fix the fibers position with respect to the window towards the PMT. The ideal case is to have all the fibers ends tangent to the cookie window and in
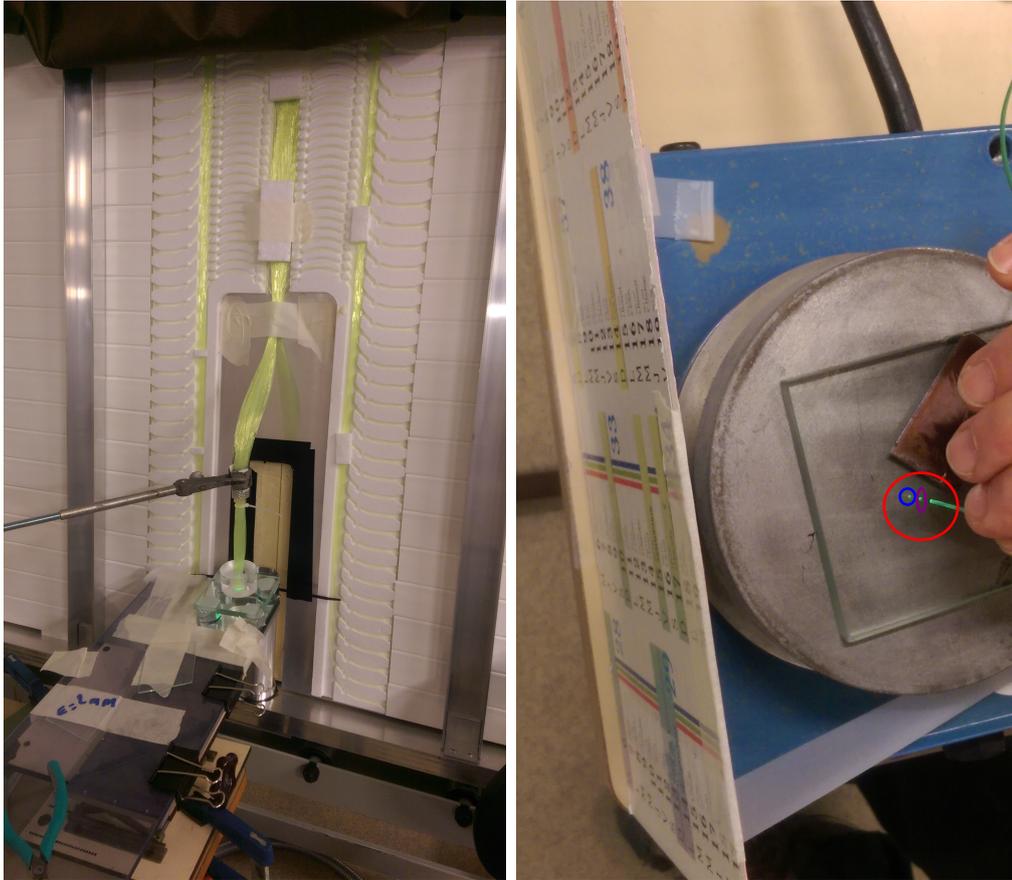
Figure 5.8: Left: Picture of the installation used to cut the fibers ends before melting the surface and then to align them before the making the cookie. There are 24 fibers in 2 modules for each SSD, summing up to 96 fibers ends (2 ends per fiber). Once all the fibers are correctly drawn inside the scintillator bars and inserted in the routers, they are fixed in place with polystyrene pieces and adhesive tape, and the fibers ends are bundled above a working station for cutting and melting. Right: Picture of a fiber's melting. A fiber's end is put in contact with the hotplate for a few second to guarantee a smooth surface. The reflection of the fibers tips, in the glass above the hotplate is used by the operator to control the state of the fiber surface. It is further checked using a magnifying glass.

contact with it, before gluing. The cookie is then carefully set back in order to insure a 2 mm gap between the fibers end and the window. This allows a better flow of the optical cement and avoids trapping air bubbles, leading to a better optical coupling. An example of a glued cookie can be seen in Figure 5.2. In practice, it turned out to be very tedious to align all the fibers in contact with the cookie's window.

With the fibers aligned in the cookie and fixed with the optical cement, the most critical step has been taken. The module is then closed while ensuring the light tightness of the detector and it can be moved to the testing area.

Quality tracking   At the Grenoble site, several operators were involved in the assembly process. To keep a continuity and traceability through the assembly of the 90 SSDs it was thus necessary to setup a detailed assembly procedure. To this end, I participated in the construction of the first few detectors while setting up an illustrated step-by-step notice on how to assemble the SSDs, from the base components up to the closed modules. New operators, some from other labs but also interns, greatly contributed to the assembly and with minimal training time thanks to the written procedure. All the pictures of this subsection 5.3.1 were taken for this assembly procedure.

Also, during the assembly of the 90 SSDs, the operators were asked to provide and fill different forms to keep track of actions taken on any detector as well as useful information: defects, scintillator number, fiber roll number,... Some of this information is then uploaded to a database shared by all the construction sites of the collaboration.
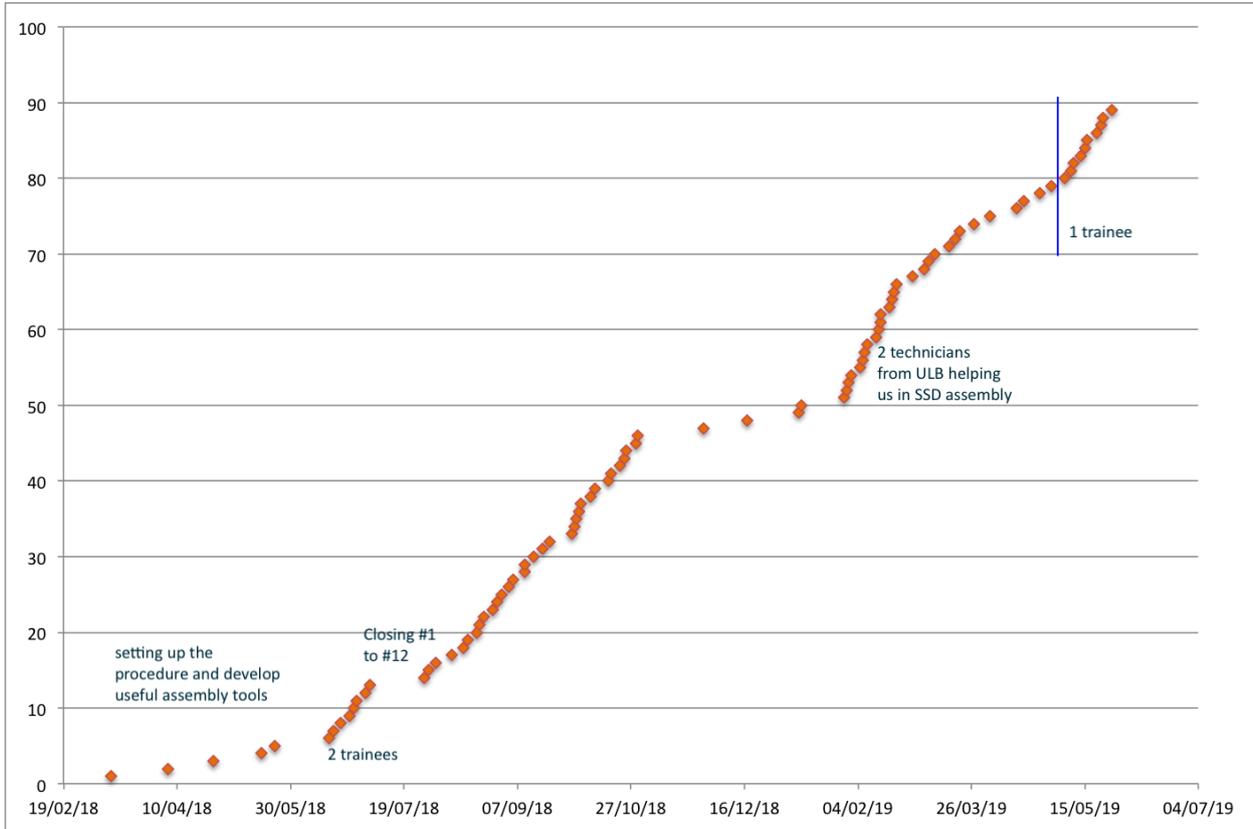
Figure 5.9: Evolution through time of the number of SSDs produced at the LPSC. The start of the assembly process was slower since the detailed procedure had to be set and because new tools were developed to accelerate the assembly. The arrival of new operators can clearly be seen in the sharp increases in production rate. The production and tests of the 90 SSD modules was finished in June 2019. All the 90 modules have been shipped to Argentina and are currently being installed on site.

5.3.2 Testing the SSDs

Introduction and mechanical tests As already mentioned, each construction site had the responsibility not only of assembling the detectors, but also of characterising the detectors response and testing for any malfunctions. All the SSDs to be installed in the argentinian pampa must be coherent as a whole detector and each individual detector must have a consistent response to a particle going through it: there needs to be an uniformity of the detector's response between individual modules and whithin a single module's signal. To be sure that the constructed SSDs comply with these conditions, a series of thorough checks have to be performed on each module. As for the assembly, these tests were carried out following a systematic procedure formalized in the form of check-lists and a series of data, logs, analysis results and plots recorded and backed up. Each SSD was tested for mechanical issues, light-leaks, general performances and thorough characterisation of the detector.

The first step is to control the mechanical integrity of the detector. This entails making sure that each part of the SSD has been assembled properly, that the aluminum top cover has been glued properly (especially in the corners), that no unwanted holes have been drilled

in the aluminum casing and that there are no missing rivets. Any light-leak (i.e photons from the outside entering the aluminum casing, not produced by the scintillators) would compromise a SSD, especially given that the detectors in the field will be completely exposed to a very bright sun. It is thus of utmost importance that the modules are checked for light-leaks during the characterization process.

Detector characterisation   The main mission of a SSD is to deliver a signal corresponding to the amount of secondary particles going through it, as the WCD under it is triggered. The output of the PMT is digitized to get the signal amplitude as a function of time, called the "trace". The obvious way to check that an SSD module behaves as expected is to measure and characterize the signal it produced when a vertical muon goes through it. In order to compare detectors built at different sites as well as to set some initial calibration parameters, it was agreed to define some metrics based on the signal of through going muons : the **M**inimum **I**onizing **P**article (MIP), **V**ertical MIP (VMIP) and **S**ingle **P**hoto-**E**lectron (SPE) values in charge. The minimum quality requirement for the SSD is to have at least 15 photoelectrons per MIP.

We already introduced the MIP in subsection 5.2.1 as the analog for the SSDs of the VEM for WCDs. A minimum ionizing particle is usually defined as the signal produced by a particle whose mean energy loss through matter is close to the minimum, usually relativistic particles are MIPs. In our test setup, the "MIP signals" are peaks in our signal traces, produced by relativistic muons going through our detectors. The VMIP value is reconstructed from a subset of MIP values selected for being produced by vertical muons. All the MIPs we select for the tests are more or less vertical (), due to the geometry of the test setup, but we will see in subsection 5.3.4 that we can use the timing information on the trigger signals to further constrain the arrival direction of the muons generating the MIPs, thereby defining VMIPs.

The SPE is the response of a PMT to a single photo-electron. Measuring the mean of the charge distribution of SPE is a widely used technique to accurately calibrate the gain of a PMT. By expressing the signals obtained with a PMT in units of mean SPE value, one can compare the signals obtained at different gains or even with different PMTs, provided one knows how to compare their quantum and collection efficiencies. Having the MIP **and** SPE values allows for a measurement of the detector's efficiency with minimal dependency on the PMT used and on the HV settings by using the MIP per SPE observable.

5.3.3   Setup description

A baseline design for the test setup was agreed upon by all the SSD assembly sites. The minimum setup requirement was to have two plastic scintillator particle detectors, one above and one below the SSD, acting as an external trigger, and a data acquisition system. A Hamamatsu PMT (similar to the one ultimately used in the field) was provided to each institute for readout. Each institute was then free to design and mount its own test setup, using the materials at its disposal.

Test setup   I will now present the setup that was used at the LPSC. During the construction, the test setup was installed in the building where the construction was done. That way, the SSDs were continuously tested through the months of construction.

The assembled modules are installed on a table, with two scintillator boards on each side, to act as an external trigger. These 4 scintillator boards[1] are narrow and cover the whole SSDs width (1500 mmx132 mmx20 mm). A PMT is installed on the SSD to read its signal

---

[1]originally used by the ALICE experiment

while the scintillator boards had built-in PMTs on both sides. The external trigger is used to select through-going muons.
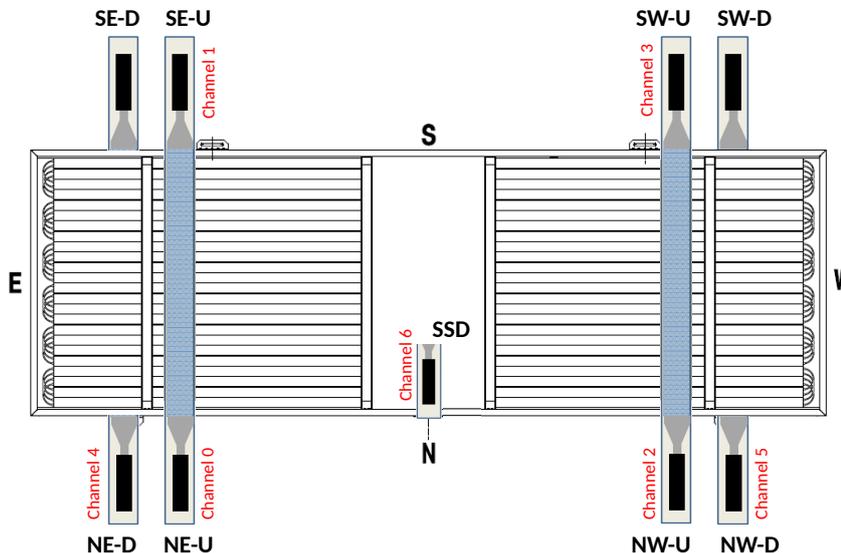
Figure 5.10: Scheme of the setup used to test and characterize the SSDs response at the LPSC. The PMTs on both sides of each scintillator bars making up the external triggers are identified as SE/SW/NE/NW, their localisations on the cardinal axes inside the hall, and -U/-D, whether they are above/below the detectors.

All 9 PMTs were powered by NIM based CAEN HV power supplies. The 8 PMTs of the external trigger (Photonis XP2020) were essentially used as a trigger and for precise timing. They were constantly supplied with the same voltage defined once for ever to provide a good efficiency while minimizing noise induced triggers and to achieve high timing accuracy on their signals. The voltage supplied to the SSD PMT has a direct impact on the signal read.

To collect the SSD light, we used one Hamamatsu R9420-10 photomultiplier. A PMT with an active ISEG base was first tested, but the electronic noise level generated by the active base was too high to properly get to the level of SPE signals. Instead we tested a PMT provided by the IPNO, same reference but with a passive base provided by Hamamatsu. We chose to supply it with a voltage > 1400 V, 1400 V at first and then at 1410 V mid-construction, to counterbalance the drop in temperature in the hall. With a lower voltage, the SPE peaks could not be resolved with the SPE search algorithm we designed (see subsection 5.3.4). A fast amplifier and a passive filter were added to get gain and baseline fluctuations compatible with the SPE measurements.

Acquisition system All 9 PMTs are then read-out using a 8 channels 12-bit WaveCatcher digitizer [179]. The WaveCatcher can read 8 channels of 1024 time-bins simultaneously, while there are 9 PMTs in the test setup, the output of the PMTs installed on the bottom scintillators were thus grouped two by two (same side together) as can be seen on Figure 5.5.

On the WaveCatcher, one can adjust the sampling frequency from 400 Msamples/s up to 4.2 Gsamples/s. Given that the WaveCatcher has a fixed number of inputs (1024), a higher sampling frequency will mean a shorter event trace. We used a sampling frequency of 400 Msamples/s for our tests, to maximize the number of potential SPEs recorded in
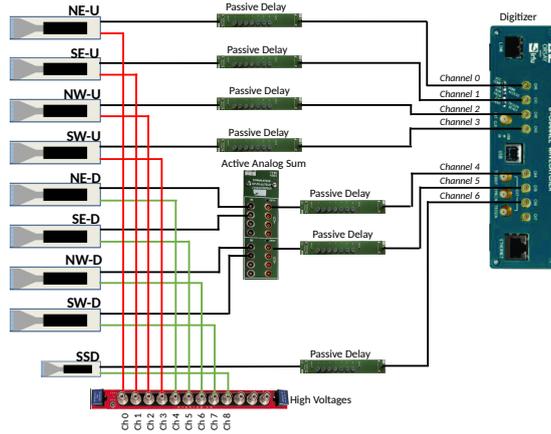
Figure 5.11: Scheme presenting the way the PMTs are read-out with the WaveCatcher digitizer.

our traces and because the resolution achieved in this configuration for the MIP peaks was sufficient for our purpose.
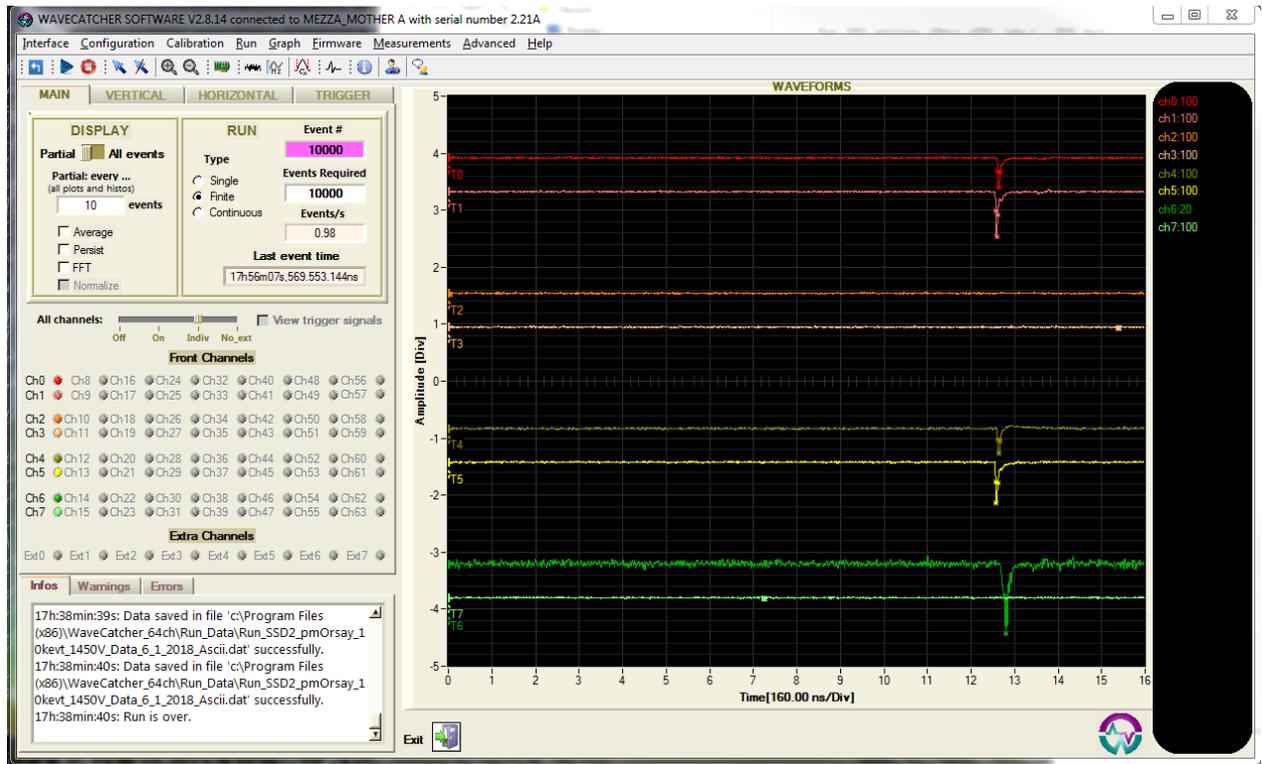
Figure 5.12: A screenshot of the WaveCatcher software, taken during acquisition. This interface allowed us to control that the acquisition was going smoothly and to check for any major lightleaks.

The WaveCatcher digitizer also offers an interface [180] to read-out the PMTs signals in real-time as can be seen in Figure 5.12. This interface is really useful and convenient, because it allows the user to control that the acquisition was started properly and is running smoothly. The WaveCatcher is also constantly recording the raw signal rate above a low threshold between triggers. This rate can be displayed as function of time. This display was used to easily detect light-leaks due to mechanical defects at the beginning of the assembly process. Using a strong flashlight (above 1100 lumen, more than a direct sunlight at noon when applied at less than 30 cm) and running it along the aluminum casing, we can check in real-time for an eventual light-leak that would produce spikes in the SSD signal and that is easily

detected as a sudden rate increase by a very large factor. This test was performed around each detectors, scanning every sensitive areas of the detectors, in particular the juncture between the assembled parts.

Additionally to those *online* functions, the WaveCatcher also provides tools for *offline* analysis of the data recorded around the triggered events. Histograms of the charge and amplitude of the signal, of the noise and of the variance during the data acquisitions are automatically produced. This information is very useful to control the quality of the constructed modules, but a more in-depth analysis is required to characterize the detectors response to an ionizing particle.

5.3.4   Description of the framework to search for SPE and MIP peaks

When an ionizing particle goes through two of the external scintillators, one on top and one on the bottom, it acts as a trigger for the WaveCatcher to record *an event*. When a trigger happens, the WaveCatcher records the signals for all 8 of its channels. As was previously stated, the user can modify the sampling frequency, but he can also choose the location of the trigger on the traces. The traces are the recorded signals from each entries digitalized and sampled in 1024 time bins of 2.5 ns. It was decided to record as much signal before the trigger as possible so that while the triggering signal (MIP peak) is fully recorded inside the trace, a large portion of the trace is available to look for SPE peaks without having to worry about trailing signal or electronics undershoot after photons from the MIP peak. For each detector, at least two runs were performed, one with the detector shielded from light by an opaque fabric cover, and one without. A "run" is a collection of signals from the WaveCatcher made up of 50 000 triggered events (typically 50 min long). The two types of "runs", covered and uncovered, were used as a further check of the modules light-tightness: if the uncovered run didn't show a significant increase in the rate of SPE peaks, it was considered as light-tight. While the traces are recorded in amplitude, the peaks values used for the calibration are expressed in charge (in pC). The relation from amplitude to charge is:

$$P_C = (P_{Amp} - Nbin \times S_0) \times \Delta t / \Omega \tag{5.7}$$

where P_C and P_{Amp} are respectively the amplitude and charge value of the peak, Nbin is the number of bins inside each trace, S_0 is the mean baseline, Δt is the duration of one bin and $\Omega = 50$ Ohms, the resistive load to the voltage divider of the PMT as well as the input impedance of the readout electronics.

With thoses traces, the analysis is done in four steps :

- check whether the trace is usable or if it is dominated by noise

- perform a sliding window search for any peaks along the trace

- categorize the peaks into SPE or MIP

- perform a fit of the peaks in each trace, then fit the peaks distributions to get values for the SPE and MIP peaks charge for each SSD

Finding peaks inside the traces: The procedure to locate peaks in the traces is illustrated in Figure 5.13. A first test is done on each trace to make sure that the trace is not dominated by noise. This is done by fitting an histogram of the baseline of the trace (i.e the signal in the traces with their peaks removed) with a gaussian function and applying a cut on the sigma value. This first cut is applied to get rid of aberrant traces for example because of
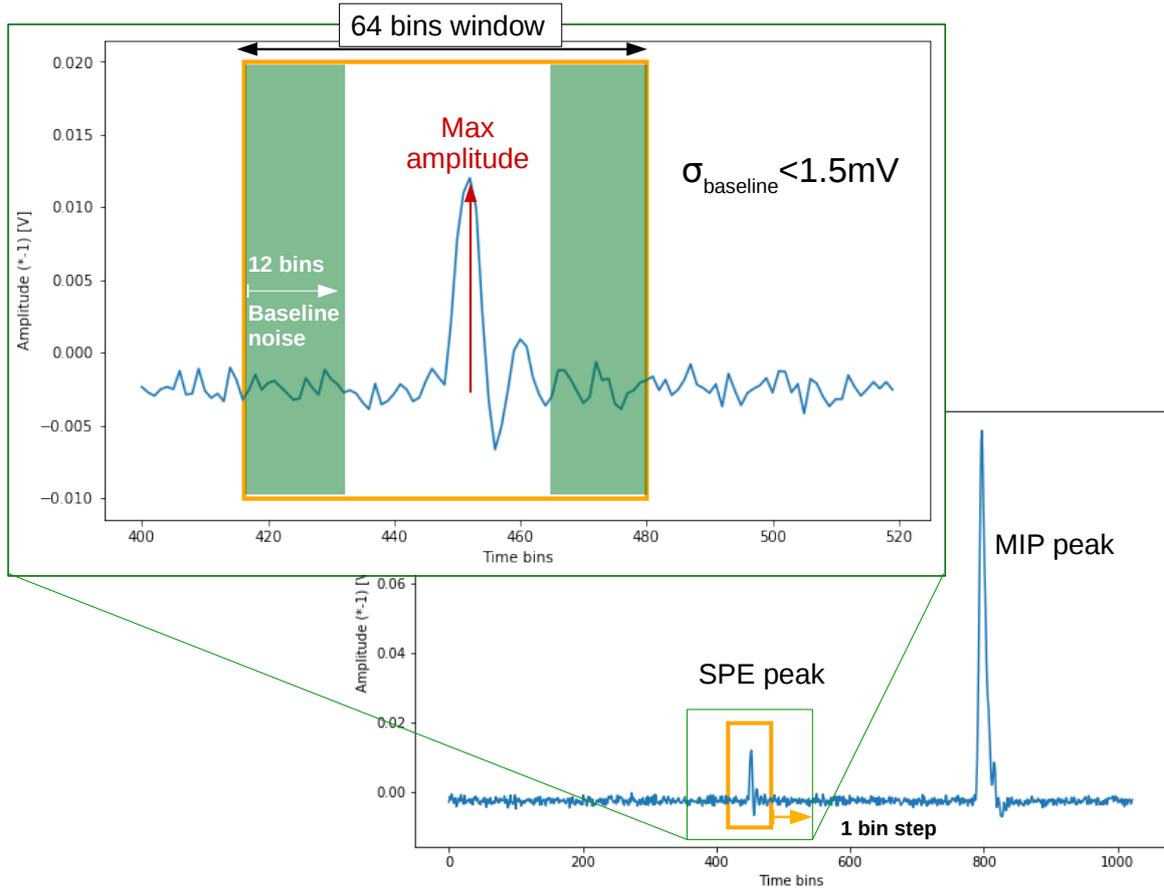
Figure 5.13: Simplified scheme of the sliding window search performed on the traces (see text).

strong EM currents bursts induced by the mechanical bridge electrical supply or other similar but hopefully rare events.

Then, starting at bin 0, a 64 bins window is run along the trace. Inside the window, we look for the maximal amplitude bin. The peak is discarded if the signal value isn't within unbiased values [-5 mV : 50 mV] or if the total charge inside the window is beyond 20 pC. The first and last 12 bins are used to fill an histogram, which is fitted to characterize the baseline noise level in that time window. If this baseline noise level is higher than 1.5 mV, the window is classified as too noisy. If the peak is not centered, the window is slid by 1 bin and the process iterates, if the trace is too noisy or if the peak is discarded, it is slid by 64 bins. Once an unbiased and centered peak is located within a quiet window, it is recorded as a peak and the search continues after a 64 bins increment.

Classifying the peaks: With all peaks found, the baseline noise is recalculated with the peaks removed and the baseline noise $\sigma_{baseline}$ is measured. The peaks located well before the trigger bin ($<$ bin 640) are tagged as SPE peaks.

SPE peaks are fitted using a user-defined function, calibrated on LED-pulser runs, where very low amplitude and very narrow light pulses where sent to the window of the PMT and read out using the test setup electronics and data acquisition. An example of a SPE peak fit on a led-pulse run is shown in Figure 5.14. The fit-function takes into account the bipolar shape of the peaks (induced by the 50 MHz low-pass filters used upstream of the WaveCatcher) and the fit is done on the whole 64 bins window to also fit the baseline in

order to subtract it in the calculation of the SPE peak charge.
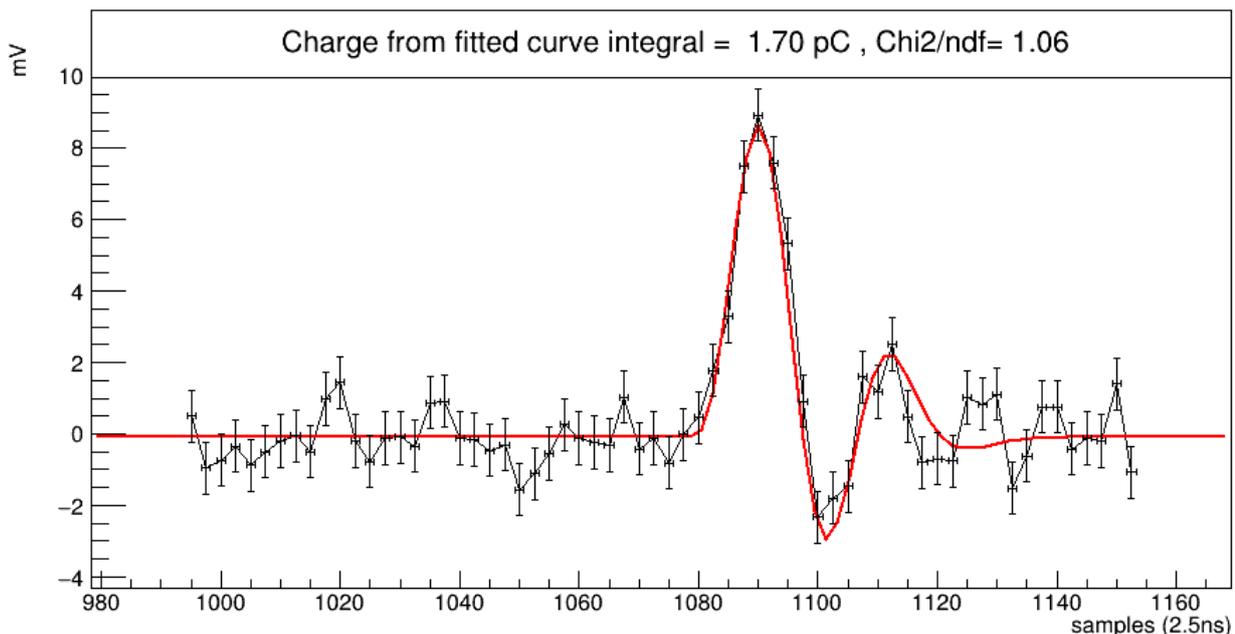
Figure 5.14: Example of a SPE peak fit. This particular pulse was taken from a LED-pulse run.

The MIP peaks are easier to find as we know precisely the bin range in which they should be located since they are caused by the triggering through-going particle. As such, we search for MIP peaks by finding the maximum between bin 750 and bin 850. For each trace we only record as MIP the highest peak in this range. With this MIP peak identified, the MIP value is calculated by summing the charge (after baseline subtraction) of the bins making-up this peak. The limits of this integration window are defined as the crossing of a $4\sigma_{baseline}$ threshold.

Vertical-MIP selections: The scintillator boards used for the external triggers are narrow but long, this means that some through-going particles triggering the data acquisition may be inclined along the boards longitudinal direction. To further constrain the selected MIP in verticality (MIP -> VMIP), a timing selection was performed. Using the time-difference between the signal reception from the PMTs on both side of the scintillator boards, we can estimate the position the ionizing particle went through. To go from signal timing to position on the boards, a narrow scintillator slab was used to get the empirical correlation factors: A_{EU}=1.441, A_{ED}=1.110, A_{WU}=0.779, A_{WD}=-0.056 and B=0.127, which gives us:

$$\mathrm{Pos}_j = (A_j - (t^{j0}_{LET} - t^{j1}_{LET}))/B \text{ with j in EU,ED,WU,WD} \tag{5.8}$$

where Pos_j is the position of the triggering particle on the external trigger scintillator boards and t^{j0}_{LET} and t^{j1}_{LET} are the *Leading Edge Time* (the start of the rise in signal) recorded by the PMTs at both ends of the external trigger scintillator boards. With the test setup, a timing resolution of 0.8 ns can be achieved, which corresponds to a resolution of 3 cm on the particle position along the board.

With this estimation of the position where the triggering particle went through the boards, we can apply a minimum distance between the position on the top board and the one on the bottom board: ΔY=13.5, to further constrain the verticality of the triggering particles. In the characterization algorithm, a MIP is tagged as a VMIP if:

$$\mathrm{abs}(\mathrm{Pos}_{EU} - \mathrm{Pos}_{ED}) < \Delta Y \text{ or } \mathrm{abs}(\mathrm{Pos}_{WU} - \mathrm{Pos}_{WD}) < \Delta Y \tag{5.9}$$
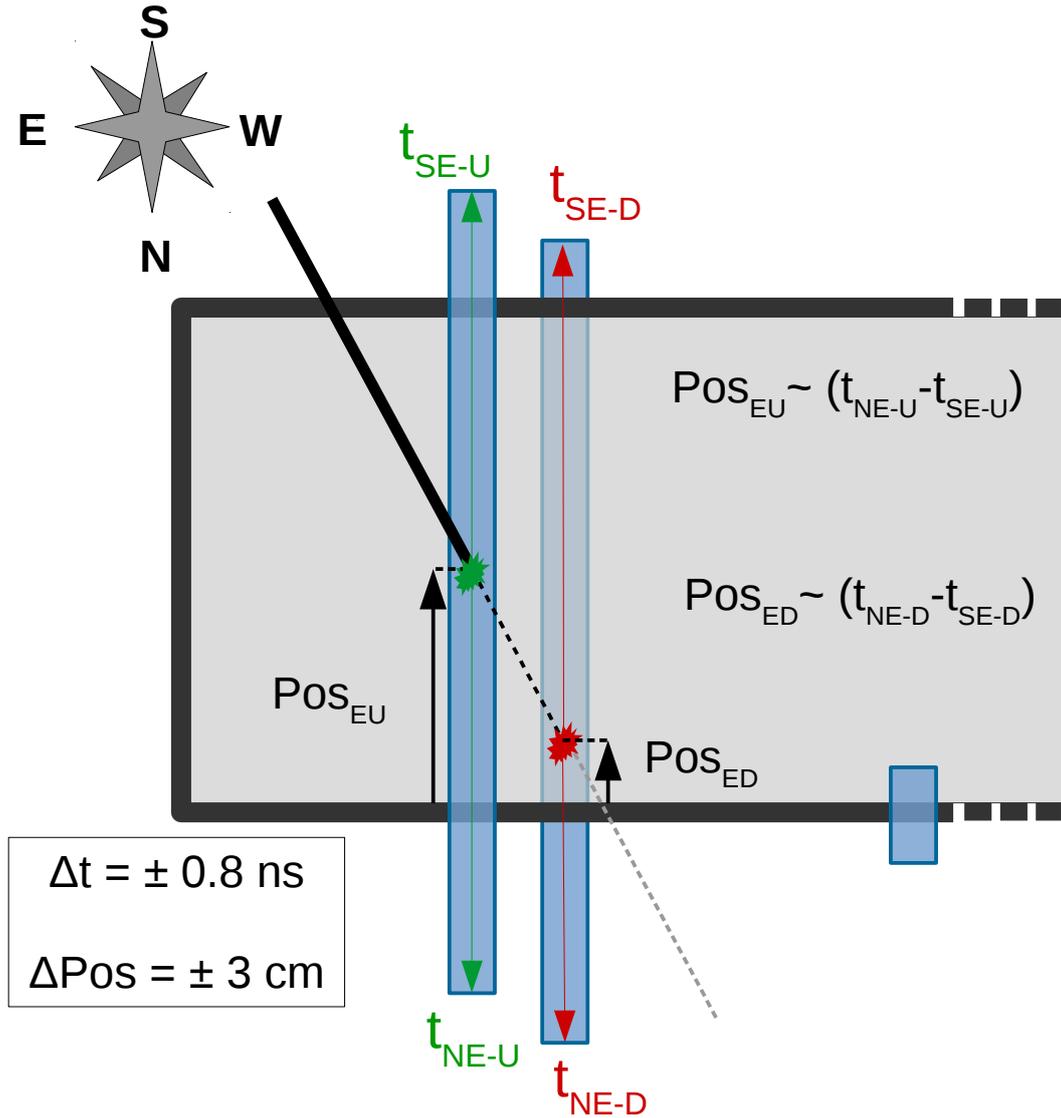
Figure 5.15: Scheme of the procedure to reconstruct the position at which a particle went through the boards to trigger the test bench. The time difference between the signals from both side of a single scintillator bar of the external trigger, can be directly linked to the position of the triggering particle.

Being able to estimate the position of the muons, allowed us to further select MIPs on their arrival direction along the longitudinal direction to zenith angle below 25° as well as control the efficiency of the SSDs across the width of the SSD. [1].

5.3.5 Results of the tests

The procedure described has been applied on all 90 SSDs constructed at the Grenoble site.

As already mentioned in subsection 5.3.1, a procedure was written at the beginning of the construction process, to guide the several operators during the construction process. When mistakes were made on the mechanical integrity of the modules, they were directly fixed and, if needed, the procedure was appended. This document was thus updated as the construction

---

[1]a scintillator bar inside the SSD was incapacitated to check that its absence was seen in the efficiency measurement

went on and as specific tools were designed.

The *online* search for light-leak proved to be extremely useful, as the few light-leaks recorded on modules were identified during this test. Two kinds of situation came up: 1. the composite panel being not properly glued to the frame in the corners, resulting in leaks and 2. unnecessary holes being drilled in the aluminum casing. In both kinds of situation, we were able to identify the source of the leaks and repair the detectors. In total, we had to repair 6 detectors at the beginning of the construction and these problems didn't occur once the issues during the assembly process were found out. The *offline* search (differences between covered and uncovered runs) did not reveal any other light-leaks however all the *leaking* SSDs were correctly identified as having a leak with the *offline* search, proving its validity.

As can be seen in Figure 5.16, the mean values of MIP and SPE per SSD are relatively stable throughout the construction. The mean MIP values per SSD are stable around 55 pC which corresponds to around 30 PE per MIP peak since the mean SPE values are around 1.9 pC. At the end of the production (SSDs number 83 and above) the mean MIP values started rising above 55 pC. The cause behind this increase is not fully understood, it might be due to a rise in the temperature, PMT aging or a scintillator light yield increase in the last batch used.
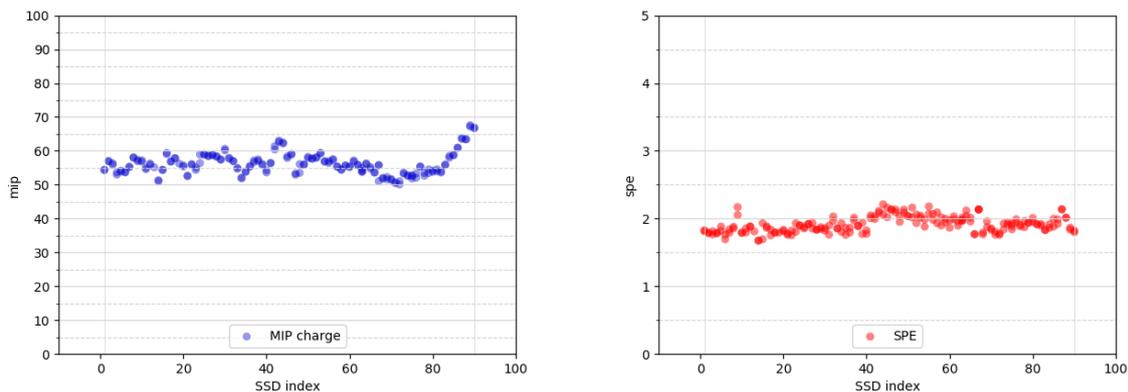
Figure 5.16: Distributions of the SSDs mean MIP and SPE values in pC by their production number. The values themselves are within the specifications set by the collaboration around 55 pC and 1.9 pC respectively. The stability of the values through the construction is also satisfactory, although we observed an increase in the mean MIP values on the last few SSDs tested.

On Figure 5.17, the mean VMIP values for each SSD and the VMIP/SPE values are displayed. The mean VMIP value for each SSD is of course directly correlated to the MIP values, hence it follows the same trends through the construction. The VMIP values are slightly lower due to the more vertical muons leaving less energy in the scintillators as they go through less matter. The more interesting VMIP/SPE value is a reliable estimator of the detector performances. Indeed, the MIP and the VMIP values in pC directly depend on the PMT dynode chain gain thus on the value of the HV applied. A less depending observable is the VMIP/SPE, in which the effect of the geometry of the setup is minimized (thanks to the cut narrowing the zenith angle distribution) but also because gain dependence vanishes in the ratio. Thus, this is considered as the proper observable allowing to compare results of test done with different setups. Note that there remain a setup dependence in the quantum and the collection efficiencies of the PMT used.

In practice, we have seen that the SPE values are highly sensitive to the search and fitting algorithm used. In particular we saw that a search for peaks based on a threshold crossing

criteria leads to SPE values varying a lot depending on the value of the threshold [1]. With the setup used in these tests, the distribution of PE peaks is dominated at low energy (below the real SPE value) by *"fake peaks"* due to noise in the traces. The temptation of course, to find a clear SPE peak in the distribution is to cut out these values. Unfortunately, a threshold set too high will also cut-out the beginning of the SPE peaks distribution, resulting in an over-estimation of the SPE value.
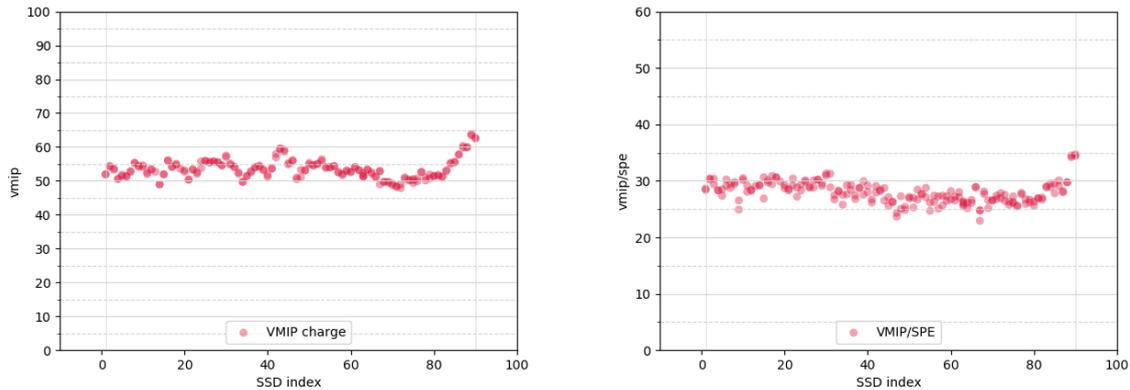
Figure 5.17: Distributions of the SSDs mean VMIP values in pC and VMIP per SPE values by their production number. The mean VMIP values are close to the mean MIP values (see Figure 5.16) but shifted down, as expected since the more vertical muons produce less light as they go through less matter inside the scintillators.

The VMIP/SPE values are the ones that are ultimately used to compare the performances of the detectors between institutes, and the values that were recorded on the 90 SSDs were satisfactory. The histogram of the VMIP/SPE values of the 90 SSDs in Figure 5.18 shows two things: 1. that the VMIP/SPE ratio is stable throughout the produced SSDs and 2. that the mean value μ=27.97 PE per VMIP. The stability is a good sign that the construction process went well and that the detectors have a similar response. The mean value is within the collaboration specification for the SSDs.

In other institutes, similar procedures have been applied to measure the same quantities. As the characterization of the detectors proceeded, the tests results, as well as relevant observations about each detector's construction process were recorded on a collaboration-wide database.

The KIT has also performed its test procedure, on a few SSDs from each assembly sites (see Figure 5.19). The test setup used at KIT was a muon tower, triggered with 3 tower planes in coincidence and the peak search algorithm is similar to the one used at the LPSC, peaks are located in the traces and classified based on their position in the traces. By performing the same measurement technique with the same setup, and especially the same PMT, it was demonstrated that the detectors built in the various assembly sites were all within the collaboration specifications and that their performances were consistent with each others.

---

[1]whether the threshold value is based on the baseline noise of the trace or if it is a set value
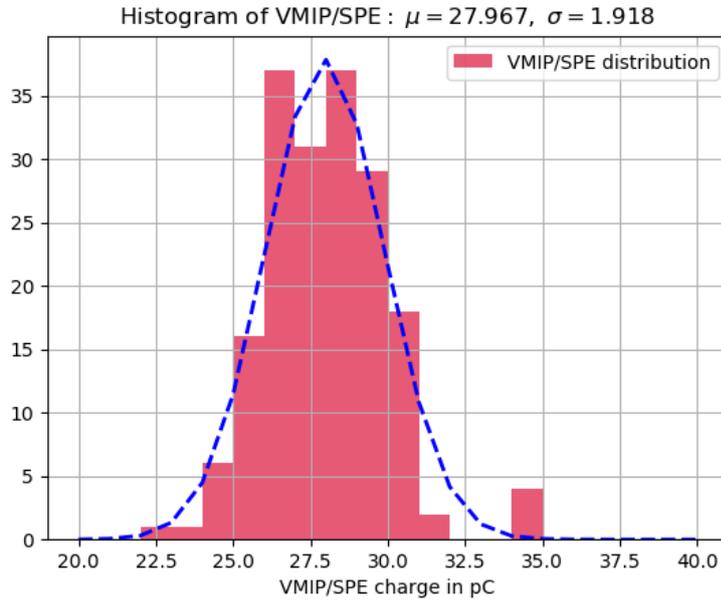
Figure 5.18: Histogram of the mean VMIP/SPE values of the 90 SSDs. The dashed blue line is the fit of the distribution with a gaussian function with parameters μ=27.97 and σ=1.92.
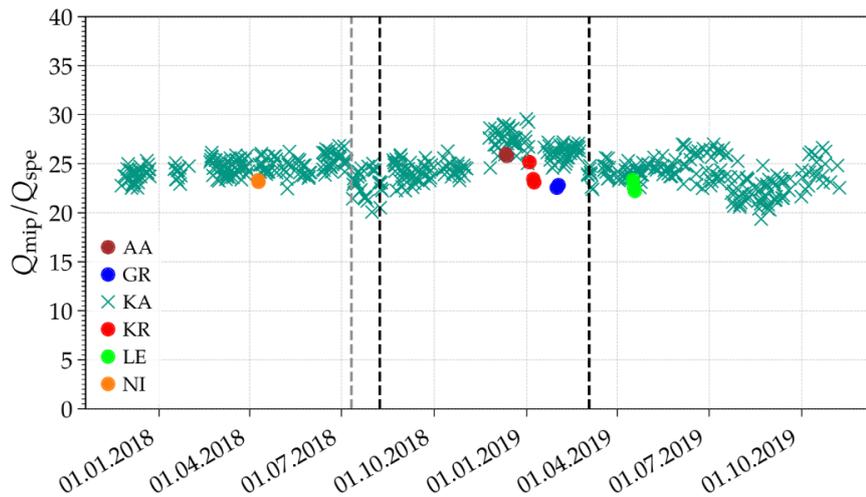
Figure 5.19: SPE per MIP values obtained with the KIT test setup, for SSDs assembled in the several assembly sites.

Chapter 6

Ultra-High Energy photon searches

Contents

In Chapter 1, the reasons why the search for UHE photons would be valuable in the identification of the UHECR sources and in the comprehension of the acceleration mechanisms and the propagation of UHECR were explained. The discriminating features of photon-induced EAS, compared to showers induced by hadronic primaries were presented in Chapter 2. In Chapter 3, the capabilities of the Pierre Auger Observatory to detect the showers and to reconstruct UHECR parameters with an unprecedented precision were described, as well as the high quality results obtained concerning the UHECR physics. In this chapter the main subject of this work, the searches for ultra-high photons, is presented in more details. Relevant observables built from EAS parameters and used for UHE photons searches are introduced. Then, the searches for UHE photons conducted by UHECRs observatories are described. Finally, the observables and analysis methods used for this particular work are introduced.

6.1 Main observables for photon searches

6.1.1 Depth of the maximum of the shower development (X_{max})

As was seen before in subsection 3.2.1, the development of a shower through the atmosphere, can be used as an indicator of mass composition, since the EM part of an EAS develops at a slower pace compared to the hadronic part,

The depth at which a shower reaches its development maximum, X_{max}, is obtained by fitting a Gaiser-Hillas parameterization (see subsection 4.2.3) on the longitudinal profile of the shower using two variables related to the width and asymmetry of the profile [181]. The X_{max} is inversely correlated to the primary's mass for a given energy (see subsection 3.1.2).

The value of X_{max} obtained with the FD is very robust since it is directly and carefully measured, and it proves to be the observable with the best discrimination power to separate the mass of the primaries for most hybrid events. However, with events detected by the SD detector only, one has to rely on other indirect estimation of the X_{max}. Other observables exist to access the information on a shower's longitudinal development, in an order to reproduce the X_{max} value obtained with the FD.

6.1.2 Radius of Curvature

One example of how to extract information on a shower's longitudinal development is to look at the radius of curvature: Rc. As was described before in subsection 3.1.3 and 4.3.1, the shower front is not planar but curved toward the first interaction point. Thus, the first particles are increasingly delayed compared to a planar shower front, the further away the station is located from the shower core (r). The shower front can be approximated by a spherical model, and estimated using the trigger time of the first particles in a station. The radius of curvature of a shower Rc can be obtained from the fit of the spherical shower front.
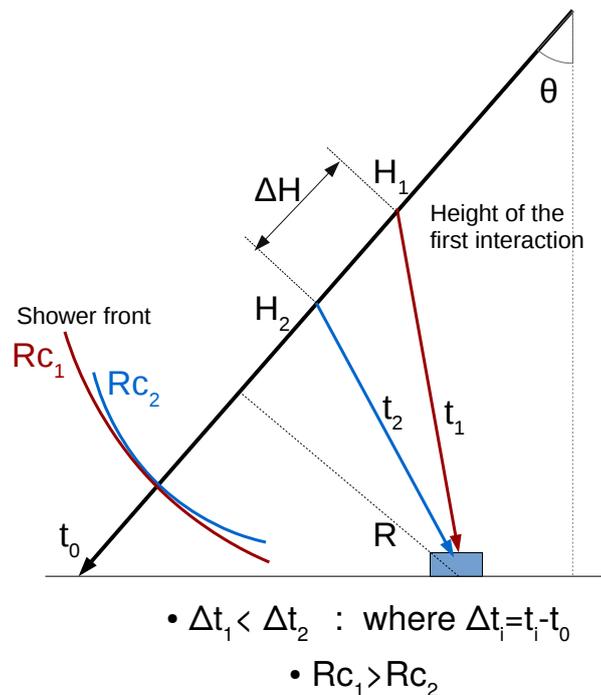
Figure 6.1: Illustration of the geometrical effects involved in the Radius of Curvature calculation.

For a particle due to an interaction at height H on the shower axis and observed in a SD

station at distance r from the shower core, the time delay Δt can be expressed as:

$$\Delta t = \frac{1}{c}\left(\sqrt{H^2 + r^2} - H\right) \propto \frac{r^2}{H} \quad (r \ll H) \tag{6.1}$$

where $H = H(\theta, E)$ is the first interaction height. In Figure 6.1, the geometrical effects involved in the measurements of Rc are presented.

The average height of the first interaction of an EAS depends on the mass of the primary, with photon-induced showers having a first interaction at higher depth (lower altitude) for a given zenith angle and primary energy. The delay Δt at a given lateral distance r will decrease with the first particles' interaction height $H(\theta, E)$. Thus, larger delays are expected for an EAS initiated by a photon primary. In addition to this geometrical effect, secondary muons can reach the ground from higher in the atmosphere, further decreasing the delays for hadronic-induced showers.

6.1.3 RiseTime

The time spread Δt of the secondary particles inside each SD station is a local sample of the shower-front thickness. For a station at distance r from the core and for secondary particles originating from interactions at heights $[H - \Delta H, H]$ on the shower axis, the time spread Δt needed for all secondary particles to reach the station increases as H decreases. By using equation 6.1, one gets:

$$\Delta t_1 \propto r^2\left(\frac{1}{H_1 - \Delta H} - \frac{1}{H_1}\right) \quad (r \ll H) \tag{6.2}$$

$$\text{and} \quad \Delta t_1 < \Delta t_2 \quad \text{for} \quad H_1 > H_2 \tag{6.3}$$

A smaller production height H entails a larger time-spread, hence, for a fixed core distance, photon-induced showers are expected to have larger Δt due to their larger X_{max}[1].

In the Auger reconstruction, this time-spread of the signals is used to build the station-level variable $t_{1/2}$ [182], the so-called RiseTime of a station, calculated as the time it takes for the station signal S to go from 10% to 50% of its integral:

$$t_{1/2} = t(S_{50\%}) - t(S_{10\%}) \tag{6.4}$$

According to the discussion above, on average and at a fixed core distance, $t_{1/2}$ is expected to be lower for showers from hadronic primaries than for photon-induced showers.

However, the position of a station on the projected shower plane also has an effect on the time-spread if the shower is not vertical ($\theta \neq 0°$). Two stations located at the same distance r from the shower core, do not sample the shower at the same stage of their development, depending on their projected position on the shower axis plane. There is an assymetry depending on a station azimuth ζ, whether it is *early* or *late* on the path of the shower front. The station with an azimuth closest to $0°$ samples the shower at its earliest and the station at $\zeta = 180°$ samples it at its latest. Furthermore, the assymetry will increase with the inclination (zenith angle θ) of the shower [183]. In Figure 6.2, the geometrical effects involved in the calculation of $t_{1/2}$ are displayed, along with a simplified shower plane to represent the azimuth angle ζ of a station.

Although the asymmetry can be used for mass composition studies [183], it has to be corrected to make shower to shower comparison. The correction to $t_{1/2}$, for a station at

---

[1]since a larger X_{max} entails that, on average, the shower particles are produced deeper in the atmosphere (larger H).

Figure 6.2: Illustrations of the geometry involved in the RiseTime calculations. The secondary particles produced over a given length ΔH reach a *late* station with a smaller angle, meaning a shorter time spread than an *early* station located at the same distance r from the shower core.

Figure 6.3: Example of the $t_{1/2}$ calculation on a data trace.

a distance r from the shower core, is done by parameterizing the shift resulting from the asymmetry, and subtracting that term to $t_{1/2}$ to get an asymmetry-corrected value, $t_{1/2}^{corr}$.

$$t_{1/2}^{corr} = t_{1/2} - g(\theta,r).cos(\zeta) \tag{6.5}$$

The value of the parameters for the function $g(\theta,r)$ were obtained by fitting the function on selected sub-samples of observed data. The subsamples used to fit the function, from which the parameters are derived, were selected by following a series of cuts to reject saturated stations and focus on a range of shower energies and r values to ensure the quality of the data:

- expected signal $S_{exp} > 10$ VEM and observed signal $S_{obs} > 3$ VEM
- energy $10\,\mathrm{EeV} < E < 30\,\mathrm{EeV}$

Figure 6.4: Dependance of $\langle t_{1/2}/r \rangle$ on the azimuth angle ζ for events with energy between $18.55 < \log(E/\text{eV}) <$ 18.70 in bins of zenith angle θ. Taken from [183].

- distance $600\,\text{m} < r < 1600\,\text{m}$

- divide in 6 bins of θ in the range $\sec(\theta)=[1.0,\ 2.0]$

In the rest of this work, unless mentioned otherwise, any mention of $t_{1/2}$ will refer to the asymmetry corrected value $t_{1/2}^{corr}$.

Construction of a discriminant observable

To combine these $t_{1/2}$ values into a discriminant observable at the event-level, one method is to estimate the value of the variable in a station located at $1000\,\text{m}$ from the shower core and at azimuth $\zeta = 0°$. This is done by parameterizing $t_{1/2}$ as a function of the distance from the core r (see equation 6.1):

$$t_{1/2} = (40 + \sqrt{a^2 + b.r^2} - a)\,\text{ns} \tag{6.6}$$

This empirical function (equation 6.6 from [184]) is then fitted on the individual stations to get the values of parameters a and b. The $t_{1/2}(r = 0)$ value is set at $40\,\text{ns}$ as that is the mean response time of a station to a particle. The value of the risetime at $1000\,\text{m}$, $t_{1/2}(1000)$ can then be calculated for each shower.

Another way of combining the information of the stations $t_{1/2}$ is the $\langle \Delta \rangle$-method [184]. In this method, the goal is also to create a discriminant observable using the $t_{1/2}$ of a station. However, instead of fitting a function on each event, the $\langle \Delta \rangle$-method first builds *benchmarks* of the mean values of $t_{1/2}$ as a function of distance r for events with energies

$19.1 < \log(E/\text{eV}) < 19.2$ in bins of zenith (a benchmark is built for each of 6 bins covering $1.0 < \sec(\theta) < 1.45$). Then the observed values of $t_{1/2}$ in each station are compared to the benchmark predictions in this particular event's energy and zenith angle and combined together to build the $\langle \Delta \rangle$ value:

$$\langle \Delta \rangle = \frac{1}{N_{st}} \cdot \sum_i \delta_i = \frac{1}{N_{st}} \cdot \sum \left(\frac{t_{1/2} - t_{1/2}^{bench}}{\sigma_{t_{1/2}}} \right) \tag{6.7}$$

where δ is the uncertainty weighted residual of the risetime, N_{st} is the number of stations, $t_{1/2}^{bench}$ is the benchmark prediction of $t_{1/2}$ and $\sigma_{t_{1/2}}$ is the uncertainty on $t_{1/2}$.
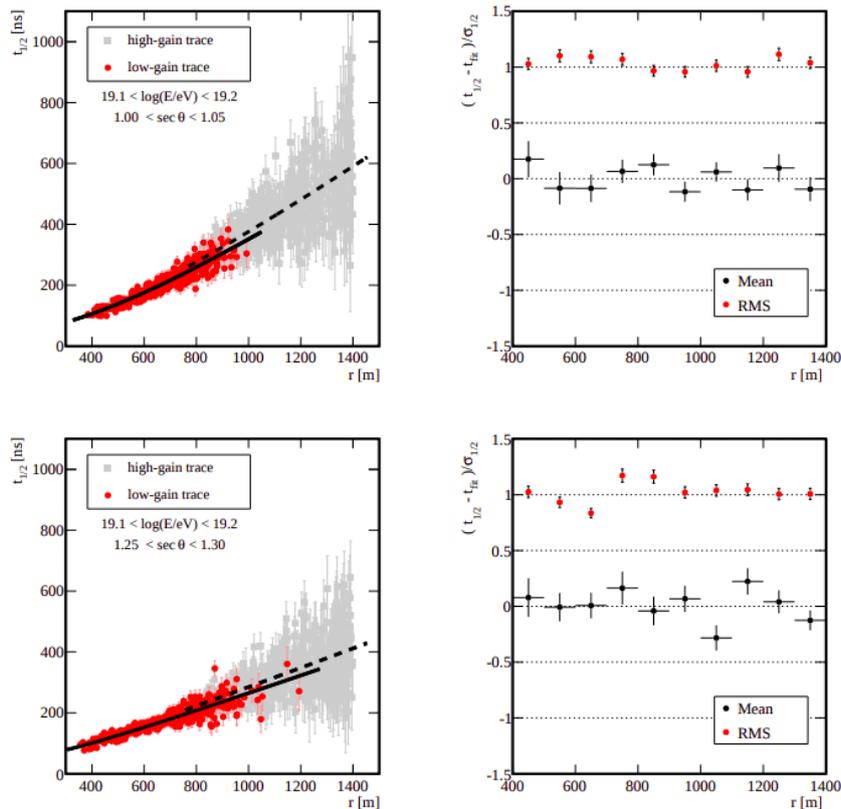
Figure 6.5: Examples of benchmark fit for $1.00 < \sec(\theta) < 1.05$ (top) and $1.25 < \sec(\theta) < 1.30$ (bottom). Taken from [184].

6.1.4 Number of triggered ground detectors

The lateral distribution function of an electromagnetic shower is steeper than the one of a hadronic shower (see subsection 3.1.3). This, combined with the fact that the SD triggers are optimized for muons, leads to the number of stations triggered (N_{st}) by a photon event being lower on average compared to the N_{st} of a hadron-induced shower of same zenith angle and primary energy. Hence, N_{st} can be used as a discriminating observable in photon-search analysis in and of itself, although it is usually implicitly used inside other, more complex event-level observables (RiseTime, Muonicity,...).

6.1.5 Lateral distribution function and S_b

The lateral distribution function (LDF) of secondary particles is used in the SD standard reconstruction to estimate the S_{1000} value of a shower (see subsection 4.3.1). However, the

parameterization of the "standard LDF" is optimized for data and does not properly fit the lateral distribution of the signals in case of photon showers, as they have, on average, steeper LDFs (see subsection 3.1.3).

This inappropriate parameterization of the LDF for photons can be used to build an event-level observable, by combining the differences in each station signal with the expectation from the data-parameterized LDF. The most straightforward way to account for the quality of the LDF fit on the stations signal would be to use the χ^2 of this fit. However, no conclusive analysis has used it as a satisfactory separation power [185].

Instead, an observable that directly compares the signal observed S_i in the stations at distance r_i from the shower core, to the signal expected from the data-fitted LDF, $\text{LDF}(r_i)$, was built, L_{LDF}:

$$L_{LDF} = log_{10} \left(\frac{1}{N_{st}} \cdot \sum_i \frac{S_i}{\text{LDF}(r_i)} \right) \tag{6.8}$$

Other observables, have been constructed to try to obtain a better separation power. Thus the information from the steeper LDF and N_{st} can be combined into a single observable, S_b:

$$S_b = \sum_i^{N_{st}} S_i \cdot \left(\frac{r_i}{r_0} \right)^b \tag{6.9}$$

where S_i and r_i are the signal and distance to the core respectively of station i, $r_0 = 1000\,\text{m}$ and b is a free parameter[186][187], that represents the shape of the LDF.

6.1.6 Muon density

The ratio of muonic to electromagnetic component of an EAS at ground-level is an indicator of its development through the atmosphere (see subsection 3.2.1). Thus, it is an indirect indicator of the primary nature, and boasts a very high separation power between hadron-induced showers and photon-induced showers.

However, the number of secondary muons at ground-level may not be directly accessible. One reason is that EAS observatories only **sample** the footprint of the shower. But also and mainly, it is because muons may not be distinguishable from other secondary particles in terms of signal shape, or be integrated together with them in the same time bins ("insufficient" timing resolution of the detectors).

6.1.7 Area-over-Peak

The Area-over-Peak (AoP) ratio is defined as the ratio between the integral of the signal and the value of its peak. The AoP is related to the spread in time of the traces, and thus provides an information on the thickness of the shower front.

Beside being used in the calibration of the WCDs in the PAO, the AoP is one of the main discriminant observable for neutrino searches[1]. It is used in other experiments as a discriminant parameter for photon searches (see subsection 6.2.1).

6.2 Current status and results of UHE photon searches

Ever since the first detection of UHECRs above GZK energies, EAS observatories have searched for UHE photons. The first results from UHE photons searches were published

---

[1]In the neutrinos searches [188] of the Pierre Auger collaboration, the AoP has been used to cut-out very inclined hadronic showers.

in 2002 by the Havera Park experiment, which reported on the photon fraction limits in horizontal showers above 10^{19} eV [189].

Since then, several UHECRs observatories, including the Pierre Auger Observatory have conducted photon searches using their data.

6.2.1 Photons searches results in UHECRs observatories

Haverah Park In 2002, the Haverah Park experiment published the first limits on the photon fraction at UHE [189], using its observational dataset of inclined events ($60° < \theta < 80°$).

The collaboration obtained upper-bounds on the photon fraction by comparing the number of events above a given energy threshold, to expectation from MC-simulations, assuming a proton, iron or photon composition.
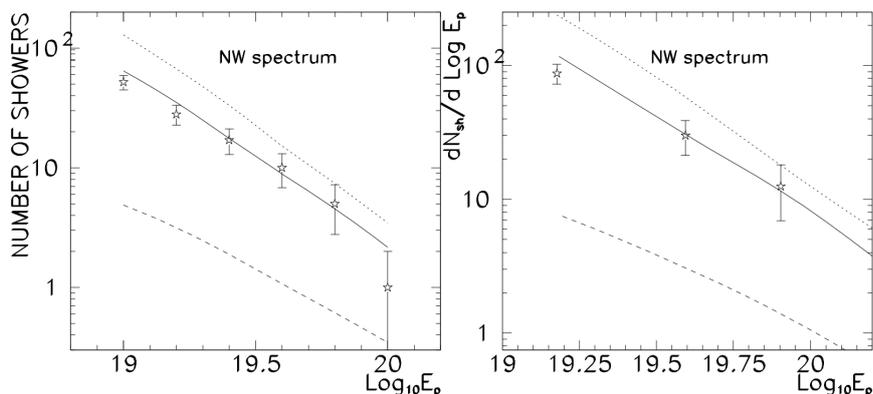
Figure 6.6: Integral (left) and differential (right) spectrum of inclined events above 10^{19} eV recorded by the Haverah Park experiment (2000). The expected spectrum for pure-composition of iron (dotted line), proton (continuous line) and photon (dashed line) are also drawn. Taken from [189]

From Figure 6.6, it was concluded that less than 48% of inclined events above 10^{19} eV can originate from photon primaries with a 95% confidence level (CL), and less than 50% with the same confidence level above 4×10^{19} eV.

AGASA The AGASA experiment reported in 2002 its own results on photon fraction upper limits [190]. The AGASA experiment had 27 muon detectors in addition to its surface array, composed of 111 stations spread over $\simeq 100$ km^2. The analysis is performed on 6 events, with energies above 1.25×10^{20} eV, and with their muon densities measured by at least 2 muon detectors. For each of the 6 events selected for the analysis, 100 photons shower were simulated, with the same primary direction as the one reconstructed for the actual event. The energy of the simulated photon-primaries were shifted up by 20% to account for the systematic underestimation of the energy of a photon-induced shower during reconstruction. Also, the simulated photon primaries energies varied from shower to shower to take into account the energy reconstruction uncertainties.

The AGASA collaboration used the muon density as the discriminant observable, by comparing the measured values to ones obtained from photon simulations, in order to assign each event a "photon-agreement" score:

$$\chi_j^2 = \frac{\left(\rho_j - \left\langle \rho_j^s \right\rangle\right)^2}{\left(\Delta\rho_j\right)^2 + \left(\Delta\rho_j^s\right)^2} \tag{6.10}$$

where ρ_j is the muon density observed in event j and $\left\langle \rho_j^s \right\rangle$ is the average muon density from photon simulations, $\Delta\rho_j$ and $\Delta\rho_j^s$ are the standard deviations of the muon densities.
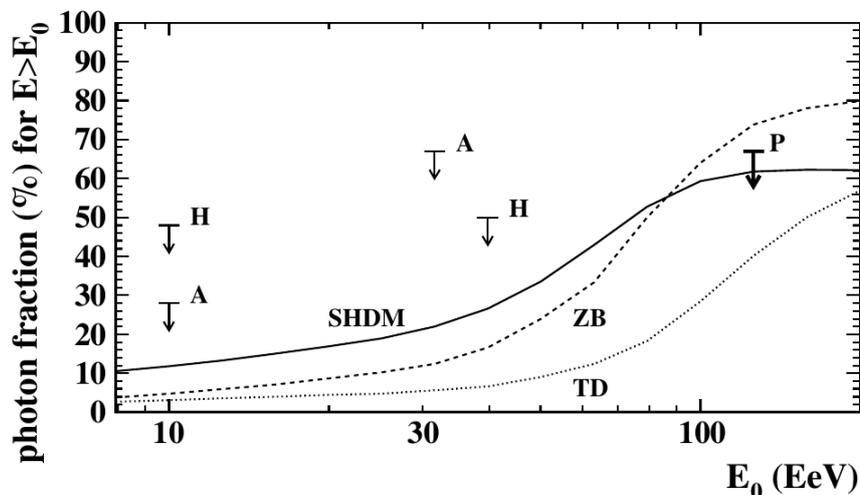
Figure 6.7: Upper limits (95% CL) on the photon fraction in cosmic-rays above $8\,\mathrm{EeV}$ produced by the AGASA (A and P arrows) and Haverah Park (H arrows) experiments along with predictions from super-heavy dark matter (SHDM), Z-burst (ZB) and topological defect (TD) models (2005). Taken from [190] and reference therein.

Using this comparison between the observed and expected muon fraction from photon simulations, it was possible to derive that the fraction of photons was below 67% in cosmic rays with energies above $1.25 \times 10^{20}\,\mathrm{eV}$ with 95% CL. This result was the first limit on the photon fraction in CR above the GZK cut-off, and it placed severe constraint on the Z-burst model. However, the spectrum from the AGASA experiment was later found out to not be compatible with the suppression found by other experiments like Hires, the PAO and TA with good statistical significance.

Yakutsk The Yakutsk experiment used a method similar to the one described for the AGASA experiment to assign a photon-probability on an event-by-event basis [191]. Instead of using the reconstructed energy of the events to parameterize their photons simulations, they used the observed signal density in the surface detectors.

The muon density at $300\,\mathrm{m}$ from the reconstructed shower core, $\rho_\mu(300)$, obtained from an empirical muon lateral distribution function is used as the discriminant variable. The analysis is then performed on an event-by-event basis to assign each shower a photon-probability p_γ.

To derive upper limits on the integral flux of photons, the Yakutsk collaboration used the relation:

$$\bar{n}(F_\gamma) = F_\gamma.A.(1 - \lambda) \tag{6.11}$$

where $\bar{n}$ is the average number of photons expected in the dataset, F_γ is the integral flux of photon primaries, A is the exposure corresponding to the dataset and λ is the efficiency, the fraction of photons "lost" (removed by selection criteria applied in the analysis).

Note that A is the effective exposure so each cut applied on the dataset during the analysis reduces this value and thus, weakens the derived upper-limits. Then, to constrain F_γ at 95% confidence level, the following relation was used:

$$\sum_n P(n)W(n,\bar{n}(F_\gamma)) < 1 - \frac{95}{100} \tag{6.12}$$
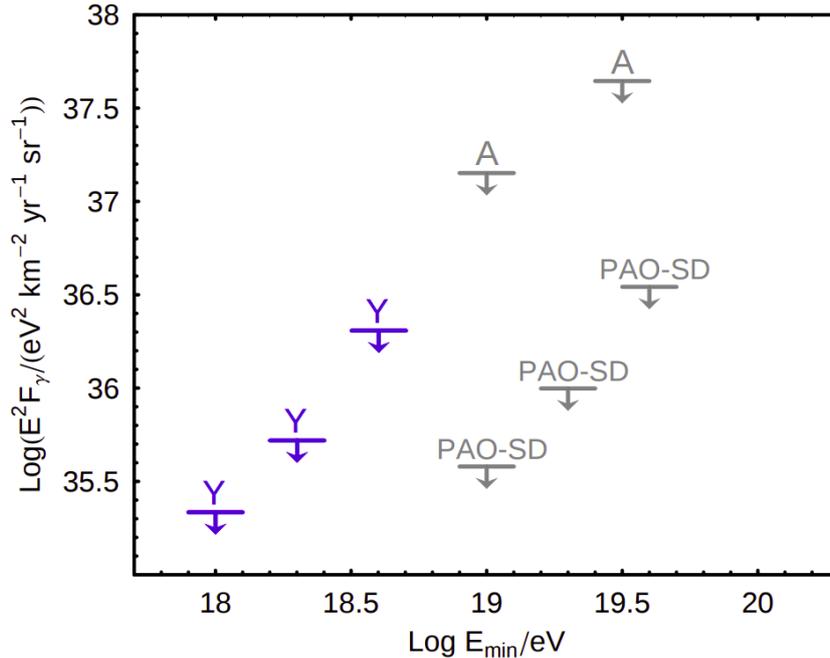
Figure 6.8: Limits (95% CL) on the differential photon flux set by the Yakutsk experiment (Y), the Pierre Auger Observatory (PAO) and the AGASA experiment (A) (results published in 2010). Taken from [191].

E_{min}	10^{18} eV	2×10^{18} eV	4×10^{18} eV
ϵ_γ	0.004	0.008	0.041
$F_\gamma \ (km^{-2}.sr^{-1}.y^{-1})$	0.22	0.13	0.13

Table 6.1: Upper-limits on the photon fraction and integral photon flux estimated above $E_{min} = 10^{18}$ eV by the Yakutsk collaboration. From [191].

where $P(n)$ is the probability of n photons being in the dataset and $W(n, \bar{n}(F_\gamma))$ is the Poisson probability of finding n photons given the average $\bar{n}$ number of photons in the dataset.

The upper-limits on the photon fraction (ϵ_γ) and the integral photon flux (F_γ) calculated for events with minimum energies E_{min} above 3 energy thresholds and with 95% confidence level are reported in table 6.1.

Telescope Array   The Telescope Array collaboration published its constraints on diffuse photons flux in 2019 [192]. In the corresponding analysis, 16 reconstructed observables were combined into a multivariate classifier. These observables are related to the discriminant features of a photon shower such as:

- the radius of curvature and width[1] of the shower front

- the shape of the lateral distribution function

- the muonic component of the showers at ground-level

Most of these variables are similar if not identical to the ones presented in the last section 6.1.

A classifier is built, using a boosted decision tree (BDT) method, combining these observables, independently for each range of energies. A look-up table was built on MC photons

---

[1]TA uses the Area-over-Peak (AoP) and a fit of its shape as discriminant parameters for photon search

E_0	$10^{18.0}$ eV	$10^{18.5}$ eV	$10^{19.0}$ eV	$10^{19.5}$ eV	$10^{20.0}$ eV
$F_\gamma\ (km^{-2}.sr^{-1}.y^{-1})$	0.067	0.012	0.0036	0.0013	0.0013

Table 6.2: 95% C.L Upper-limits on the integral photon flux calculated by the TA collaboration. (From [192])

and used to build a function E_γ to estimate the hypothetical photon-energy of the primaries, from the reconstructed energy values of the events. The analysis is performed on $E_\gamma > E_0$ for $E_0 = 10^{18.0}, 10^{18.5}, 10^{19.0}, 10^{19.5}$ and $10^{20.0}$ eV, so a BDT classifier is trained for each range of energies using MC proton as background and MC photon as signal. The classifier then assigns a value ξ to each event according to its "photon-likeness". Using the MC protons and MC photons datasets, a ξ_{cut} value is set[1], above which a real event will be classified as a photon-candidate.

The upper-limits on the photon fraction and integral photon flux of the TA analysis are calculated in the same way as the Yakutsk analysis (see equations 6.12 and 6.11). The collaboration obtains the limits shown in figure 6.9, the values for each bin of energy are reported in table 6.2.
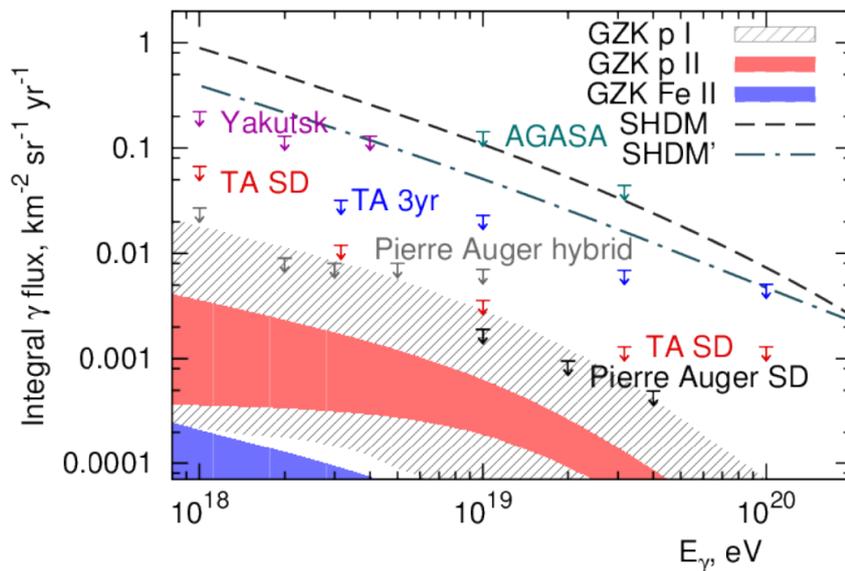
Figure 6.9: Upper limits on the integral photon flux reported by the TA experiment in 2019 (red arrows) along with the upper limits set by the PAO (SD: black arrows, hybrid: grey arrows), the AGASA experiment (light blue arrows) and Yakutsk (purple arrows). Predictions from super-heavy dark matter scenarios (SHDM/SHDM' dashed lines) and GZK scenarios are also drawn. From [192] and reference therein

6.2.2 Photons searches with the Pierre Auger Observatory

The Pierre Auger collaboration has also been conducting ultra high energy photon searches, and published limits derived on the diffuse photon flux (first publication [193] in 2008), as well as flux upper limits from photon point sources. The searches using only the surface detector of the PAO are the most directly comparable to the ones described above, we will thus start by detailing them, before going over the search conducted with the hybrid detector.

Auger SD All the SD-only photon searches have used one or several of the observables presented in 6.1. The most recent photon search conducted using the PAO data with the

---

[1]For the TA analysis, the ξ_{cut} value is zenith dependant.

SD-only uses two observables, the $\langle\Delta\rangle$-risetime, $\langle\Delta\rangle$, and the modified photon LDF, L_{LDF}. The datasets used in this analysis include: reconstructed data from the surface detector between 01/01/2014 and 30/06/2018, MC simulated showers initiated by proton primaries and photon primaries (pre-showered and non-preshowered). Several selection criteria are applied on the datasets in order to ensure the quality of the reconstruction of the analysed events:

- 6T5 flag (see subsection 4.3.1) and energy $E_\gamma > 10^{19.0}$ eV, to ensure that the trigger is 100% efficient and that the exposure does not change with the energy.

- at least 4 candidate stations with signal above 6 VEM to ensure a good reconstruction of the shower.

- zenith angle $30° < \theta < 60°$, to mainly have showers with X_{max} above ground.

The observables used in the analysis, are normalized to the expectations from the MC photon dataset. For each event with parameters (S_{1000},θ), the values of the observable x are compared to the average value expected for an event with the same parameters in the MC photon simulations dataset $\bar{x}(S_{1000},\theta)$. Both $\langle\Delta\rangle$ and L_{LDF} are thus modified into g$\langle\Delta\rangle$ and gL_{LDF} so that their distributions for the MC photon dataset are centered on 0 ($\overline{\langle\Delta\rangle}=\overline{L_{LDF}}=0$) and had standard deviation equals to 1 ($\sigma_{\langle\Delta\rangle}=\sigma_{L_{LDF}}=1$) using the relation 6.13.

$$gx = \frac{x - \bar{x}_\gamma(S_{1000},\theta)}{\sigma_x(S_{1000},\theta)} \tag{6.13}$$

This is done to have the photon median in the principal component analysis (PCA) be zenith- and energy-independent. The PCA techniques are a set of methods to explore data distributions. The core of these methods is to find the correlations between the observables, build axes on the lines representing the correlations [1] and draw the distribution of the datasets in these new axes. The simplest example of a PCA is the projection of a dataset in a 2D parameter space, where the PCA is just a rotation of the axes along the highest correlation line. The output of such an analysis is then the projected value of an event on this principal component axis. A simple example of a PCA with two random gaussian observables is shown in figure 6.10

The PCA for this analysis is set up on the training burn sample[2] and on the non-preshowered MC-photons (see figure 6.11), to calculate the principal component axis and define the photon-cut, i.e. the value on the PC axis projection above which an event is considered as a photon candidate. The photon-cut is defined as the median of the unpreshowered MC photon showers on the PC axis. The preshowered MC photon events are excluded from the observables rescaling and the PCA construction, but they are taken into account to calculate the limits on the photon flux.

The projection of the burn sample and MC-photon showers in the analysis on the PCA main axis is shown in figure 6.12. The figures shown correspond to an analysis of 2015 [194]. An update of this analysis was done in 2019 [195], with the following results: among the real events, 4 passed the photon cut above $E_\gamma > 10^{19}$ eV, 2 above 2×10^{19} eV and none above 4×10^{19} eV. These numbers of candidates are consistent with the background expectations.

Thus, upper-limits on the photon flux were derived (see 6.13), assuming, as a conservative assumption, that all the candidate events are signal and not background for each range of energy $E_\gamma > E_0$ at a 95% CL, assuming a E^{-2} spectrum (see table 6.3).

This analysis has the advantage of using rather simple observables, that are easy to study and of not relying too much on the simulations (in particular the simulated hadronic showers).

---

[1]Another way of visualizing the problem is to consider the axes as the lines around which the observables
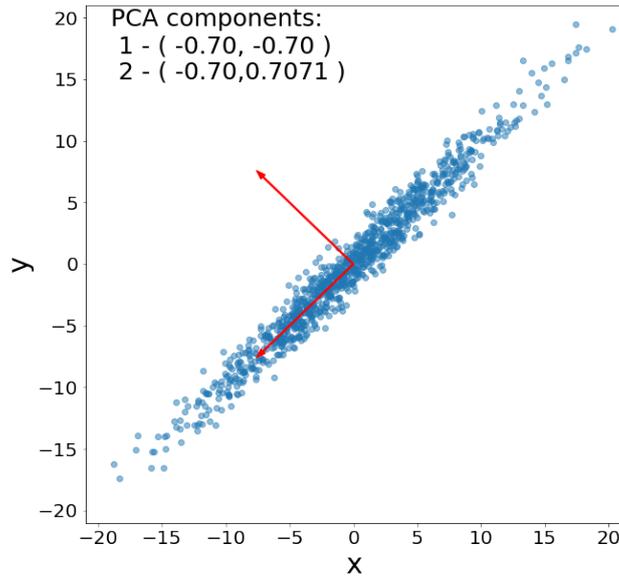
Figure 6.10: A simplified example of a principal component analysis on a 2D gaussian distribution. The blue points are randomnly drawn on a 2D gaussian distribution. The PCA applied on this distribution returns the two axes drawn as red arrows.
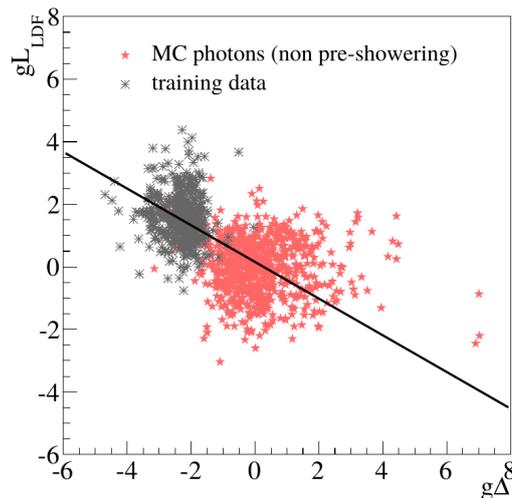
Figure 6.11: The principal component analysis of the Auger-SD photon search. In black are the burn sample events and in red the MC photons events (excluding the preshowered ones). The datasets are displayed in the rescaled variables space: on the horizontal axis, $g\langle\Delta\rangle$ and on the vertical axis, gL_{LDF}. The main axis of the PCA is drawn as the black line. Taken from [194]

The drawbacks of this method is that it requires a lot of quality selection criteria which lowers the available exposure for the analysis and that the observables are normalized with the reconstructed photon energy which is reconstructed with a bias and an uncertainty of 30%.

Auger Hybrid   As already explained in section 4.2, the Pierre Auger Observatory is an hybrid experiment, composed of a surface detector and a fluorescence detector. The dataset

---

are the less widely distributed.

^{2}the burn sample is a sample of the whole dataset used before performing the final analysis to test an already mature analysis.
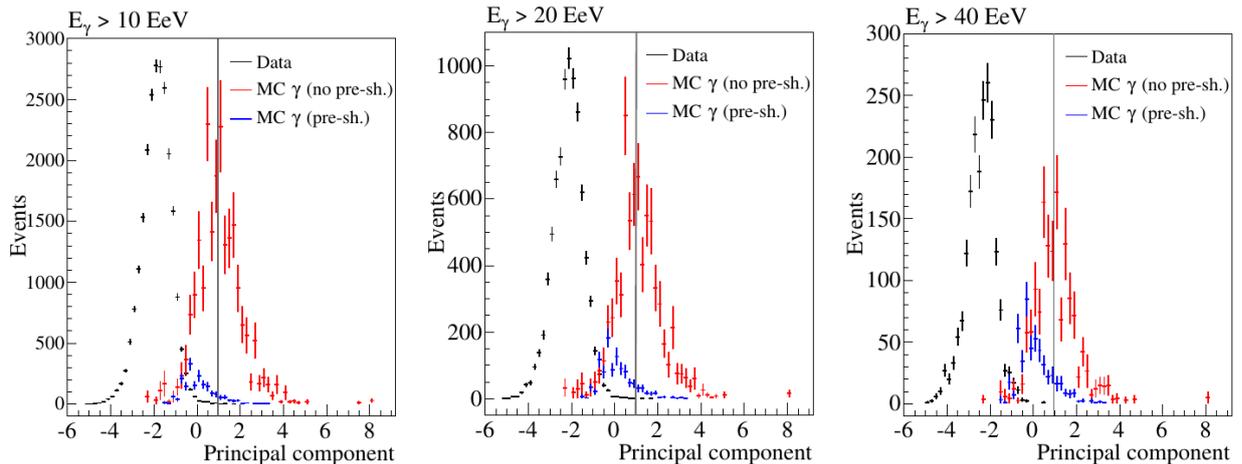
Figure 6.12: The distributions of the projection on the principal component axis of the events above $E_\gamma = 10^{19}$, 2×10^{19}, 4×10^{19} in the Auger-SD photon analysis. In black are the real showers, in red are the MC-photon showers (non-preshowered) and in blue are the preshowered MC-photon showers. Taken from [194]

E_0	$10^{19.0}$ eV	$2 \times 10^{19.0}$ eV	$4 \times 10^{19.0}$ eV
F_γ $(km^{-2}.sr^{-1}.y^{-1})$	1.9×10^{-3}	1.0×10^{-3}	0.49×10^{-3}

Table 6.3: Upper-limits on the integral photon flux calculated in the Auger-SD analysis. (From [194])

made of the events recorded with both detectors is called the hybrid dataset and is particularly interesting as it contains all the informations on its development about each shower. As mentioned in section 4.2, the hybrid dataset contains much less events than the SD one since the fluorescence detector has a lower duty-cycle and also due to quality selection criteria applied to warrant a good reconstruction of the hybrid event, especially for the measurement of the X_{max}.

The hybrid photon search uses three observables to separate p-showers and γ-showers whose discrimination power has been highlighted in section 6.1:

- the height of maximal development of the shower reconstructed by the FD, X_{max}.

- an indicator of the showers LDF, S_b.

- the number of candidate stations in the event, N_{st}.

The dataset used for the hybrid search [187] is composed of three parts: (1) reconstructed hybrid data from 01/2005 to 12/2013, MC-simulations of showers initiated by (2) photon and (3) proton primaries. A series of selection criteria are applied on the datasets to ensure that only events recorded under suitable conditions for both detectors are selected and that the reconstructed showers have a properly reconstructed geometry and profiles. Other selections are performed to make sure that the reconstructed discriminant observables are reliable. Only events with $E_\gamma > 10^{18}$ eV are selected for this analysis. In the hybrid analysis, the energy is calculated by fitting a Gaisser-Hillas profile on the measured X_{max} and then integrating over this fitted form. Then a correction is applied to account for the invisible energy. This correction is different for EM showers (1%) compared to hadronic primaries (10-15%) [196]. The energy E_γ used in the hybrid analysis is calculated using this method, under the assumption of each event being a photon (EM showers correction for invisible energy).

E_0	$10^{18.0}$ eV	$2 \times 10^{18.0}$ eV	$3 \times 10^{18.0}$ eV	$5 \times 10^{18.0}$ eV	$10^{19.0}$ eV
F_γ $(km^{-2}.sr^{-1}.y^{-1})$	0.029	0.009	0.008	0.008	0.007

Table 6.4: Upper-limits on the integral photon flux calculated in the Auger-hybrid analysis. (From [197])

The discriminating observables are then combined into a multivariate analysis (MVA) using a boosted decision tree (BDT) classification algorithm. A subsample of the MC-proton and MC-photon simulations have been used as background and signal respectively to train the classifier. To take into account the observables dependencies on energy and zenith angle, these two parameters (E, θ) are added as complementary inputs to the BDT. The events are weighted according to a E^{-2} spectrum in the analysis. Similarly to the Auger-SD analysis, the median of the MC-photons distribution on the BDT response value axis was selected as the photon cut.

This analysis was applied on recorded hybrid events in ranges of energy: $E_\gamma > 10^{18}$, 2×10^{18}, 3×10^{18}, 5×10^{18}, 10^{19} eV. In total, 3 events had BDT response values above the photon cut, making them photon candidates, all in the first energy range (10^{18} eV$<E_\gamma<2 \times 10^{18}$ eV). This is within the nuclear background estimations. The hybrid search has been extended down to lower energies (E$<10^{18}$ eV) through the use of the 750 m array and the HEAT telescope with the same discriminating observables [195]. Upper-limits on the photon flux at 95% CL are derived from the BDT response, for each energy range (see table 6.4 and figure 6.13).
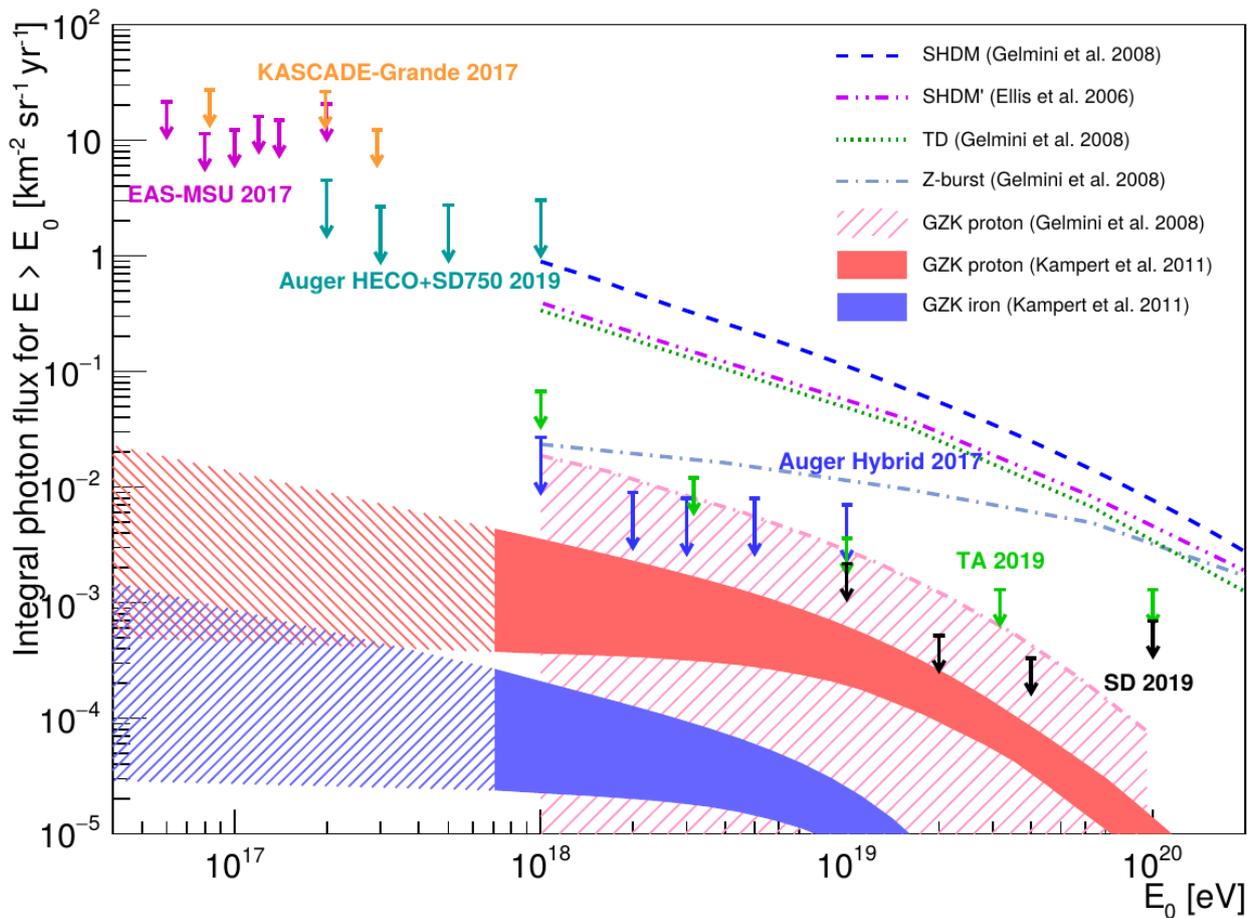
Figure 6.13: The latest results of the Auger searches (hybrid, hybrid with 750m array, SD) are shown, as well as the limits obtained by EAS-MSU [198], Kascade-Grande [199] and TA [192]). The predicted flux from top-down scenarios and GZK process are also shown (see [195] and references therein) Taken from [195]

Unfortunately, no UHE photon detection was confirmed and no evidence of GZK cosmogenic

photons has been reported so far. Thus the Pierre Auger Observatory has been able to set the lowest upper-limits to the integral photon flux over energies from 2×10^{17} eV to 10^{20} eV. The AugerPrime upgrade will improve the separation power of the 1500 m SD-array, by allowing a measurement of the muonic component of the showers.

6.3 Design of a new multivariate analysis

In the current photon searches, many quality selection criteria are applied to the data before performing the discriminant analysis. I designed a new multivariate analysis using the observables retrieved from the SD in an efficient way. The aims of this analysis are:

- to use discriminating observables based on the estimation of the muonic component and built at the station-level.

- to use a reconstruction method of the event-level observables that requires a minimal amount of quality selection criteria, leading to an improved photon efficiency.

In the first part, I am describing the observables chosen to be used in the multivariate analysis, showing their potential discriminating power. Then, I will give more technical information on the algorithms I tested and I used to obtain the discrimination between proton showers and photon showers

6.3.1 *Muonicity*

Since the muonic component is lower in a photon shower compared to a hadronic one, a procedure to evaluate the muon content in a station, called the *muonicity* method has been introduced by collaborators [200]. The procedure to determine the *muonicity* is described here.

As a secondary muon goes through a WCD, it produces a peak in the recorded signal of a station (the VEM trace). By performing a search on the traces to find such peaks, one can estimate the muonic component of an EAS. It is important to note however that high energy gammas in EAS can produce peaks indistinguishable from the ones produced by muons (see Figure 7.1).

The idealized shape of the signal produced by a muon inside a WCD is a sharp increase over one or two FADC bins, followed by a slow, quasi-exponential decay-tail. For such a signal, a normalized deconvolution filter can be applied, to isolate each peak while retaining its integral value. The deconvolution can be written as:

$$c'_k = \frac{c_k - \alpha.c_{k-1}}{1 - \alpha} \tag{6.14}$$

$$\text{where } \alpha = exp(-\Delta t/\tau) \tag{6.15}$$

where $\alpha = exp(-\Delta t/\tau)$ and τ is the decay-time of the signal inside the WCD.

When applied, on an idealized signal (a one bin signal rise followed by an exponential decrease, see figure 6.14), this filter disentangles the muon peaks and removes the decay-tail to integrate the signal into a single bin.

In reality however, the signal is altered by several features:

- the signal is not collected instantly and the rising time can cover one or two bins before the signal reaches its maximal value. From this effect, the integral signal is spread into several bins in the transformed signal.

Idealized muon signal
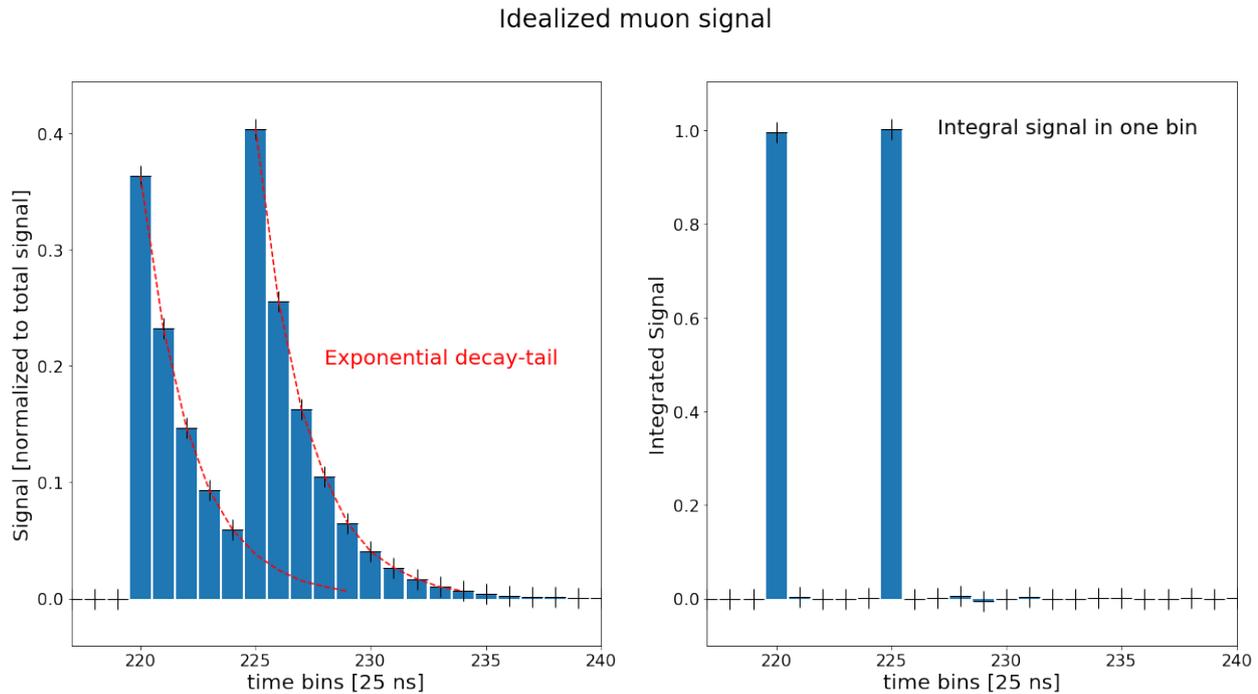
Figure 6.14: Idealized signal from a muon going through the WCD (left) and the same signal with the deconvolution filter applied (right).

- the VEM trace has noise fluctuations especially around the peaks, which are reflected in the transformed signal.

- if the decay-time τ is under-(over-)estimated, the transformed signal will have a positive (negative) decreasing exponential tail.

The search for muon peaks that is performed on the deconvoluted trace must take those 3 features into account.

The deconvolution filter is applied on the averaged VEM trace and an iterative process is started on the time-bins (k) of the transformed signal:

- each bin k is counted as a peak-candidate only if its signal c'_k is above a minimum signal S_{min}. This limits the number of peak-candidates due to noise fluctuation as a peak-candidate.
  (1): $c'_k > S_{min}$

- a peak-candidate is a peak only if its signal c'_k is larger than the values of the adjacent bins. That way only the peak with the highest value is counted in the case of a cluster of bins due to a peak signal not properly integrated into a single time bin.
  (2): $c'_k > c'_{k-1}$ **and** $c'_k > c'_{k+1}$

- the signal S_k of a peak, is the sum of the bin signal c'_k and the highest adjacent signal value $max(c'_{k-1},c'_{k+1})$ to recover part of the integrated signal that spills into adjacent bins due to the features mentioned before
  (3): $S_k = c'_k + max(c'_{k-1},c'_{k+1})$

- the peaks signal S_k values of the VEM trace are summed into the station variable S_{peaks} if they are above a threshold value S_{thr}.

The normalized distributions of the values S_k for MC protons and MC photons for different zenith angle ranges are compared in figure 6.15. A muon hump is visible in the MC proton distribution, around 1 VEM and drifting upward with increasing zenith angle. These humps are due to the muons contribution, and as less muons are expected in EM showers, the MC photon distribution doesn't exhibit such behavior. Also, the hump for more vertical showers is around 1 VEM, as expected since the VEM is defined as the signal produced by a through-going muon hitting the station vertically.
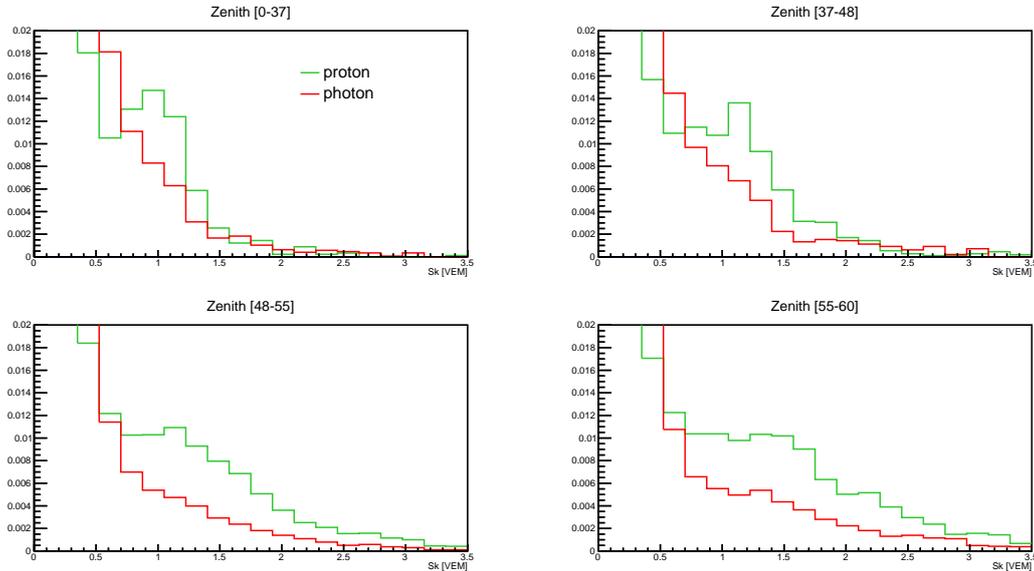
Figure 6.15: Normalized distributions of the S_k values calculated for MC-protons (in blue) and MC-photons (in red) for showers with S_{1000} values between 10 VEM and 20 VEM for different ranges of zenith angle.

With this algorithm applied, one S_{peaks} value is assigned to each station. For a shower with given parameters (S_{1000}, θ), a station at a given distance r from the reconstructed shower axis will have, on average, a smaller S_{peaks} value if the primary is a photon: $S_{peaks}^{pr} > S_{peaks}^{ph}$

6.3.2 Smoothing

In the last section, the "spikiness" of the VEM traces was used to estimate the muonic component of a shower. Flip the problem and it becomes obvious that the "smoothness" of a trace can also be an indicator of the muonic ratio of an event. The *smoothing* method has been developed by collaborators [201] and further developments have been performed on it [202]. The original method is used in this work and it is described here.

The aim of the *smoothing* procedure is to separate, inside the VEM trace, the contribution to the signal coming from the electromagnetic and muonic component of the shower by applying a moving average filter.

The following steps are applied on the averaged VEM trace:

- for each time bin k, measure the smoothed signal c_k^{sm}, calculated as the normalized integral of the VEM trace on a $2.N_{bin} + 1$ window centered on bin k.

  (1): $c_k^{sm} = \sum_{i=k-N_{bin}}^{k+N_{bin}} \frac{c_k}{2.N_{bin}+1}$

- if the smoothed signal c_k^{sm} is inferior to the original signal S_k, the bin is replaced by the smoothed one in the VEM trace.

  (2): if $c_k^{sm} < c_k$, then $c_k = c_k^{sm}$
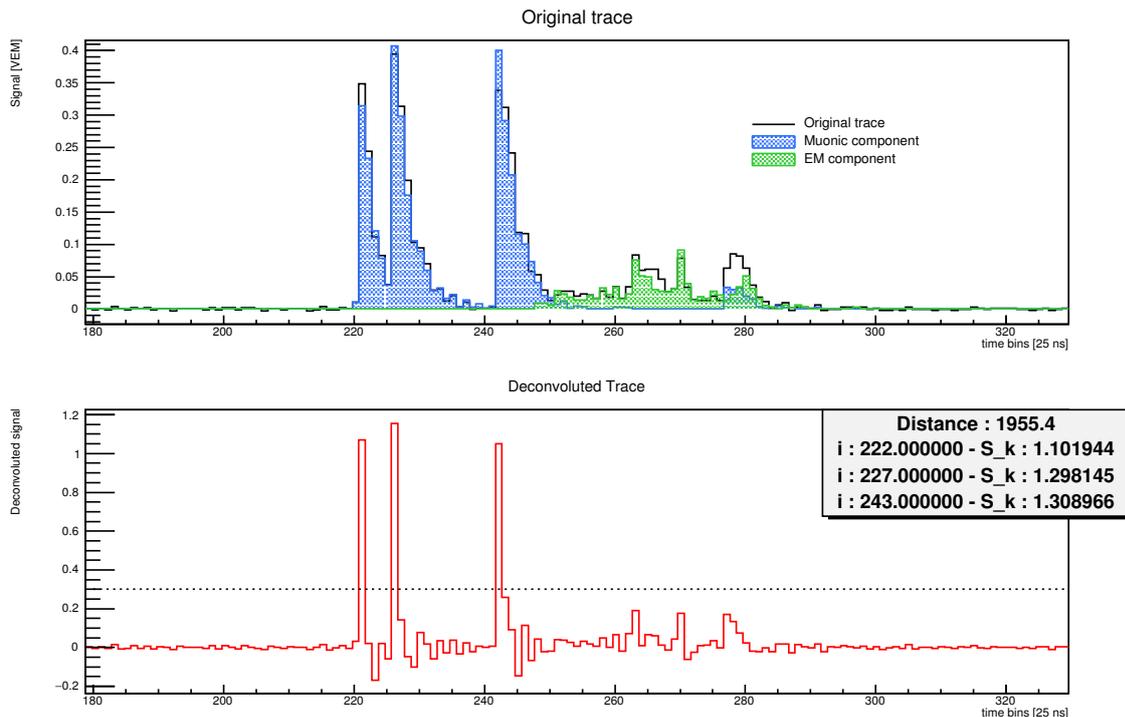
Figure 6.16: The *muonicity* process applied on a simulated VEM trace, with the muonic and EM components highlighted in different colors.

- once each bin has passed the first two steps, the process restarts, for N_{iter} iterations, each time applied on the newly smoothed trace.
  (3): $c_k \rightarrow (N_{iter}) \times (c_k(\mathbf{1})(\mathbf{2})) \rightarrow (c_k^{sm})_{N_{iter}}$

- the total smoothed signal S_{tot}^{sm} is then normalized by the total signal S_{tot} to get $1 - f_\mu$.
  (4): $f_{EM} = 1 - f_\mu = \sum\limits_k \left(\frac{c_k^{sm}}{c_k} \right) = \left(\frac{S_{tot}^{sm}}{S_{tot}} \right)$

The f_μ value is an estimation of the fraction of the muonic component in the total contributions to the signal of a station. This variable is expected, for a given shower (S_{1000}, θ) and a given station (r) to be on average lower for a photon-initiated shower.

Obviously, the observables obtained from the *muonicity* and the *smoothing* methods are expected to be correlated. However, the two observables are complementary. Indeed they are most effective on different ranges of distance: *muonicity* at larger distances and *smoothing* closer to the core. Also, with both methods, a precise measurement of the muonic signal inside the traces is not possible. Instead, the aim is to extract a correlation between the observables and the muonic signal. Thus, combining two estimators of the muonic signals provides a better estimation overall. In practice, it has been verified that the separation power is better with both observables instead of just one.

6.3.3 Building the multi-variate analysis

Through the rest of this work, the term **station-level variables** is used to call the variables sensitive to the proton/photon discrimination assigned to each **station**. The term **event-level observables** is referring to the observables that are obtained by combining the **station-level variables** of each event into one single value.

Through the *muonicity* and *smoothing* method, variables (S_{peaks} and f_μ respectively) are calculated and assigned to each station selected for the analysis. A standard way of combining
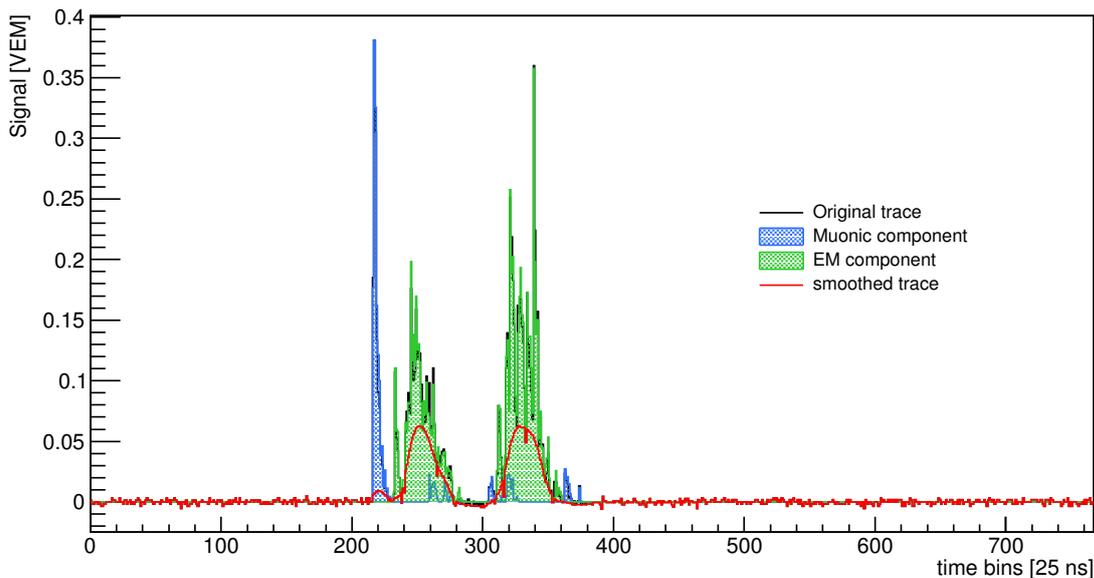
Figure 6.17: The *smoothing* process applied on a simulated VEM trace, with the muonic and EM components highlighted in different colors.

station-level variables into event-level observables is through the use of a primary-specific function or benchmark (see 6.1.3). The function or benchmark is fitted (trained) on a specific dataset, either signal or background (tagged[1]) to make a prediction of the discriminant variable value given the event and station parameters. Then, the predictions are compared to the calculated values of the station-level variable and the differences (or another combination of the two values) are summed into an event-level observable.

In this work, a similar approach was selected, using randomized decision trees based algorithms (see paragraph Decision Tree) to make the discriminant station-level variables predictions. For both S_{peaks} and f_μ, the variables used to make the predictions are:

- the station distance from the shower core r

- the zenith angle of the shower θ

- the expected signal at $1000\,\text{m}$ from the shower core S_{1000}

The azimuth angle was tried as an additional parameter to predict the variables but no significant improvement in the prediction precision was found. A model is trained for each station-level variable $X(r, \theta, S_{1000})$ using MC-proton simulated events to predict the value X_{pr}^{pred}. The dataset used to train the models consist of 2/3 of the stations, selected randomly. The remaining 1/3 is used to validate the prediction.

Instead of building a benchmark based on a parameterized function to use as a model to predict the variable values, a machine learning algorithm was chosen, to decrease the number of quality cuts on the training datasets. To explain the considerations that went into choosing the algorithms to build the models, some basic notions of Machine Learning need to be introduced.

---

[1]sometimes untagged data can directly be used to train the function or benchmark, under the assumption that the potential signal is too rare to impact the fitting process.

Terminology

- scaler: a scaler is a tool used to rescale a dataset, upstream of a machine learning (ML) algorithm, for example by centering the distribution on 0 and normalizing the variation to 1. It can also reverse the scaling downstream of the algorithm. Most ML algorithms work better with scaled variables.

- supervized/unsupervized machine learning: when the behaviours of the variables chosen to describe the dataset are sufficiently understood and some rules or laws are applied to guide the machine learning, it classifies as supervized ML. Sometimes however, one might just want to try to classify a dataset by finding correlations, relations between features (clustering) or explore the dataset based on statistical considerations (visualization or dimensionality reduction). In these cases, it is classified as unsupervized ML. In this work, supervized algorithms are used.

- classifier/regressor: when the aim of the model is to classify the events into a set of labeled categories, the model is called a classifier (discrete output values). If the aim is to predict a quantity based on inputs, it is a regressor (continuous output values).

- loss and cost function: the loss function is used to estimate the error on the prediction of each data point in the training sample. The cost function is the combination of these losses over the whole training dataset. Many learning methods perform an optimization of a cost function over a training dataset

- hyperparameters: each kind of machine learning algorithm has a set of options that can be modified by the user. They are generically called the hyperparameters of a model.

Decision Tree

The decision tree algorithm is an extension of sequential selection criteria analyses, that trains a model using recursive partionnning on a dataset. Lets take a simple example: Building a binary decision tree for classification on a dataset *data* with events tagged as signal s or background b, and parameters ($x1,x2,x3$) can be separated in steps:

- layout the *data*, the *trunk*, all events and all parameters ($x1, x2, x3$) and find the cut that provides the best separation between s and b. For example, let's say $x1 > 0.5$ is the single cut that separates best s and b for the whole dataset.

- this cut becomes a node, and two subsets of *data*, *branches*, are created: *dataA* with all events where $x1 > 0.5$ and *dataB*, with all events that passed the cut.

- the process is now applied once again on both subsets. The best separation cut found can be on a parameter on which a selection has already been performed earlier in the tree.

- once a branch reaches a termination criteria, it becomes a *leaf*. Each node has a purity score p, calculated using an impurity function, in this example, the ratio between the number of events tagged as s in the subset and the total number of events in the subset: $p = \frac{s}{s+b}$. A purity score close to 1 means the events in this leaf are more likely to be signal, and the more likely to be background if p is closer to 0.

- the termination criteria must be setup before training the decision tree and are usually: minimum leaf size, tree depth, perfect classification or no improvement from further splitting.

Once the tree is trained on a tagged dataset (s,b), events from another dataset, not used for training, can be sent through the tree to be assigned a purity value p. If p is above a given value (e.g. 0.5 in this example), the event is tagged as signal, otherwise it is tagged as background.
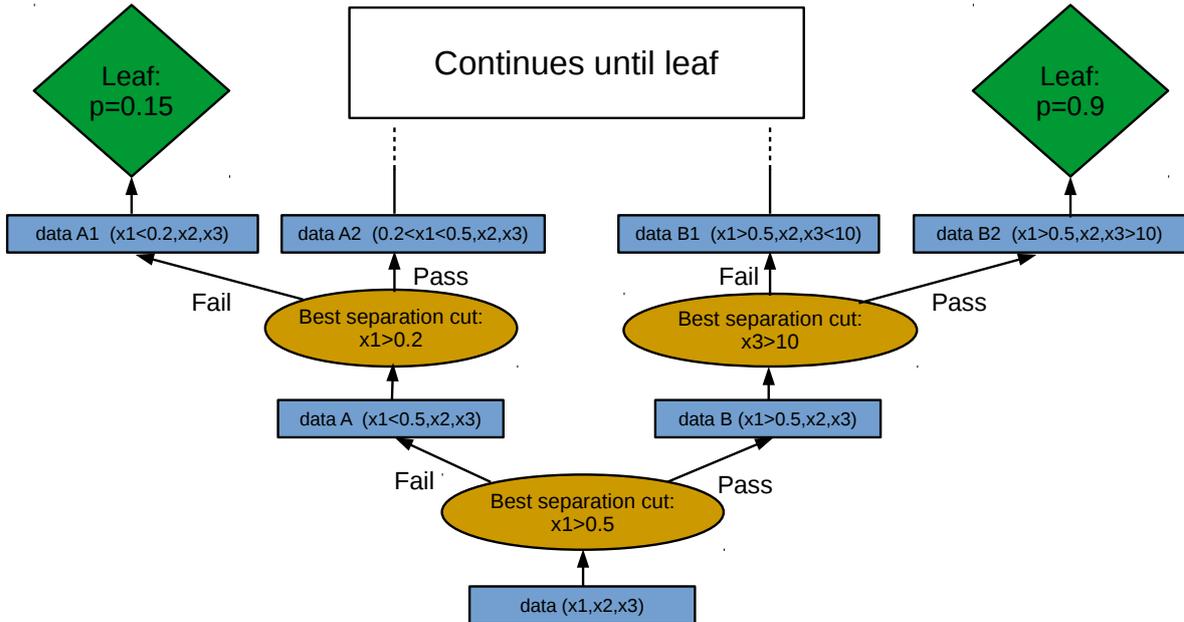
Figure 6.18: Simplified scheme of a decision tree training.

This is the simplest example of a binary decision tree for classification, but all subsequent types of decision tree algorithm rely on the same principles. To train a decision tree for regression, the main change is the impurity function used to define the cut in each node. Where for a classification problem, the separation is done on discrete values (i.e signal or background), in a regression problem, the values are continuous and the criteria to split nodes must thus be different. A common choice of impurity function for a regression tree node is the weighted mean squared error (MSE). It writes as:

$$MSE(k) = \frac{1}{N_k} \sum_{i \in k} (y^i - \hat{y}_k)^2 \tag{6.16}$$

where N_k is the number of events at node k, y^i is the prediction for event i at node k and $\hat{y}_k$ is the mean value of the predicted values for all events in node k.

ExtraTree Classifier and Regressor

Beyond the simple decision trees, a more advanced multivariate analysis technique commonly used in particle physics are the boosted decision trees (BDT). In BDTs instead of training a single decision tree for a long time and trying to have it return the highest purity scores possible, the aim is to combine weaker trees together. The way the weaker trees are combined

in BDT is called the boosting and many boosting algorithms exists (adaptive boosting, gradient boosting,...).

The ExtraTrees algorithm is a type of randomized decision tree algorithm, where each tree is trained on a subset of data and either all or part of the features. The subset of data or features are created using bootstrapping techniques. All the trees are then combined into a single estimator by averaging their predictions. Combining shallower trees reduces overfitting, as single decision trees have a tendency to overfit, and it also reduces variances, as the errors of each trees are somewhat decoupled since they are trained randomly and individually. In the Extra Tree Classifier/Regressor algorithms, the node splitting threshold for each feature is first drawn randomly and then, the most discriminant one is selected from those randomnly drawn thresholds. In the implementation of these algorithms, each tree is called an estimator and the number of estimators ($n_{estimators}$) is one of the most important hyperparameter to configure for this work's analysis.

NearestNeighbors

Nearest Neighbors algorithms are fairly simple prediction algorithms, where the target quantity (labels in classification, continuous values in regression) is estimated from the values of the datapoints "closest" to it inside the training subset. The distance between datapoints is the standard Euclidian distance in the parameter space. The number of datapoints used to make the prediction, as well as the weights each datapoint is asigned, are two of the hyperparameters used to configure the model.

Although fairly simple, these kinds of algorithms can be really effective for large datasets under two conditions:

- the dataset has as a "compact" parameter space, where there are enough neighbors to make a reliable prediction for each datapoint.

- for regression, the values y to predict have continuous distribution.

The simplicity of this type of algorithms also allows for quick training and prediction relative to other, more complex algorithms.

For each variable var, a model is trained on MC proton simulation to predict the value var_{pred} from a set of parameters (r,θ,S_{1000}). The event-level observable is then calculated for each event as:

$$obs = \frac{1}{N_{st}} \cdot \sum_{i \in N_{st}} \frac{var}{var_{pred}} \tag{6.17}$$

Using this relation, the station-level variables (var) S_{peaks} and f_μ become the event-level observables (obs) *Muonicity* and *Smoothing*.

In this chapter, the main observables used for photon discrimination in UHECRs data were presented, as well as the searches for UHE photons in EAS observatories. For this thesis work, the emphasis is put on using the muonic component inside the SD traces. To that end, the *muonicity* and *smoothing* methods are used to build station-level variables that are correlated to the muonic signal. To reconstruct event-level observables from the station-level variables, ML regressors are used to emulate a proton-benchmark, making predictions of the variables values. These proton-like predictions are then compared to the calculated values of the variables and combined into event-level observables. Since this method can be applied to any station-level variable that is correlated to the mass-composition of UHECRs, the

stations risetime values $(t_{1/2})$ are also used in the analysis, to reconstruct an event-level *RiseTime* value. A multivariate analysis was built, combining three discriminant observables: *Muonicity*, *Smoothing* and *RiseTime*. In the following chapter, the results of this analysis are presented.

Chapter 7

Data Analysis

Contents

In this chapter, studies of the station-level variables presented in the last chapter are presented, along with their behaviour. Afterwards, the event-level observables built from these station-level variables are studied and the multi-variate analysis combining them for proton/photon discrimination is introduced. The results of this proton/photon discrimination, applied on MC simulations are first presented, before the same analysis is applied on a fraction of real data. Finally, an extrapolation of these results in terms of photon-flux limits and perspectives for photon searches in UHECRs data are presented.

7.1 Presentation of the datasets and tools

Before delving into the analysis, this section (re-)introduces the lexicon and presents the datasets used in this analysis.

- the station-level variables are the discriminant variables calculated/measured on individual stations.

- the event-level observables are the discriminant observables calculated by combining the station-level variables.

- a proton(/photon) shower is an extensive air shower (EAS) initiated by a proton (/photon) primary.

There are three datasets used in the multi-variate analysis (MVA) that was designed:

- proton showers reconstructed from simulations (MC-protons)

- photon showers reconstructed from simulations (MC-photons)

- real data reconstructed with the Surface Detector (SD) of the Pierre Auger Observatory (PAO)

The reconstructed simulated showers are produced using:

- shower simulation: CORSIKAv75600

- high-energy hadronic interactions: EPOS-LHC

- low-energy hadronic interactions: Fluka (v. 2011.2c)

- thinning: $1e^{-6}$

- slope of the energy distribution: -1

- energy range: $18.5 <\log(E/eV)< 20.5$

- zenith angle: $0° < \theta < 70°$

- number of events: 150k photons, 42k protons

- preshower option on for photons

- detector simulation and reconstruction: OFFLINE trunk version 31300

The real events dataset is a burn sample of around 2% of all the events reconstructed with only the SD of the Pierre Auger Observatory between 01/01/2004 until 31/12/2017. The burn sample is actually a concatenation of two parts:

- Part A: SD data from 01.01.2004 until 14.05.2013, random extraction of events from a pre-selected sample (pre-selection as used for the Auger SD spectrum presented at the ICRC 2013: T4 and 6T5 required, reconstructed zenith angle $< 60°$, reconstructed LDF available, reconstructed energy > 3 EeV) - 54490 events

- Part B: SD data from 15.05.2013 until 31.12.2017 with the events selected so that their SDId number is a multiple of 50 - 45809 events

To have a homogeneous dataset, it was thus decided to apply the same cuts to Part B of the burn sample as well.

The simulations available for this analysis contained events with reconstructed energies between $10^{18.5}$ eV to $10^{20.5}$ eV, so all real events with reconstructed energies below $10^{18.5}$ eV were discarded from the analysis. This is the main cut applied on the burn sample as it removes $\sim 97\%$ of the events in the burn sample.

The showers are reconstructed using the vertical reconstruction method of the OFFLINE so only events with zenith angle $0° < \theta < 60°$ are selected. The cuts applied on the data are also applied on the simulations for consistency.

The shower variables from the Offline software and the *Muonicity* and *Smoothing* methods have been developed by collaborators. In addition, I fully designed the analysis presented in this chapter. The regression and classification algorithms used are from the scikit-learn Python module [203].

7.2   Study of the variables and parameters

7.2.1   Parameters and station-level variables study

For both S_{peaks} and f_μ, determined respectively by the *Muonicity* and the *Smoothing* methods (see subsection 6.3.1 and subsection 6.3.2), the initial motivation was to use them to estimate the fraction of the signals due to the muonic component of the showers. The most relevant way to assess the quality of the station-level observables is thus to compare the values of S_{peaks} and f_μ to the "true" values of the muonic signals obtained in MC-simulations. It is important to keep in mind though that those two variables provide at best a rough estimation of the muonic signal, and the actual goal is to have a correlation between them and the muon signal. Indeed, both methods (*Muonicity* and *Smoothing*) scan the "topology" of the VEM traces to find features that are associated with either the muonic or the EM component.

The two variables cannot perfectly separate the muonic from the EM component because of multiple phenomena:

- high energy gammas can produce peaks in the VEM traces indistinguishable from muon peaks (see bottom figure in figure 7.1).

- if too many particles produce a signal in a trace inside a few ~ 25 ns time-windows, it becomes impossible to distinguish each particle contribution (see top figure in figure 7.1).

The calculation of S_{peaks} and f_μ depends on parameters S_{min}, S_{thr} (see subsection 6.3.1) and N_{bin}, N_{iter} (see subsection 6.3.2) respectively. A tuning of these parameters was thus performed, to select the best values of the parameters, based on the correlation factor of S_{peaks} and f_μ and the relative difference between the variables and the true muon component in the traces, using MC-simulations. The selection process also took into account the features underlined in previous studies done in the collaboration. The values that optimize the correlation factors and at the same time produce coherent values of the variables are $S_{min} = 0.3$, $S_{thr} = 0.7$, $N_{bin} = 8$ and $N_{iter} = 4$.

S_{peaks} and f_μ are based on features sensitive to the muonic content inside the shower. However, the time-spread of the shower front has been underlined in subsection 6.1.3 as a good discriminant variable and it seemed worthwhile to add a variable based on it in the analysis. The *RiseTime* was thus added as a supplementary discriminant variable to the
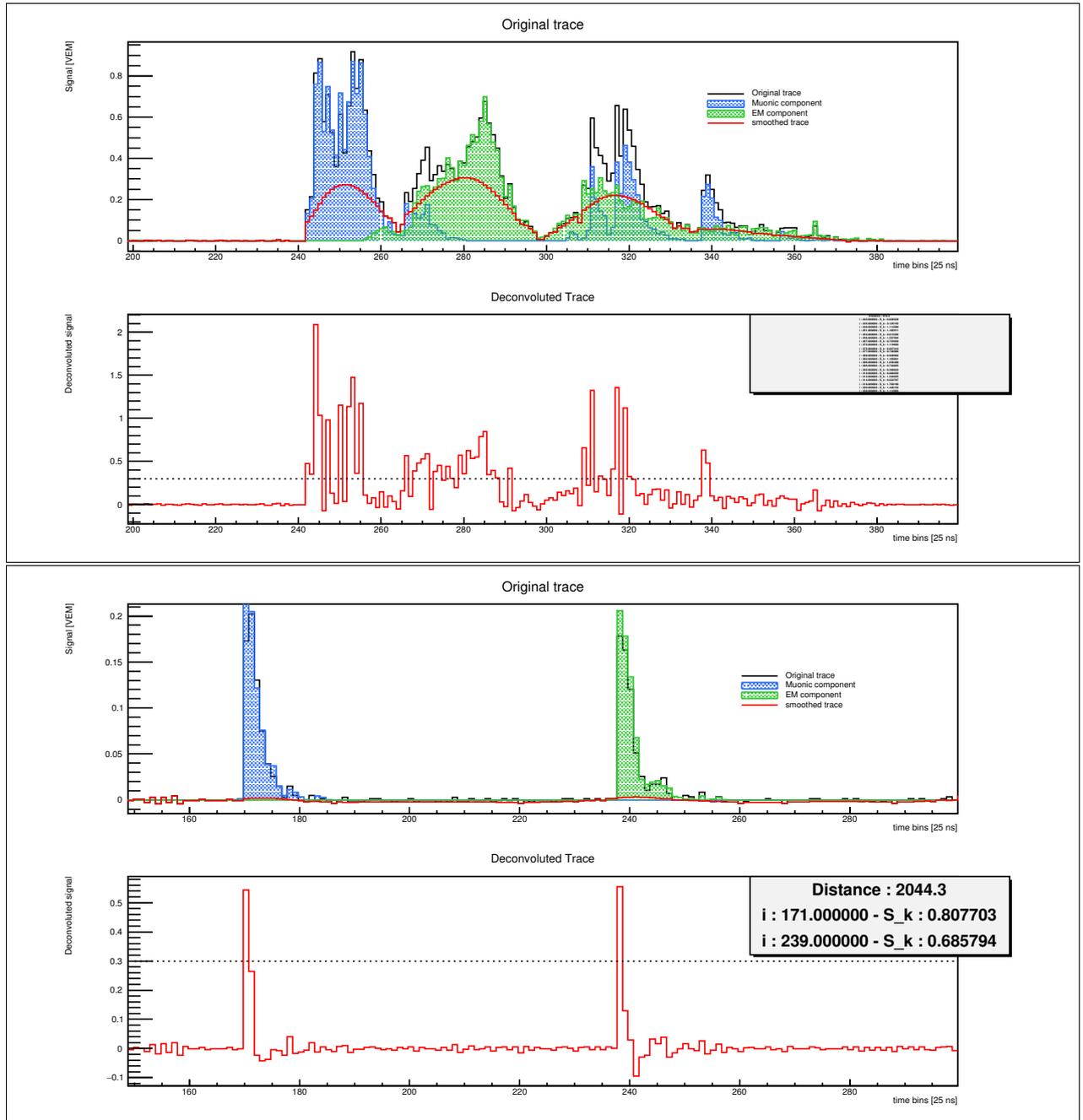
Figure 7.1: Examples of "*bad*" traces for the *Muonicity* and *Smoothing* methods. In both frames, the original VEM traces are shown above the deconvoluted traces, calculated using the *Muonicity* method, while the *smoothed* traces (calculated according to the *Smoothing* method) are drawn over the VEM traces, in red. The dotted line in the deconvoluted traces represent the S_{min} value above which a bin is considered a peak by the *Muonicity* method. The frame on the top-right of the deconvoluted traces indicates the bin-number and the S_k value of the deconvoluted peaks selected by the *Muonicity* method. Top: the VEM trace shows a *crowded* trace, where the muon signals are bundled together and the *Muonicity* method is unable to properly separate the muon peaks. Bottom: in this VEM trace, around bin 240, a photon produces a "*muon-like*" peak, and both the *Muonicity* and the *Smoothing* methods treat it as a muon signal. One can see that for the top figure, the method does not work as it finds too many peaks compared to the true number of muon peaks.

analysis, after the whole analysis-framework was built using S_{peaks} and f_μ. The goal was to "plug" this new variable to the existing analysis, thus, no specific study was performed here on the definition/calculation of the *RiseTime* value. In the rest of this chapter, when the station-level variables are defined as "calculated" (var_{calc}) for the *RiseTime* it refers to the

value given in the reconstructed shower files.

Parameter space of the analysis

The relevant parameter space of the stations (distance of the station to the shower core in the shower plane r, total signal in VEM S_{tot}, expected signal at $r = 1000\,$m from the LDF-fit S_{1000} and zenith angle θ) was scanned, to only include in the analysis stations in which the variables S_{peaks} and f_μ had a correlation with the muonic component. For the $RiseTime$, a correlation is expected at the event-level with the X_{max} of the event. The parameter space on which the analysis is performed is selected to optimize the correlation of the variables with the muonic signal. The three discriminant observables were then carefully studied to find the appropriate cuts which balance the optimal parameter space for each of them. From this study, the following cuts were selected to be applied, prior to the analysis:

- θ: 0° to 60°, because the parameters used are based on the vertical reconstruction of the showers, and it is only valid for this range of zenith angles.

- no criteria to be applied on S_{1000} or S_{tot}.

- r: only stations between $500\,$m and $3500\,$m from the shower core were selected for the study. This is a conservative choice, made so that all the stations used are within a parameter-space where all the variables are properly reconstructed.

Although no further quality selection criteria are applied on θ and S_{1000}, the separation is expected to be better at higher θ and at higher S_{1000}, because more stations are triggered for such showers and because for inclined showers the muon fraction is more important.

7.2.2 Station-level variables predictions

With these variables S_{peaks}, f_μ and $RiseTime$ reconstructed for each station in the selected parameter space, machine learning algorithms are used to predict the station-level variables as a function of the parameters (r, θ, S_{1000}) of each station.

Two types of regression algorithms are studied, to search for the optimal hyperparameters and test the quality of their predictions on simulated proton and photon showers.

Choosing the hyperparameters of the regression algorithm

Algorithms from the $scikit - learn$ module were tested to build the best predictive model for the station-level variables. Many stations are available to train the models and the predictions are made based on three parameters (r, θ, S_{1000}), thus two algorithms proficient for problems with large training datasets were tested, the $ExtraTreesRegressor$ (ETR) and the $KNearestRegressor$ (KNR) (based on the Nearest Neighbors method) (see subsection 6.3.3).

Prior to any training or prediction, the datasets are always standardized using a scaler, as the estimators (classifiers and regressors) perform better on normally distributed input variables. The scaler chosen for this analysis centers each input variable distribution on the median and scale the spread with the interquantile range. Using the median and interquantile range, instead of the mean and variance (as one could expect to reproduce a normal distribution) allows a better handling of outliers. The scaler object remembers the transformation and does the reverse operation on the model predictions.

Each algorithm was applied on the MC-protons to test the quality of the predictions and select the best one. The scorer, to quantify the quality of the prediction is the coefficient of determination $r2$. Intuitively, $r2$ quantifies the proportion of variance of the prediction target that can be derived from the training parameters. A model that fully describes the prediction targets from the parameters will get a score of 1, a model that always guesses the right values but does not vary with the input features will get a score of 0.

Given y the prediction target, $\widehat{y}$ the prediction from the model and $\mathbf{y}$ the mean value $\sum_i y_i/n$, $r2$ is computed as (i being one of n datapoints):

$$r2(y,\widehat{y}) = 1 - \frac{\sum\limits_{i=1}^{n} y_i - \widehat{y_i}}{\sum\limits_{i=1}^{n} y_i - \mathbf{y}} \tag{7.1}$$

For both ETR and KNR algorithms, many hyperparameters can be changed to optimize the model. The most notable choice of hyperparameters are:

- the weighted mean squared error is chosen as the criteria used to perform the leaf split in ETR algorithms;

- the number of estimators combined to make a single prediction was scanned through to find the optimal value for each variable between $n_{estimators}$=[1, 2, 3, 4, 5, 10, 20, 30, 40, 50, 80, 90, 100, 110, 120, 150, 200, 300];

- for a given datapoint A to predict, the weight assigned to each neighbor K is chosen to be inversely proportionnal to the distance between K and A, in the parameter space;

- the number of neighbors used to make a single prediction was scanned through, to find the optimal value for each variable between $n_{neighbors}$=[1, 2, 3, 4, 5, 6, 10, 20, 30, 50, 80, 90, 100, 110, 120, 150, 200, 300].

The effects of the hyperparameters $n_{estimators}$ and $n_{neighbors}$ are very similar, as they both define the number of individual estimators being combined for each datapoint. It thus makes sense to directly compare the evolution of the performances of each model along the values of these two parameters, as can be seen in figure 7.2.

Both kinds of algorithms were tested extensively, for each station-level variable, by performing a study over a grid of hyperparameters: each set of hyperparameters is cross-validated by splitting the training dataset into 3 subsamples and training a model on each combination of 2 subsamples (see Figure 7.3). The performance of the trained models for each set of hyperparameters are evaluated on the averaged values from the grid search.

Prediction quality of the algorithms

A $KNeighborRegressor$ was trained for each of S_{peaks}, f_μ and $RiseTime$, with values of $n_{neighbors}$= 30, 100 and 200 respectively. The same was done for $ExtraTreesRegressor$ models, with values $n_{estimators}$=50, 50, 50. The quality of the station-level variables predictions was evaluated with the averaged score "$r2$" over the 3 models trained on the folded training dataset.

To ensure that the predicted variables behaved in the same way as the calculated ones, the distributions of both over the parameters space were always plotted and checked (see an example in Figure 7.4).
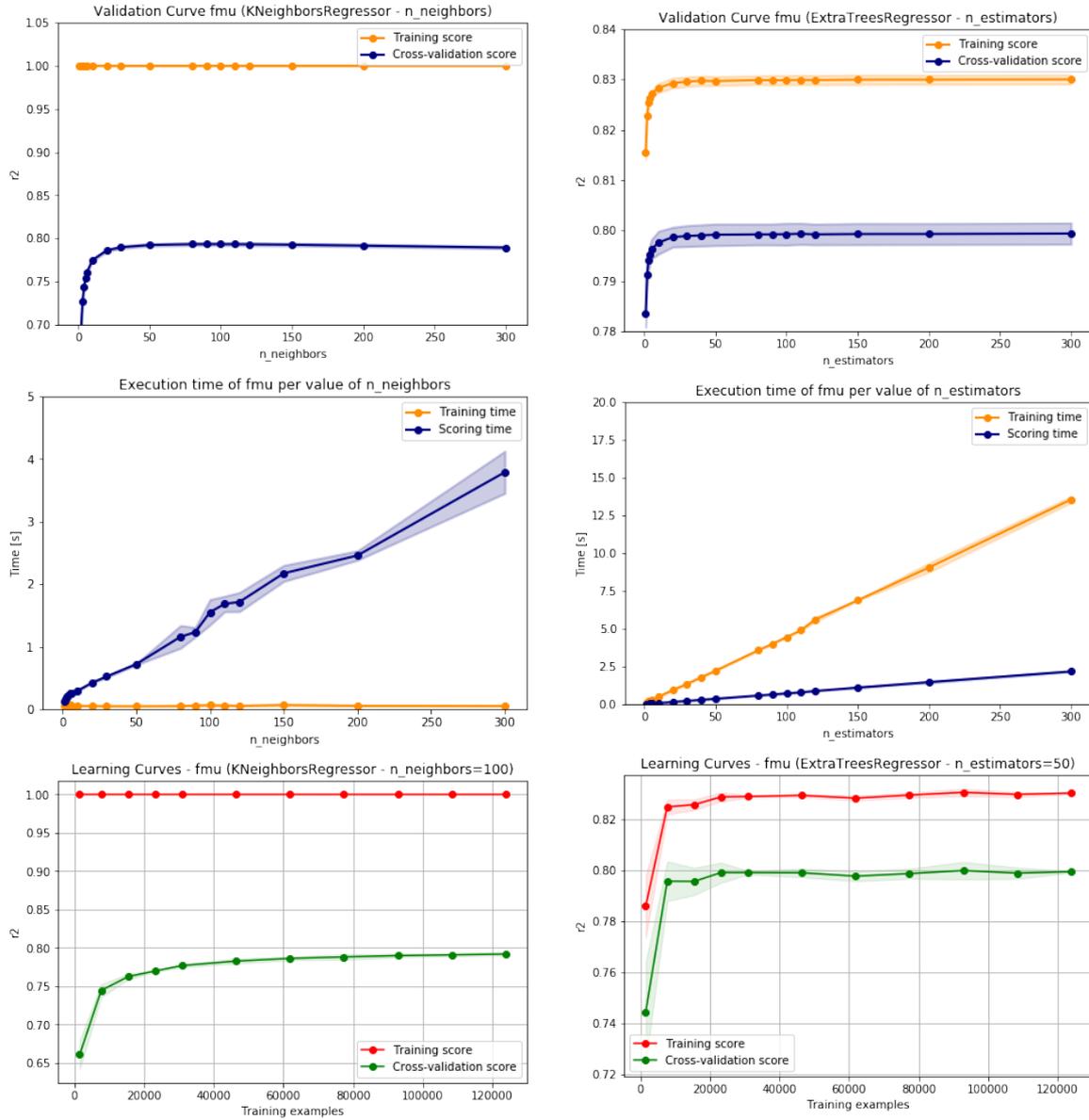
Figure 7.2: The curves used to compare the *KNearestRegressor* algorithm (on the left) with the *ExtraTreesRegressor* algorithm (on the right) to select the best regressor for f_μ predictions. From top to bottom: (1) the validation curves of the score value $r2$ of the models as a function of the hyperparameter values, (2) the execution time needed to train and score with the models as a function of the hyperparameter values and (3) the learning curve, the score value of the model with increasing number of samples.
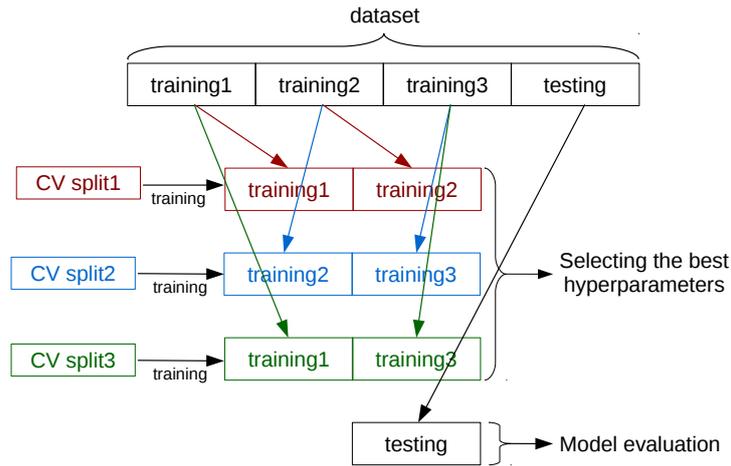
Figure 7.3: Scheme of the cross-validation (CV) process performed on the ML-models to select the best hyperparameters.
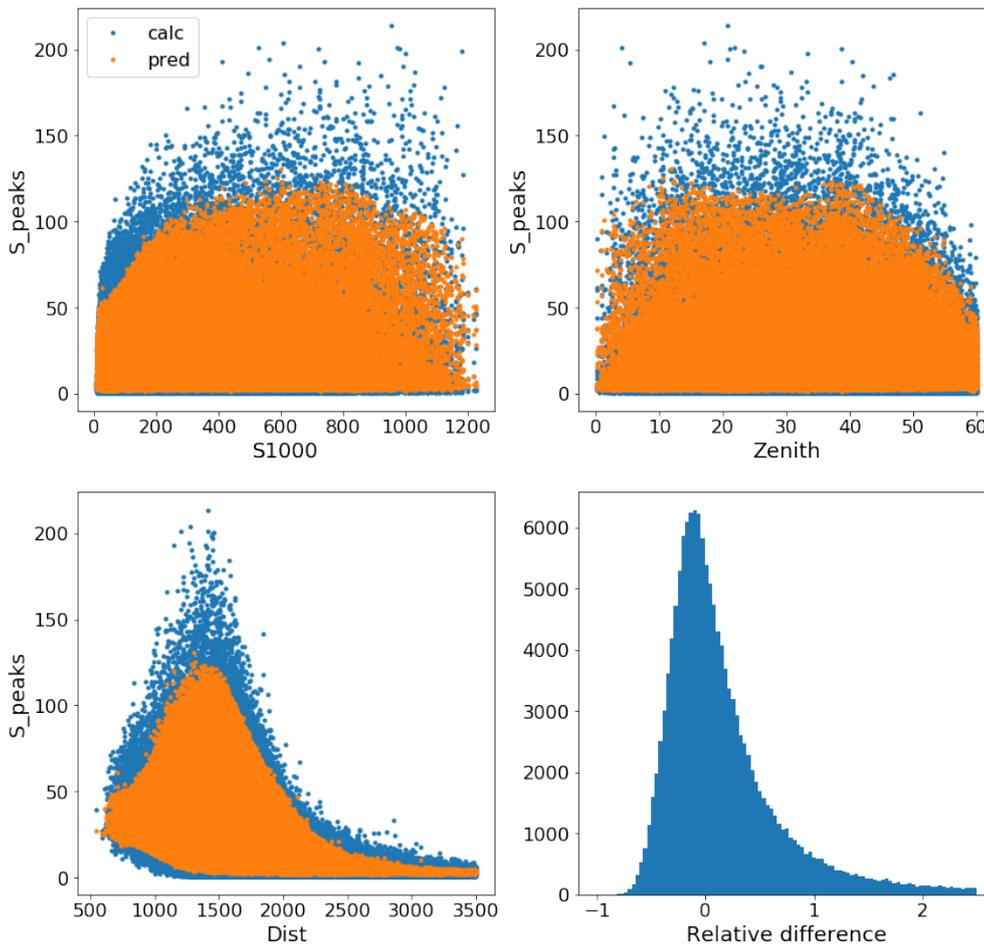
Figure 7.4: Distributions of calculated values (in blue) and predicted values using the KNR algorithm (in orange) of S_{peaks} for MC-protons. From top to bottom and left to right, the figures represents: the distributions of the S_{peaks} values as a function of (1) S_{1000}, (2) θ and (3) r. The bottom-right figure (4) represents the relative difference between the S_{peaks} values calculated and the ones predicted by the regressor.

In general, the models have a tendency to make predictions closer to the bulk of the distributions, which is expected since the predicted values are calculated by combining the values of the closest neighbors, and there are obviously more neighbors towards the center of

the distributions.

	S_{peaks}		f_μ		$RiseTime$	
	KNR	ETR	KNR	ETR	KNR	ETR
Mean Test Score	0.9041	0.9133	0.7933	0.7992	0.3297	0.3286
Std Test Score	0.0006	0.0007	0.0016	0.0021	0.0037	0.0033
Training Time (s)	0.0557	5.0532	0.0504	5.5817	0.0451	6.4941
Scoring Time (s)	0.628	0.7442	1.5665	0.8502	2.3643	1.0796

Table 7.1: Table of the performances of the $KNeighborsRegressor$ and $ExtraTreesRegressor$ models trained for each variable (S_{peaks}, f_μ, $RiseTime$). The mean and std test score are respectively the mean and the standard deviation of the score value obtained from the test sample over the cross-validation. The training and scoring times are the times needed to train and apply the model respectively, where "apply" means making the predictions on the test sample.

As expected the $RiseTime$ predictions have a low "$r2$" value since the parameters used for the predictions (S_{1000}, θ, r) were selected specifically for the other two variables. The poor quality of the predicted $RiseTime$ values will lower the separation power achievable with this observable only, since both proton and photon predictions are farther away from the calculated values whereas the separation stems from the fact that the predictions are "*proton-like*". Nonetheless, as it is presented further, the $RiseTime$ still provides substantial photon/hadron separation.

The two types of algorithms show similar results, and the choice of one algorithm over the other is not straightforward. The $KNeighborsRegressor$ algorithm was selected, in part because of the simplicity of its concept and also because it proved to provide consistent and stable results over many iterations and on sparser datasets. Another, not critical but noteworthy point is that the KNR training is much faster than the ETR training. For another analysis, where the number of predictions is several orders of magnitude above this present one, the ETR algorithm which makes predictions faster, at the cost of a more expensive training, could be a more appropriate choice. In particular, for a potential *online* analysis, where the events would be analyzed semi-continuously, the models would just have to be trained once and then could be applied in a very time-efficient manner.

Another conclusion from the very similar prediction qualities calculated with both algorithms, is that they "saturate" (the scoring values plateaued) before getting a perfect score. This shows that the limiting factor in the prediction is not the models themselves but more likely that the parameters used to make the predictions are not sufficient to reproduce the behaviours of the station-level variables. Either the values of the variables depend on a parameter X which is not provided to the model and thus it will never be able to perfectly predict the variable values, or there is just too much randomness or noise in the variables calculation making them unfittable by a function.

7.2.3   Analytical functions

A flaw often pointed out against the use of ML models is that they are black boxes. To ensure the validity of the method, and to justify the use of the machine-learning predictions, analytical functions predicting the values of the station-level variables as a function of r, were fitted on the MC-proton dataset, in bins of (S_{1000},θ). The analysis steps are the same using these analytical functions as the one using ML models, but instead of training the models, the functions parameters are fitted on MC-proton simulations, in bins of (S_{1000}, θ). The fitted parameters are stored, and used to predict the values of the station-level variables.

Each station-level variable (S_{peaks}, f_μ and $RiseTime$) is fitted in bins of (S_{1000}, θ) as a function of r. The functions are fitted on the average values of the station-level variables in bins of distance. The binning of the dataset must strike a balance between having a grid as thin as possible and a number of stations in each bin large enough to have an accurate average. If a bin has less than $n_{entries} = 10$, the stations in this bin are removed from the analysis.

The functions used to fit the station-level variables as a function of distance r are:

- S_{peaks}: $f_{S_{peaks}}(r) = C_{S_{peaks}} + A_{S_{peaks}}.exp(-\frac{(r-1000\,\mathrm{m})}{B_{S_{peaks}}})$

- f_μ: $f_{f_\mu}(r) = C_{f_\mu} + A_{f_\mu}.log(\frac{r}{B_{f_\mu}})$

- $RiseTime$: $f_{RiseTime}(r) = A_{RiseTime}.log(\frac{r-0}{B_{RiseTime}}) - A_{RiseTime}.exp(\frac{r-C_{RiseTime}}{B_{RiseTime}})$

where A_{var}, B_{var} and C_{var} are the parameters used to fit the functions (different for each bin and for each variable).

The $RiseTime$ distribution is usually fitted analytically only up to $r = 1400\,\mathrm{m}$ [184], so a new function was designed to account for the wider parameter space of the analysis. It is the sum of two components, a logarithm to describe the slow increase up to a given r (θ dependant) and a negative exponential, added to describe the decrease at higher r. In Figure 7.5, examples of the analytical functions fits are presented. The predictions reproduce the expected behaviour, with the S_{peaks} values following the lateral distribution of the signals as it is correlated to the stations total signal. The f_μ distribution increases with r towards a maximum value of 1, since the EM component is expected to be concentrated in the central part of the shower, so as the EM component gets weaker, the fraction of muons tends to 1.

The parameterizations could be optimized further, for example past analyses have built a global benchmark for the $RiseTime$ where the parameters of the functions depend explicitly on the zenith angle θ. But the analytical functions described here have been built simply to check the behaviours of the variables and compare (see next subsection 7.2.4) the ML predictions with analytical ones to make sure that they behave as expected from the shower physics.

To quantify the quality of the analytical models, the mean relative error on the prediction is used, and compared to the Mean Test Score of the ML models. The quality of the analytical models predictions is lower than the ML models ones, but they reproduce the station-level variables quite well (see Figure 7.6).

7.2.4   Proton versus photon predictions

The proton/photon discrimination stems from the way the predictive models are trained. Indeed, both the machine learning and the analytical models are only trained on MC-protons. Consequently, the predictions made by the models of the station-level variables, based on the stations distance to the shower core and the S_{1000} and θ of the event, are expected to be closer from the *true*/calculated values for proton-induced shower than for photon-induced showers.

The choice of training on MC-protons instead of MC-photons was motivated by the perspective of training the models on real data. This is not possible with the burn sample, as it contains too few events to correctly train the regressors. It is known [204] that the hadronic interaction models used to simulate air showers do not properly reproduce the number of muons which are experimentally measured in EAS. Thus directly using the data to train the classifier would circumvent this issue.
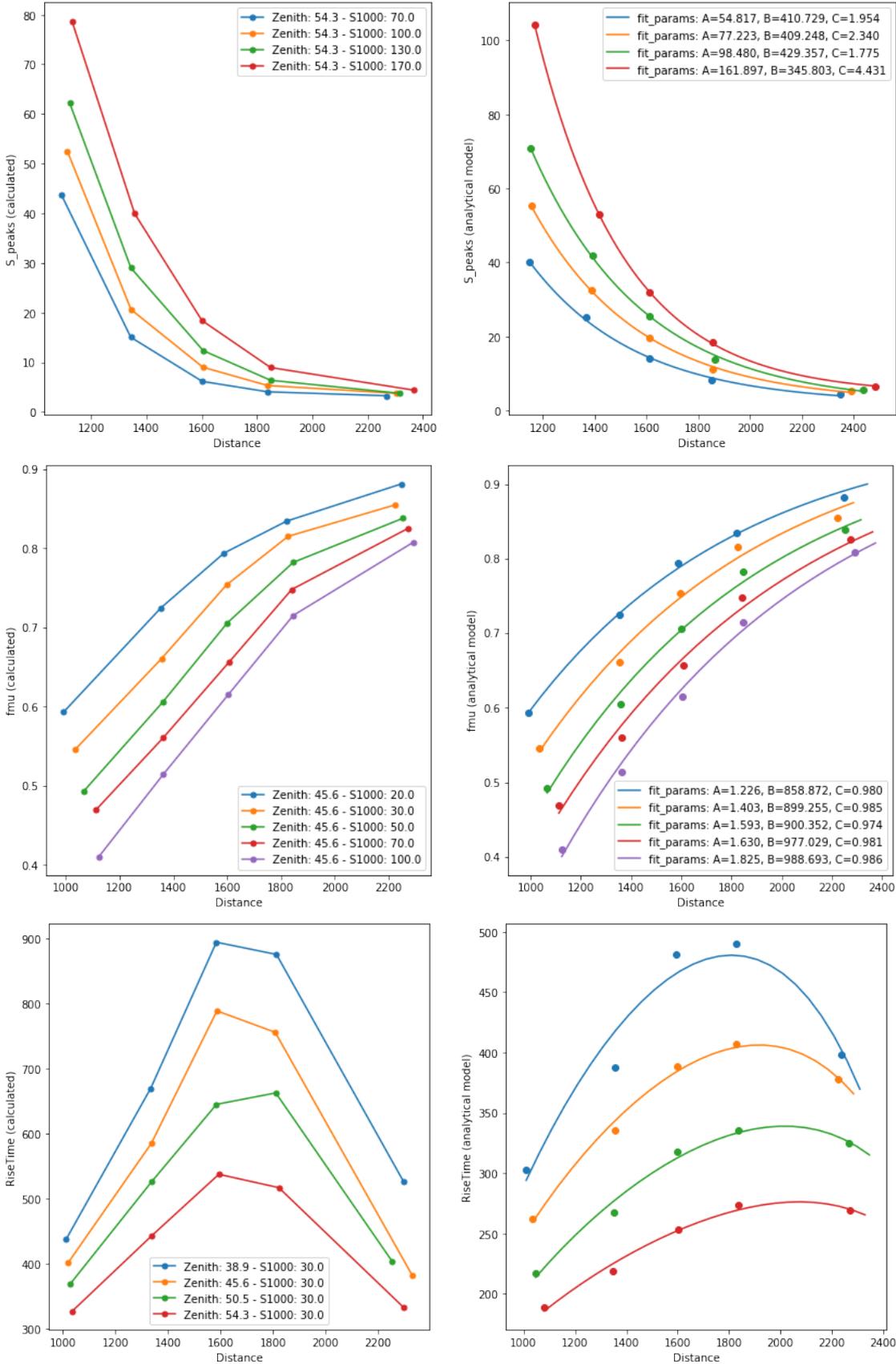
Figure 7.5: Examples of the analytical functions fits on the three station-level variables, from top to bottom: S_{peaks}, f_{μ} and $RiseTime$. On the left: the mean values of the variables in bins of r. On the right: the same mean values with the analytical functions fitted.

The relative difference between the predicted and the calculated values of the variables is thus larger for photons for S_{peaks} and f_μ because less muons are expected in photon-induced showers and for the $RiseTime$ because its values are expected to be lower for protons.
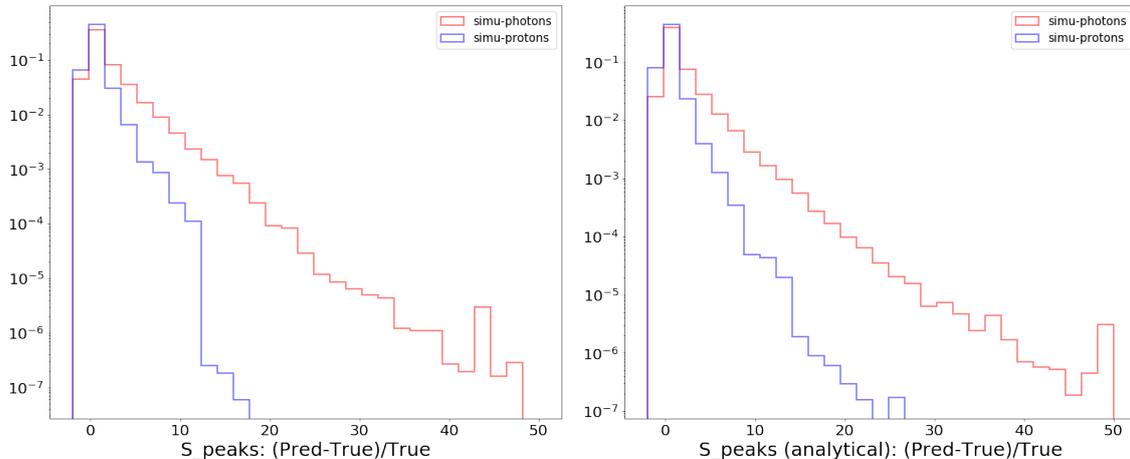
Figure 7.6: Normalized distributions of the relative error of S_{peaks} for protons (blue) and photons (red). Left: ML models. Right: Analytical models.

	MRE (ML)		MRE (analytical)	
	protons	photons	protons	photons
S_{peaks}	-7.2×10^{-4}	3.2×10^{-1}	1.4×10^{-1}	4.1×10^{-1}
f_μ	6.5×10^{-3}	7.2×10^{-2}	-2.7×10^{-2}	3.7×10^{-2}
$RiseTime$	1.1×10^{-2}	-6.3×10^{-1}	-1.8×10^{-2}	-6.5×10^{-1}

Table 7.2: Table of the mean relative errors (MRE) for each station-level variable in the MC-protons and MC-photons distributions. The MRE calculated with the ML predictions are compared to the ones obtained with the analytical models.

The predicted values of the station-level variables are consistently better in the machine-learning case compared to what is obtained with the analytical models (see table 7.2). The ML-reconstructed observables are thus selected for the analysis. In the following parts of this chapter, only the analysis performed using the ML station-level variables is discussed.

7.3 Multi-Variate Analysis

To combine the variables into a photon/hadron discrimination analysis, the station-level variables must be treated at the event-level. A classifier is then trained on the event-level observables to distinguish between photon and proton showers and it assigns a probability to each event of being a photon shower or not.

7.3.1 Event-level observables

In both MC-proton and MC-photon datasets, each selected station (i) of an event is assigned a *proton*-prediction of the station-level variables. To combine these variables (var) into observables at the event-level (obs), the predicted values of the variables (var_{pred}) are divided

by the calculated values (var_{calc}) and averaged over all the stations (N_{st}) in an event:

$$obs = \frac{1}{N_{st}} \sum_{i}^{N_{st}} \frac{var_{pred}(i)}{var_{calc}(i)} \tag{7.2}$$

The station-level variables S_{peaks}, f_μ and $RiseTime$ are combined for each event to calculate the $Muonicity$, $Smoothing$ and T_{Half} values respectively. Each event-level observable is calculated using the ML-models (obs) and the analytical models (obs_f). By comparing the distributions of the two kinds of observables, and confirming that they behave similarly, we were able to verify the "physical sense" of the ML-built observables.
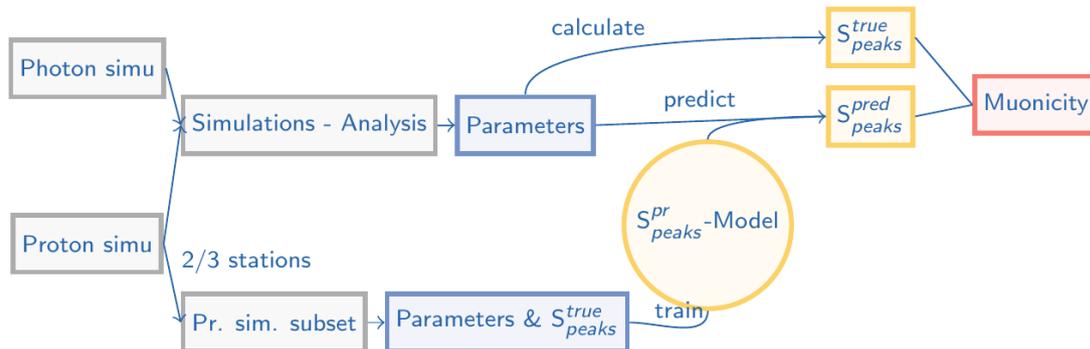
Figure 7.7: Scheme of the observable reconstruction procedure, example of the $Muonicity$ observable.

7.3.2 Weighting the events

The MC-events are simulated by drawing the MC-energy of the primaries from an uniform distribution. To reflect the behaviour expected from real proton or photon primaries, the simulated events need to be reweighted based on their MC-energies.

The rescaling is done according to a spectral index energy distribution: $E^{-\gamma}$ (see Figure 7.8). The choice of a value γ is made to match the latest results on the composition-sensitive spectrum fit of UHECRs [205] so the proton simulations are rescaled according to a $\gamma = 3$ energy spectrum. The choice of a value γ for the photon simulations is harder to justify as we do not have data on which a spectrum fit could be performed to estimate the spectral index. However, given a UHECR emission spectrum following a E^{-1} power law, the propagation effects produce a UHE photon spectrum close to a power law with $\gamma = 2$ (see subsection 2.3.2) For this analysis, we selected the value $\gamma = 2$ as it is the most commonly seen in the literature.

7.3.3 Building the classifier

With the event-level observables calculated (according to equation 7.2), a classifier is trained on 18000 MC-protons and 18000 MC-photons, randomly drawn, to distinguish between the two kinds of showers using these observables and event-parameters S_{1000} and θ.

As with the prediction of the station-level variables, multiple machine-learning models were tested. The same family of algorithms were studied in detail: $KNeighborsClassifier$ and $ExtraTreesClassifier$. This decision was justified by the work done to establish the procedure to build the station-level variable , and mainly by the ability of these algorithms to perform well with large datasets like the MC-simulations we are using in both problems. The scoring metric used for the classifiers in this step of the analysis is the "$precision$" defined as:
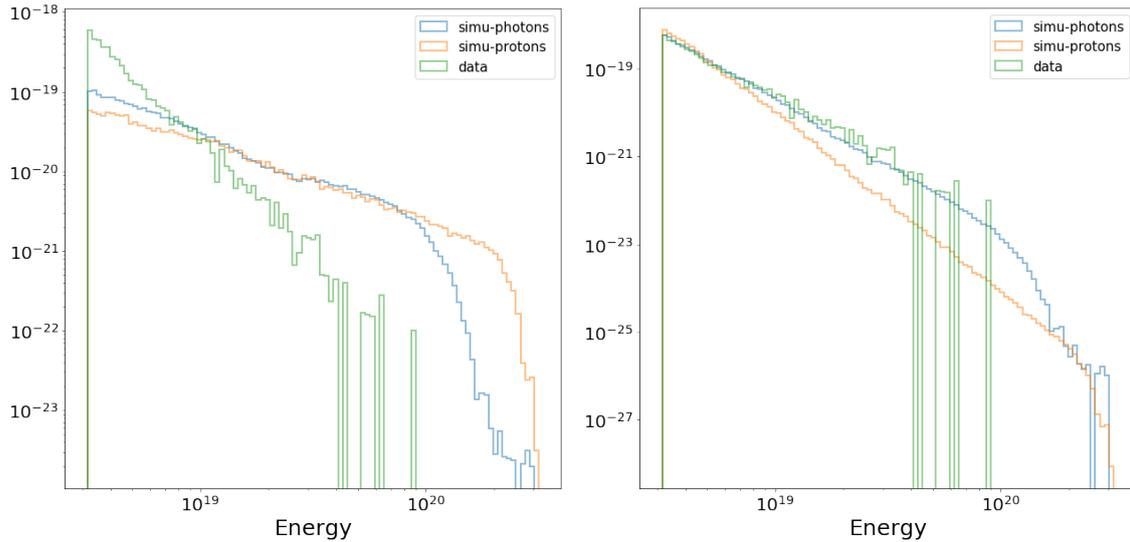
Figure 7.8: Normalized distributions of the reconstructed energies for MC-protons, MC-photons and burn sample events unscaled (left) and with the MC-protons and MC-photons rescaled on a $E^{-\gamma}$ spectrum with $\gamma = 3$ and $\gamma = 2$ respectively. To compare the three distributions they had to be normalized to 1, the vertical axis is thus the probability densities estimate.

$precision = tp/(tp + fp)$ where tp is the number of true positive (number of photons correctly identified as photons) and fp the number of false positive (number of protons misclassified as photons). This metric is appropriate for our objective: it penalizes a misclassified proton and thus leads the model into producing an output with a photon population as pure as possible.

The ETC model works better than the KNC model (see Figure 7.9), as it has a higher "cross-validation" score, i.e. the score obtained on the test data (in opposition to the training data used to fit the model). It was thus selected as the model used to build the classifier. The classifier is trained on the same number of MC-protons and MC-photons.

The output used from the classifier is the predicted class probability. This is the probability the classifier assigns to an event being a photon or not. The probability is in fact a discrete value, as one estimator returns one probability value. It is thus important to select a value of $n_{estimators}$ large enough to have an almost continuous distribution of the classifier probability (clf-proba) for the events. The number of estimators was set to $n_{estimators} = 300$.
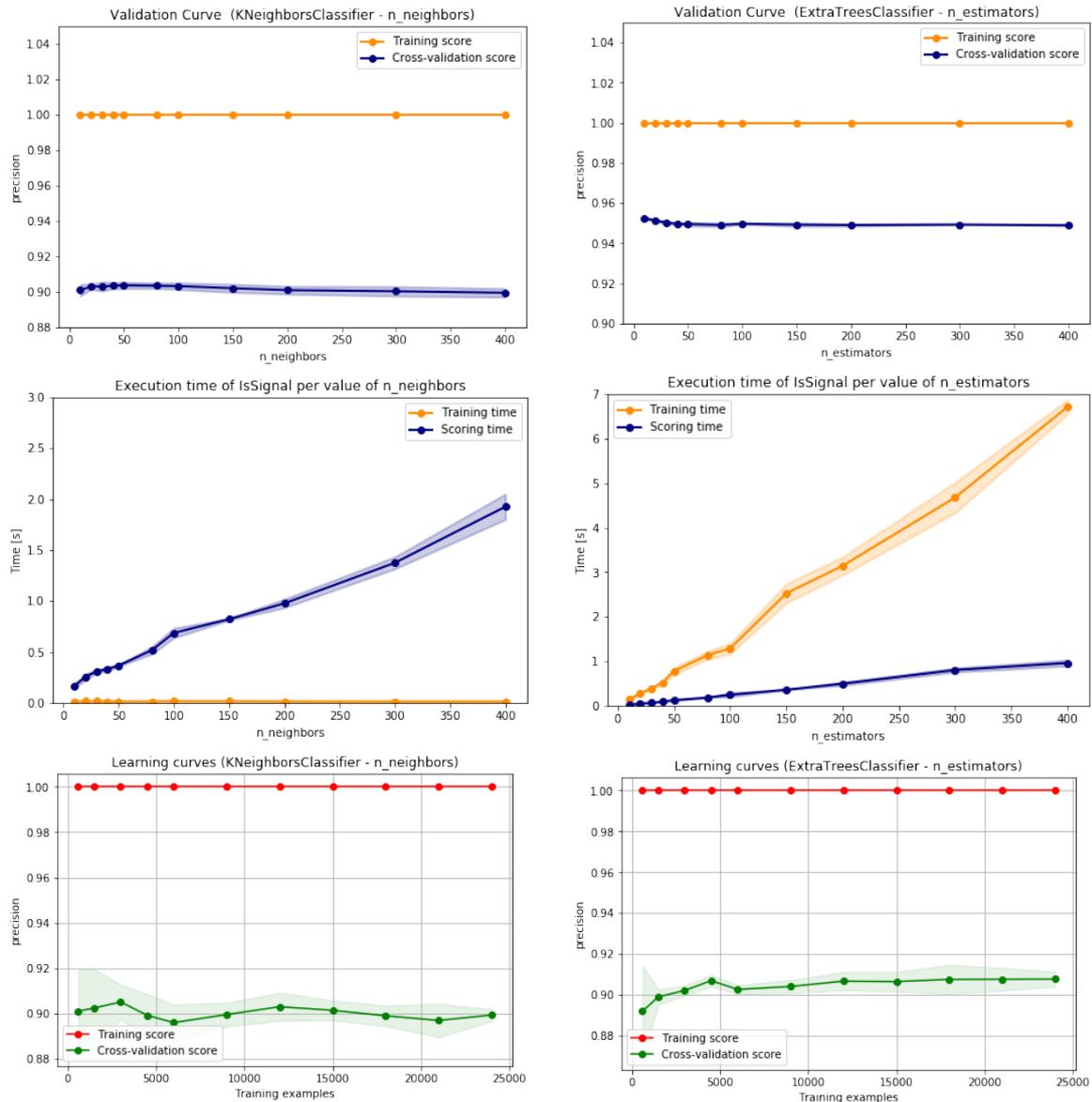
Figure 7.9: The curves used to compare the *KNeighborsClassifier* algorithm (on the left) with the *ExtraTreesClassifier* algorithm (on the right) to select the best classifier. From top to bottom, (1) the validation curve of the score value (*precision*) of the models as a function of the hyperparameter values, (2) the execution time needed to train and score with the models as a function of the hyperparameter values and (3) the learning curve, the score value of the model with increasing number of samples.

7.4 Results of the analysis

In this section, the results of the analysis are presented. First, the expectations from applying the analysis on the simulations are detailed. Then, the results of the analysis applied on the burn sample are discussed. Finally, perspectives for the analysis are explored.

7.4.1 Expectations from the MC-simulation datasets

The trained classifier can now be applied to all of the MC-simulations that were not part of the training dataset. The main tool to assess the quality of the trained classifier is the receiver operating characteristic (ROC) curve, which displays the background rejection as a function of the signal efficiency. For this analysis, the background rejection is the ratio of protons not classified as photons and the signal efficiency is the ratio of photons correctly classified as photons (with the appropriate weights, as discussed in 7.3.2). These two quantities depend on the threshold of classifier probabilities (clf-proba$_{thr}$) above which an event is treated as a photon candidate.

In an UHE photon search, the aim is to improve the background rejection as the flux of hadronic primaries is expected to be much higher compared to the UHE photons flux. The 50% signal efficiency (S.E) is often used as a reference to compare between different analyses. Using the event distribution to set the classifier threshold allows the comparison between different analysis and the median of the photon distribution is conservative enough. At 50% signal efficiency[1], the present designed analysis rejects $\sim 99.87\%$ of the background (see Figure 7.11). This rejection power goes beyond that for lower signal efficiency, but it becomes impossible to quantify due to a lack of statistics.

The same classifier when trained with the analytical observables displays inferior performances, with a background rejection around $\sim 99.75\%$ at 50% signal efficiency (see Figure 7.11).

The number of background events due to misclassification of protons can be estimated from the weighted background rejection obtained from the ROC curves, i.e $n_{bkg} = n_{events} * (1 - 0.9987)$.

This background rejection power is extremely important as it leads to an estimation of the irreducible background due to misclassification in the analysis. Unfortunately, its precision is limited by the test statistics used, and in particular by the number of MC-protons available in the testing sample (testing sample: 3646 MC-protons). This is further amplified by the individual event weighting performed to account for the expected energy spectrum. Indeed, a single event above the 50% S.E threshold can penalize the background rejection and thus the background estimation greatly. In addition, from experimental results the composition of UHECRs is expected to be a mix of proton and heavier primaries, so the actual background should be lower than for a pure proton flux, since the shower from heavier primaries have, on average, a larger muon fraction. Also, it has been shown that the hadronic models used to produce the UHECR showers simulations do not fully reproduce the behaviour of data in term of the muon component, as already mentioned. Any interpretation based on the results obtained with this multivariate analysis are thus to be considered within this context.

Nonetheless, this value provides a conservative estimation of the expected number of background events in Auger data. Since the aim of photon-search analyses is to find very rare events, a high confidence level must be sought to make a valid physical claim, hence it was decided to take the most conservative approach for background estimation and limit-setting.

---

[1]The value of the classifier probability output that corresponds to the median of the weighted MC-photon distribution.
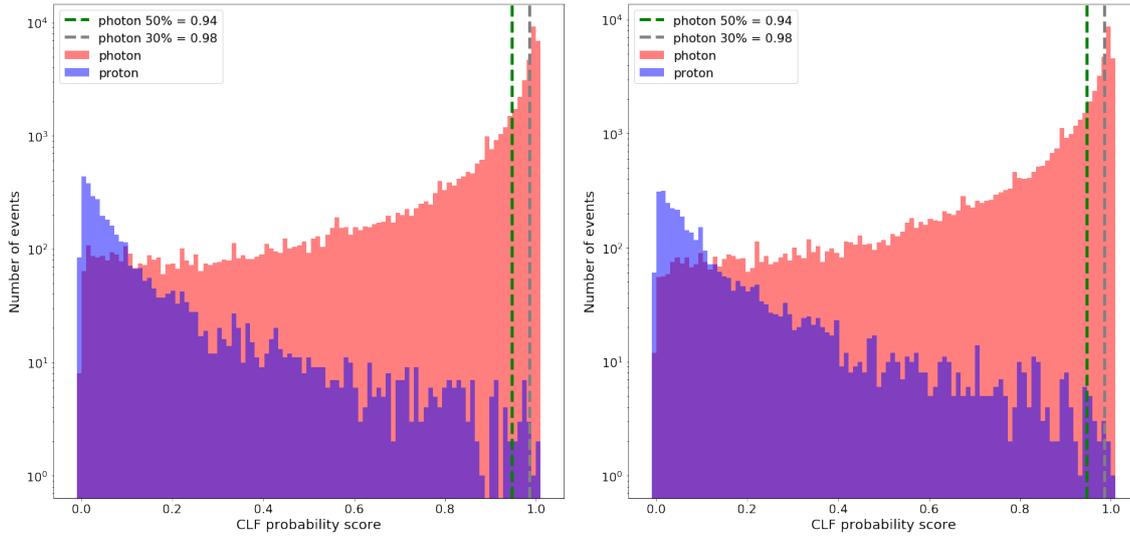
Figure 7.10: Distributions of the classifiers probability scores assigned to the MC-simulations. On the left, the probabilities are those assigned with the ML-observables and on the right, the ones assigned with the analytical observables. Only events not part of the training dataset are included in these histograms. In blue and red are respectively the MC-protons and MC-photons. The green dashed line corresponds to the median of the photon distribution (i.e. 50% signal efficiency), the grey dashed line corresponds to the 30% signal efficiency line.
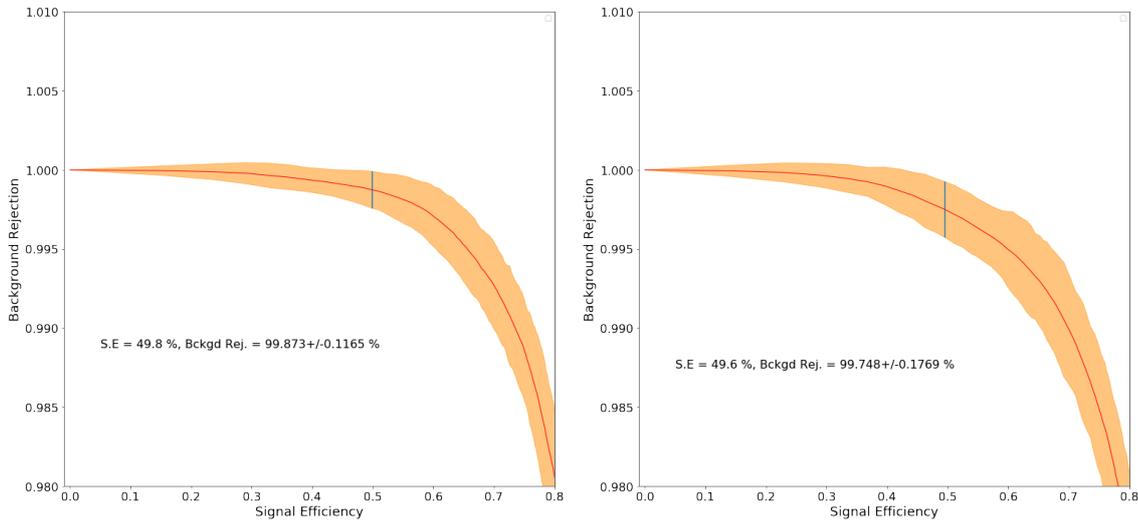
Figure 7.11: Zoomed-in ROC curves of the classifiers built with the ML-observables (left) and the analytical observables (right). The ROC curve (red line) shows the number of protons correctly classified as background (i.e not photons) for a given signal efficiency (S.E). The S.E is the fraction of photons passing the cut. To maximize the purity of the prediction, the background rejection must be as close to 100% as possible, but lowering the S.E means lowering the number of potential signals (i.e photons) in the analysis. The training of both classifiers is performed 100 times to estimate the errors on the predictions. The spread of the response is displayed as the orange area and the vertical blue line highlights the spread at 50% S.E.

A source of background comes from leading π^0 events whose first interaction produces a neutral pion, carrying most of the energy and as a result producing a mostly electro-magnetic shower [127]. This phenomenon is negligible as it is very rare[1], but it is nonetheless present in the simulated proton showers.

7.4.2 Application of the analysis on a burn sample of data

The whole analysis described in this chapter has been applied on the burn sample described at the beginning of the chapter (see section 7.1). A few events are discarded during the analysis either because none of their stations passed the selection cuts applied on the station, or because the event-level observables are not reconstructed properly (for example due to a null prediction).

Cut	Event remaining after the cut	% remaining post-cut
Burnsample	100299	1.69%
Event is 6T5	94095	93.81%
$\theta \in [0°,60°]$	92361	98.16%
Proper st-lvl variables calculation	92211	99.84%
E$> 10^{18.5}$ eV	2458	2.67%
Event has $\geq$ 2 selected stations	2457	99.96%
Proper st-lvl variables predictions	2457	100%
Proper evt-lvl observables calculation	2441	99.35%

Table 7.3: Table of the evolution of the number of events in the burn sample after application of the selection cuts or analysis steps. The "Proper" variables calculation/prediction and observable calculation cuts rejects cases in which either the calculation or the prediction returns a null or infinite value.

The analysis applied on the 2441 burn sample events selects one event above the 50% S.E threshold, that should thus be treated as a photon candidate (SdId 7426428). From the model testing, one could expect $2441 \times (1 - 0.9985) \sim 3.5$ background events due to misclassified protons. So this number of found events is within expectations.

Event 7426428 was studied, to make sure that it is a background event and, if this is the case, understand why the analysis assigns it a high probability score. This event has a small zenith angle $\theta = 14°$ and has a reconstructed energy $E = 3.55 \times 10^{18}$ eV. Due to its low zenith angle, the signals in the stations close to the core are concentrated within a short time window which makes the *Muonicity* and *Smoothing* methods unable to disentangle the components (see Figure 7.1). In addition to that, the station further away from the core has peaks that can be interpreted as muon peaks but they are below the minimum value S_{min} for S_{peaks}.

Event 7426428 happens to be an hybrid event, also detected by the FD. This allows the cross-check with the measurement of the shower profile through the atmosphere, which is an observable with high hadron/photon discrimination power. The measured dE/dX matches with the expectations from hadronic simulations with a maximum measured below $800 \, \text{g.cm}^{-2}$ (see Figure 7.13). This observation, confirms that this event is a background event and not a real photon event.

---

[1]a proton shower has a probability of producing a π^0 in the first interaction of about $\sim 12\% - 14\%$ depending on the hadronic interaction model used. Then the inelasticity decreases exponentially, leading to the probability of a leading π^0 carrying $\sim 80\%$ of the primary energy being around $\sim 10^{-4}$. See [206].
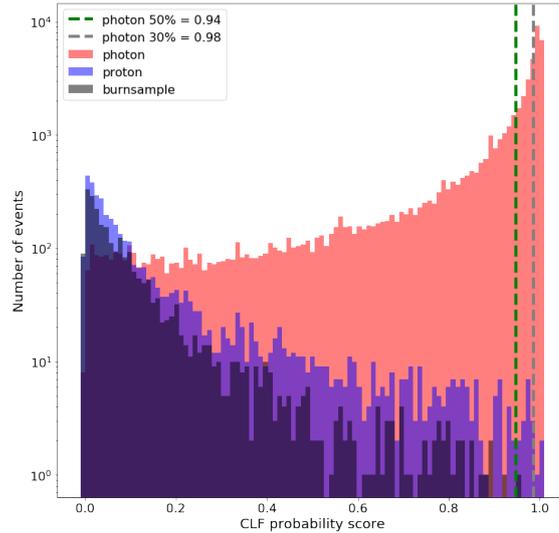
Figure 7.12: Same figure as the Figure 7.10 with the events from the burn sample added in black. One event from the burn sample is above the 50% signal threshold and should be treated as a photon candidate.
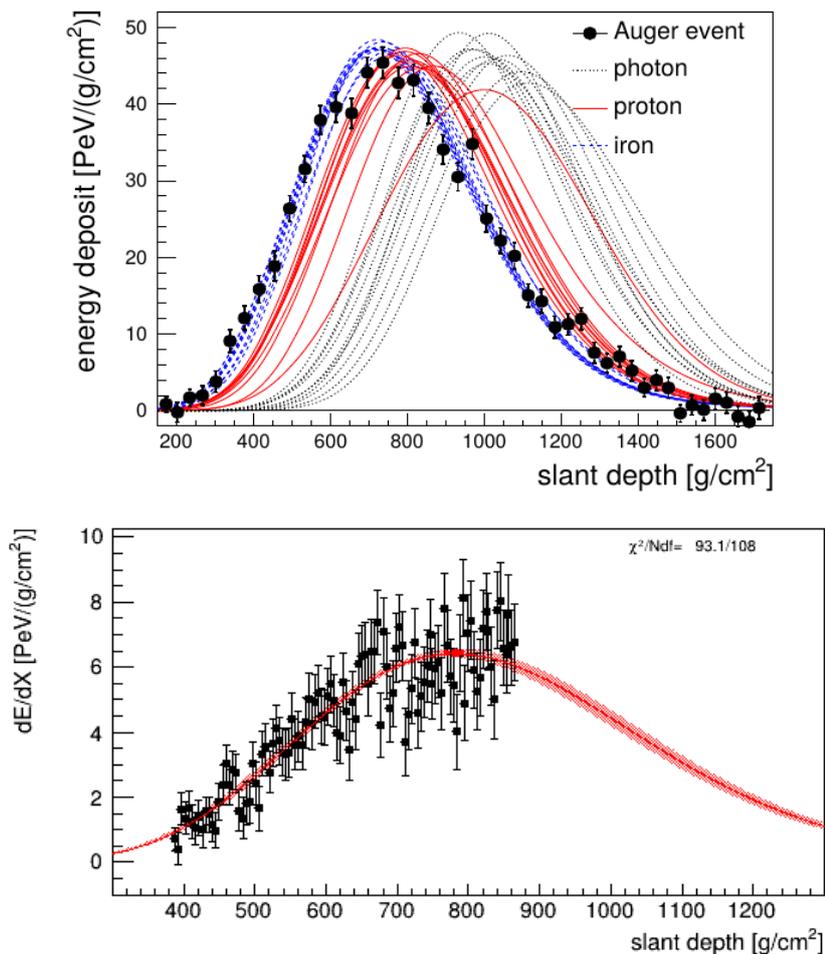
Figure 7.13: Top: Examples of longitudinal development profiles for MC-simulations of iron, proton and photon primaries. For hadronic primaries the maximum of dE/dX is expected below $\sim 820\,\mathrm{g.cm^{-2}}$, while for photons it is expected around or above $900\,g.cm^{-2}$. Bottom: Longitudinal development profile of event 7426428 measured with a FD telescope located Los Morados. This distribution of the dE/dX values as a function of the slant depth matches the hadronic-induced showers expectations from MC-simulations, with a maximum slightly below $\sim 800\,\mathrm{g.cm^{-2}}$.

Since the number of photon candidates (n=1) is compatible with the MC-proton background expectations (b=3.5), upper-limits on the integral photon flux can be derived. The occurrence of signal events expected in this analysis are independent from one-another and are not correlated in time, they thus follow a Poissonian distribution and the Feldman-Cousins method can be used. Using the Feldman-Cousins method, the upper-limits on the number of photon candidate at 95% confidence level (C.L) $N_\gamma^{95\%}$ can be calculated along with the upper-limit on the sensitivity of the analysis $S^{95\%}$. The background is also expected to be Poisson-distributed hence the Feldman-Cousins method can be used to derive the 95% C.L. upper and lower limits on the expected background from the burn sample result: $0.05 < n_{bkg}^{b.s.} < 5.14$.

For $E > 10^{18.5}$ eV:

- for $(n_0 = 1, b = 5)$: $N_\gamma^{95\%}$=1.88, $S^{95\%}$=6.32

- for $(n_0 = 1, b = 3.5)$: $N_\gamma^{95\%}$=2.08, $S^{95\%}$=5.62

- for $(n_0 = 1, b = 0)$: $N_\gamma^{95\%}$=5.14, $S^{95\%}$=3.09

These three expected background values (b=0, 3.5, 5) have been used to calculate the 95% C.L upper limits on the integral photon in bins of energy using the relation:

$$\Phi^{95\%}(E > E_0) = \frac{N_\gamma^{95\%}(E > E_0)}{\text{eff}_\gamma.\varepsilon.A} \tag{7.3}$$

where A is the exposure of the dataset, ε is the selection efficiency, i.e. the ratio of events selected for the analysis compared to the total dataset, eff_γ is the signal efficiency (i.e equal to 0.5 since the threshold is set at the median of the photon distribution).

The exposure is calculated from the number of active 6T5 elementary cells in the array multiplied by the detection area of each cell integrated over time. For events with 6T5 reconstruction above $10^{18.5}$ eV between 01/01/2004 and 31/12/2017, the exposure is 56787.95 km^2.sr.yr. The assumption was made that the energy of the events does not impact the calculation of the exposure in this analysis. This assumption is valid since for 6T5 events above $10^{18.5}$ eV the trigger efficiency is 100%. At lower energies or for less stringent event-level triggers, the exposure can experience significant variations along the energy spectrum (lower exposure if the trigger efficiency is not 100%). It is still needed to point out however that some of the quality cuts[1] applied on the data can have a small energy dependence that could slightly affect the exposure. These effects are supposed to be negligible though, as the cuts potentially concerned by this phenomenon only discard 2.7% of the events.

The selection efficiency is directly calculated from table 7.3, by multiplying the remaining fraction of events at each step of the analysis, at the exception of the energy cut, as it doesn't affect the selection efficiency: $\varepsilon = 0.0154$. Note that this selection efficiency considers the fact that the burn sample is used as a 98% cut.

The upper-limit on the integral photon flux above $10^{18.5}$ eV at 95% C.L is then, for the b=0 case (most conservative):

$$\Phi^{95\%}(E > 10^{18.5}\,\text{eV}) = \frac{5.14}{0.5 \times 0.0154 \times 56787.95} = 1.17 \times 10^{-2}\,\text{km}^{-2}.\text{sr}^{-1}.\text{yr}^{-1} \tag{7.4}$$

The results for each value of expected background can be found in table 7.4. The most conservative value is selected.

The analysis has also been applied in bins of energies $E > E_0$ for $E_0 = 10^{19.0}$ and $10^{19.5}$ eV. In both of these bins, no photon candidate was found and given the low number of events

---

[1]The 6T5 cut is already considered to calculate the exposure.

Values of expected background b	0	3.5	5
$N_\gamma^{95\%}$ ($n_0 = 1$)	5.14	2.08	1.88
$S^{95\%}$	3.09	5.62	6.32
$\Phi^{95\%}(E > 10^{18.5}\,\text{eV})$ (km^{-2}.sr^{-1}.yr^{-1})	1.17×10^{-2}	5.3×10^{-3}	4.3×10^{-3}

Table 7.4: Effect of the expected background on the analysis results following the Feldman-Cousins method. For three values of expected background $b = 0, 3.5, 5$, three quantities are calculated: (1) $N_\gamma^{95\%}$, the number of expected photon candidate at 95% C.L.; (2) $S^{95\%}$, the upper-limit on the sensitivity at 95% C.L.; (3) $\Phi^{95\%}(E > 10^{18.5}\,\text{eV})$, the integral photon flux above $10^{18.5}\,\text{eV}$ at 95% C.L.

from the burn sample in each bin, the expected background is 0 ($b = 0$). The 95% C.L. upper-limits on the integral photon flux have thus been calculated, and the results can be found in table 7.5.

E_0 (log/eV)	18.5	19.0	19.5
n_0	1	0	0
$N_{gamma}^{95\%}$ ($b = 0$)	5.14	3.09	3.09
$\Phi^{95\%}(E > E_0)$ (km^{-2}.sr^{-1}.yr^{-1})	1.17×10^{-2}	7.07×10^{-3}	7.07×10^{-3}

Table 7.5: Results of the analysis in three bins of energy, E (log/eV) $> 18.5, 19.0, 19.5$. Three quantities are represented: (1) n, the number of photon candidate in the dataset; (2) $N_{gamma}^{95\%}$ ($b = 0$), the number of expected photon candidate from the Feldman-Cousins method assuming a background $b = 0$ at 95% C.L.; (3) $\Phi^{95\%}(E > E_0)$, the integral photon flux above E_0 at 95% C.L.

It is important to keep in mind here, that these upper-limits are very conservative due to the lack of events in these energy bins. The upper-limits above $10^{20}\,\text{eV}$ cannot be calculated since there are no events with reconstructed energies above $10^{20}\,\text{eV}$ in the burn sample.

The aim of the strategy followed to design this analysis was to limit the number of selection criteria as much as possible to minimize the loss of exposure. The difficulty is to maximize the purity of the signal prediction of the classifier, i.e. to minimize the number of false-positive (a proton or a hadron classified as a photon).

This balance between the number of selection criteria applied and purity of the classification must be achieved for any analysis in which there is a low signal/background ratio. This work has shown that ML algorithms are efficient and practical tools to achieve a better balance as they are suited for high dimensionality problems.

7.4.3 Perspectives for the analysis, application of SSD data

The existing photon searches performed by the Pierre Auger collaboration set the most stringent upper-limits on the photon flux at the ultrahigh energies. The analysis detailed in the present work, as some others which are also developed, aim at further refining the photon shower identification, to increase their effectiveness.

The results presented in the previous section are obtained using only 2% of the whole SD dataset. By applying the designed ML analysis process on the whole dataset and assuming that the burn sample is unbiased, the upper-limits on the integral photon flux (around $\sim 10^{-2}\,\text{km}^{-2}$.sr^{-1}.yr^{-1}) could be brought down to values in the order of magnitude of the limits set by the Auger analysis currently used to search for UHE photons with the SD events (around $\sim 10^{-3}\,\text{km}^{-2}$.sr^{-1}.yr^{-1}).

Increasing the statistics of the available simulated showers, in particular the proton ones, would increase the performance of the designed analysis. Indeed, the classifier could be improved with a larger training sample, and a larger testing sample would allow a more robust estimation of the background.

Different Auger analyses have underlined that the models currently used to simulate the showers have difficulties reproducing the measurements, in particular concerning the shower muon content, which is dependent on the hadronic interactions in the shower development [204]. So the drawback of the presented analysis process is its dependence on the proton showers simulations. Using real data to build the models would lessen the dependence on simulated dataset. Indeed, in this case, only the MC-photons simulations, which are not dependent on the knowledge on the hadronic interactions models at very high energy, would be needed to build the classifier. However, the present ML analysis, as it has been designed, requires a large number of events to perform the training, and unfortunately the available burn sample does not contain enough data in the selected parameter space.

The MVA described in this chapter is self-consistent, and it also shows the potential of an estimator of the muonic content as a discriminant observable in a photon search. With the AugerPrime project, the photon searches performed with the Pierre Auger Observatory will have the opportunity to include a more robust estimator of the muon content of the showers at ground-level, thanks to the simultaneous sampling of the showers with the WCDs and the SSDs. Preliminary estimation on prototypes gives a resolution of 5% on the reconstructed number of muons [158]. In this framework, it will be worthwhile to upgrade the present MVA-analysis by including information from the SSD.

In the coming years, the Pierre Auger Observatory will continue to collect data, increasing the available exposure for the photon search analyses. With the AugerPrime upgrade, new composition sensitive observables will be available, further increasing its shower-by-shower primary identification capabilities. With such identification power, one can hope to identify UHE photons in the UHECR flux; if no detection of UHE photon signal is achieved, the upper-limits on their flux derived by the Pierre Auger collaboration will continue to be improved, allowing to constrain the GZK scenario as well as the UHECR source models.

The results presented in the last section are obtained using only 2% of the whole PAO SD-only dataset. By using the whole dataset and assuming that the burn sample is unbiased, the upper-limits on the integral photon flux could be brought down to values in the order of magnitude of the limits set by the most recent PAO SD-only analysis.

Another way to improve the analysis would be to improve the statistics on the MC-simulations, especially the number of MC-proton events. The classifier could be improved with a larger training sample and a larger testing sample would allow a more robust estimation of the background. The drawback is that this analysis is dependant on the UHE showers simulations, which are known to not fully reproduce the observed showers [204]. Using data to build the models would lessen this dependence on simulations. This way, only the MC-photons simulations would be needed to build the classifier and EM showers simulations are better understood than hadronic showers simulations. However, this analysis needs a large number of events to perform the training and the burn sample available does not contain enough showers (in the parameter space selected for this analysis).

The photon searches performed with the PAO will be able to use a more robust estimator of the muon content (preliminary estimation on prototypes gives a resolution of $\sim 5\%$ on the reconstructed number of muons [158]) of the showers at ground-level with the AugerPrime upgrade, thanks to the simultaneous sampling of the showers with the WCDs and the SSDs. The MVA described in this chapter is self-consistent, but it also shows the potential of an estimator of the muonic content as a discriminant observable in a photon search.

In the coming years, the Pierre Auger Observatory will continue to collect more data, increasing the available exposure for the photon search analyses. With the AugerPrime upgrade, new composition sensitive observables will be available, further increasing the shower-by-shower primary identification capabilities of the PAO. The existing photon searches performed with the PAO set the most stringent upper-limits on the photon flux at the ultra-high energies. Furthermore, this work and other aim at further refining them, to increase their effectiveness. The upper-limits on the flux of UHE photons with the PAO will continue to be pushed down, either constraining the GZK scenarios or a positive UHE photon signal will be observed.

Chapter 8

Conclusion

This thesis work focuses on the search for ultra high energy photons in the experimental framework provided by the Pierre Auger Observatory (PAO), which is currently the leading experiment to study ultra-high energy cosmic rays (UHECR). Being designed as an hybrid instrument, it is composed of a fluorescence detector (FD) and a surface detector (SD). The latter is an array of 1660 water cherenkov detectors (WCD) deployed on $3000\,\mathrm{km}^2$ that samples at ground the secondary particles of extensive air showers (EAS) produced when UHECRs interact with the atmosphere.

The objective of the major part of my thesis work was to design a new multi-variate analysis (MVA) to search for ultra-high energy photons in the flux of UHECR above 3×10^{18} eV detected with the SD of the PAO. This MVA aims at enhancing discrimination power between EAS produced by hadrons to those generated by UHE photons, from the estimation of the fraction of the signal at ground due to the muonic component of the EAS. To that end, variables correlated to the muonic fraction of the signal recorded by the SD are calculated at the individual detector level, before using machine-learning (ML) regression algorithms to build one discriminant observable for each detected shower. Usually, analytical functions are used to build benchmarks representing the value of the variables at the individual detector level as a function depending on shower reconstructed parameters. For this work, ML algorithms were used instead of analytical functions for three main reasons: firstly, the number of selection criteria to be applied is reduced; secondly, the method allows a continuous prediction over the parameter space instead of training multiple functions, and thirdly the ML method and its implementation present the advantage of being flexible (i.e. any photon-specific feature can be used since there is no need to manually build a specific analytical function). The analysis has been designed and its performances evaluated from simulated showers.

The UHE photon search was applied to 2% of all the data collected by the SD between 01/01/2004 and 31/12/2017, this corresponds to the event sample used by the collaboration to test photon search analyses. At 50% signal efficiency (S.E.), the background rejection power of the MVA is 99.87%. Out of 2441 preselected events, 1 photon candidate was found. It was checked that this event is not an actual photon through a complementary observable from the FD, however it was still taken into account to calculate the limits. The Feldman-Cousins upper-limits at 95% confidence level on the photon flux above $E > E_0$ with $\log(E_0) = (18.5, 19.0, 19.5)$ have been derived assuming no expected background (most conservative option) and are respectively: $\Phi^{95\%}(E > E_0) = 1.17 \times 10^{-2}, 7.07 \times 10^{-3}, 7.07 \times 10^{-3}$ $\mathrm{km}^{-2}.\mathrm{sr}^{-1}.\mathrm{y}^{-1}$.

The performances of this MVA analysis can be improved. First by increasing the exposure, since the number of photon candidate is not the limiting factor in this analysis and also by increasing the number of simulated showers to be used in the models training. Second, by adding new observables. Indeed I have shown, through the addition of the a new observable

taking the timing information into account to the MVA, that any discriminant feature at the individual station level can be used to create a new discriminant observable using the framework designed for this analysis.

The capacity of the PAO to search for UHE photons will also be increased by the ongoing upgrade of the Observatory, the AugerPrime project. In this work, I presented my contributions to this upgrade through the successful setup of an assembly process and a testing procedure for the scintillator surface detectors (SSD) at the LPSC. The main quality criteria selected to compare the detectors is the single photoelectron (SPE) per vertical minimum ionizing particle (VMIP). It quantifies the amount of light produced by the SSDs as a vertical atmospheric muon goes through it. The SPE per VMIP values can be compared between institutes as they are only weakly dependent on the test-setup used. These test results shared between the collaboration sites that took part in the SSDs assembly are stored on a common database and will allow a better monitoring and traceability of the detectors once they are installed at the Observatory. All 90 SSDs that had to be constructed at the LPSC passed the requirements from the collaboration. They were shipped to the Pierre Auger Observatory and are currently being deployed.

The use of information provided by the SSDs will improve the SD estimation of the EAS muonic component. Indeed, it will add another independent measurement of the EAS secondary particles at the ground to the one performed with the SD WCD. In particular, new discriminant variables could be added to the MVA presented in this work. Overall, the AugerPrime upgrade will improve the capability of the detector to perform shower-to-shower primary mass identification. In the coming years, running with the upgrade will increase the exposure available for UHE photon analyses (and UHECRs in general).

The improved UHE photon search analysis combined with the larger exposure will lead to one of two scenarios: either UHE photon detection is made or the upper-limits on the photon flux at UHE are brought further down. An UHE photon detection would open a new window of astronomy. The observation of these neutral particles would allow us to locate the nearby sources of UHECRs. It would also shed light on the limits of the sources acceleration power. In either scenarios, a better performing UHE photon search will allow us to put better constraints on astrophysical models such as: UHECRs propagation models, models of UHECRs and TeV photons sources and super heavy dark matter models. Finally, it would further increase the role of UHECR observatories in multi-messenger campaigns.

Bibliography

[1] Charles A Coulomb. Troisieme Mémoire sur l'Electricité et le Magnétisme. In: *Histoire de l'Académie Royale des Sciences, l'Acad 'emie Royale des Sciences Paris* (1785), pp. 612–638 (cit. on p. 8).

[2] De Angelis Alessandro. Domenico Pacini and the origin of cosmic rays. In: *CERN Courier* (May 2012), p. 52. URL: https://cerncourier.com/a/domenico-pacini-and-the-origin-of-cosmic-rays/ (cit. on p. 8).

[3] Pierre Auger, P. Ehrenfest, R. Maze, J. Daudin, and Robley A. Fréon. Extensive Cosmic-Ray Showers. In: *Rev. Mod. Phys.* 11 (3-4 July 1939), pp. 288–291. URL: https://link.aps.org/doi/10.1103/RevModPhys.11.288 (cit. on p. 8).

[4] Bruno Rossi. Misure sulla distribuzione angolare di intensita della radiazione penetrante all'Asmara. In: *Supplemento a la Ricerca Scientifica* 1 (1934), p. 579 (cit. on p. 8).

[5] Kurt Schmeiser and W Bothe. Die harten Ultrastrahlschauer. In: *Annalen der Physik* 424.1-2 (1938), pp. 161–177 (cit. on p. 8).

[6] W Kolhörster, I Matthes, and E Weber. Gekoppelte Höhenstrahlen. In: *Naturwissenschaften* 26.35 (1938), pp. 576–576 (cit. on p. 8).

[7] Enrico Fermi. On the origin of the cosmic radiation. In: *Physical review* 75.8 (1949), p. 1169 (cit. on p. 8).

[8] G. T. Zatsepin and V. A. Kuzmin. Upper limit of the spectrum of cosmic rays. In: *JETP Lett.* 4 (1966). [Pisma Zh. Eksp. Teor. Fiz.4,114(1966)], pp. 78–80 (cit. on pp. 8, 18, 54).

[9] K. Greisen. End to the Cosmic-Ray Spectrum? In: *Phys. Rev. Lett.* 16 (17 Apr. 1966), pp. 748–750. URL: https://link.aps.org/doi/10.1103/PhysRevLett.16.748 (cit. on pp. 8, 54).

[10] J. Linsley. Evidence for a Primary Cosmic-Ray Particle with Energy 10^{20} eV. In: *Phys. Rev. Lett.* 10 (4 Feb. 1963), pp. 146–148. URL: https://link.aps.org/doi/10.1103/PhysRevLett.10.146 (cit. on p. 8).

[11] Flyś Eyes collaboration. The Cosmic-Ray Energy Spectrum Observed by the Flyś Eye. In: *Astrophysical Journal* 424 (Mar. 1994), p. 491 (cit. on pp. 8, 34).

[12] S. Yoshida et al. The cosmic ray energy spectrum above 3.10^{18} eV measured by the Akeno Giant Air Shower Array. In: *Astroparticle Physics* 3.2 (1995), pp. 105 –123. URL: http://www.sciencedirect.com/science/article/pii/0927650594000363 (cit. on p. 8).

[13] T Sanuki, M Motoki, H Matsumoto, ES Seo, JZ Wang, K Abe, K Anraku, Y Asaoka, M Fujikawa, M Imori, et al. Precise measurement of cosmic-ray proton and helium spectra with the BESS spectrometer. In: *The Astrophysical Journal* 545.2 (2000), p. 1135 (cit. on p. 9).

[14] HS Ahn, PS Allison, MG Bagliesi, JJ Beatty, G Bigongiari, PJ Boyle, TJ Brandt, JT Childers, NB Conklin, Stephane Coutu, et al. Measurements of cosmic-ray secondary nuclei at high energies with the first flight of the CREAM balloon-borne experiment. In: *Astroparticle Physics* 30.3 (2008), pp. 133–141 (cit. on p. 9).

[15] O Adriani, GC Barbarino, GA Bazilevskaya, R Bellotti, M Boezio, EA Bogomolov, M Bongi, V Bonvicini, S Borisov, S Bottai, et al. Cosmic-ray electron flux measured by the PAMELA experiment between 1 and 625 GeV. In: *Physical Review Letters* 106.20 (2011), p. 201101 (cit. on p. 9).

[16] M Aguilar, Giovanni Alberti, B Alpat, A Alvino, G Ambrosi, K Andeen, H Anderhub, L Arruda, P Azzarello, A Bachlechner, et al. First result from the Alpha Magnetic Spectrometer on the International Space Station: precision measurement of the positron fraction in primary cosmic rays of 0.5–350 GeV. In: *Physical Review Letters* 110.14 (2013), p. 141102 (cit. on p. 9).

[17] Particle Data Group. Review of Particle Physics. In: *Phys. Rev. D* 98.3 (2018), p. 030001 (cit. on p. 9).

[18] The Pierre Auger Collaboration. Features of the energy spectrum of cosmic rays above 2.5×1018 eV using the Pierre Auger Observatory. In: *Accepted for publication in PRL* (2020) (cit. on p. 9).

[19] J. Beatty and S. Westerhoff. The Highest-Energy Cosmic Rays. In: *Annual Review of Nuclear and Particle Science* 59 (Nov. 2009), pp. 319–345 (cit. on p. 9).

[20] B Peters. Primary cosmic radiation and extensive air showers. In: *Il Nuovo Cimento (1955-1965)* 22.4 (1961), pp. 800–819 (cit. on p. 10).

[21] W. D. Apel, J. C. Arteaga-Velàzquez, K. Bekk, M. Bertaina, J. Blümer, H. Bozdog, I. M. Brancus, E. Cantoni, A. Chiavassa, F. Cossavella, and et al. Ankle-like feature in the energy spectrum of light elements of cosmic rays observed with KASCADE-Grande. In: *Physical Review D* 87.8 (Apr. 2013). URL: http://dx.doi.org/10.1103/PhysRevD.87.081101 (cit. on p. 10).

[22] R Aloisio, V Berezinsky, and P Blasi. Ultra high energy cosmic rays: implications of Auger data for source spectra and chemical composition. In: *Journal of Cosmology and Astroparticle Physics* 2014.10 (2014), p. 020 (cit. on p. 10).

[23] Roberto Aloisio, Veniamin Berezinsky, Pasquale Blasi, Askhat Gazizov, Svetlana Grigorieva, and Bohdan Hnatyk. A dip in the UHECR spectrum and the transition from galactic to extragalactic cosmic rays. In: *Astroparticle Physics* 27.1 (2007), pp. 76–91 (cit. on p. 10).

[24] Glennys R Farrar, Michael Unger, and Luis A Anchordoqui. Origin of the ankle in the ultra-high energy cosmic ray spectrum and of the extragalactic protons below it. In: *arXiv preprint arXiv:1512.00484* (2015) (cit. on p. 10).

[25] Enrico Fermi. On the origin of the cosmic radiation. In: *Physical review* 75.8 (1949), p. 1169 (cit. on p. 11).

[26] JA Peacock. Fermi acceleration by relativistic shock waves. In: *Monthly Notices of the Royal Astronomical Society* 196.2 (1981), pp. 135–152 (cit. on p. 12).

[27] AM Bykov and IN Toptygin. Effect of shocks on interstellar turbulence and cosmic-ray dynamics. In: *Astrophysics and space science* 138.2 (1987), pp. 341–354 (cit. on p. 12).

[28] WI Axford. "The acceleration of galactic cosmic rays". In: *Symposium-International Astronomical Union*. Vol. 94. Cambridge University Press. 1981, pp. 339–358 (cit. on p. 12).

[29] Dimitrios Giannios. UHECRs from magnetic reconnection in relativistic jets. In: *Monthly Notices of the Royal Astronomical Society: Letters* 408.1 (2010), pp. L46–L50 (cit. on p. 13).

[30] Elisabete M. de Gouveia Dal Pino, Rafael Alves Batista, Grzergorz Kowal, Tania Medina-Torrejón, and Juan Carlos Ramirez-Rodriguez. Magnetic Reconnection, Cosmic Ray Acceleration, and Gamma-Ray emission around Black Holes and Relativistic Jets. In: *arXiv e-prints*, arXiv:1903.08982 (Mar. 2019), arXiv:1903.08982. arXiv: 1903.08982 [astro-ph.HE] (cit. on p. 13).

[31] Maxim Lyutikov and Rachid Ouyed. Inductive acceleration of UHECRs in sheared relativistic jets. In: *Astroparticle Physics* 27.6 (2007), pp. 473–489 (cit. on p. 13).

[32] Mikhail V Medvedev. Constraint on electromagnetic acceleration of highest energy cosmic rays. In: *Physical Review E* 67.4 (2003), p. 045401 (cit. on p. 13).

[33] Claire Guépin, Benoît Cerutti, and Kumiko Kotera. Proton acceleration in pulsar magnetospheres. In: *Astronomy & Astrophysics* 635 (2020), A138 (cit. on p. 13).

[34] M Takeda, N Hayashida, K Honda, N Inoue, K Kadota, F Kakimoto, K Kamata, S Kawaguchi, Y Kawasaki, N Kawasumi, et al. Extension of the cosmic-ray energy spectrum beyond the predicted Greisen-Zatsepin-Kuz'min cutoff. In: *Physical Review Letters* 81.6 (1998), p. 1163 (cit. on p. 13).

[35] Zoltán Fodor, Sándor D Katz, and Andreas Ringwald. Z-burst scenario for the highest energy cosmic rays. In: *arXiv preprint hep-ph/0210123* (2002) (cit. on p. 13).

[36] V Berezinsky, M Kachelriess, and A Vilenkin. Ultrahigh energy cosmic rays without greisen-zatsepin-kuzmin cutoff. In: *Physical Review Letters* 79.22 (1997), p. 4302 (cit. on p. 13).

[37] VA Kuzmin and VA Rubakov. Ultra-High Energy Cosmic Rays: a Window to Post-Inflationary Reheating Epoch of the Universe? In: *arXiv preprint astro-ph/9709187* (1997) (cit. on p. 13).

[38] M Kachelriess. The rise and fall of top-down models as main UHECR sources. In: *arXiv preprint arXiv:0810.3017* (2008) (cit. on p. 13).

[39] Alexander Vilenkin and E Paul S Shellard. *Cosmic strings and other topological defects*. Cambridge University Press, 2000 (cit. on p. 13).

[40] Pijushpani Bhattacharjee and Günter Sigl. Origin and propagation of extremely high-energy cosmic rays. In: *Physics Reports* 327.3-4 (2000), pp. 109–247 (cit. on p. 13).

[41] BR Dawson, M Fukushima, and P Sokolsky. Past, present, and future of UHECR observations. In: *Progress of Theoretical and Experimental Physics* 2017.12 (2017), 12A101 (cit. on p. 13).

[42] DV Semikoz. Constraints on top-down models for the origin of UHECRs from the Pierre Auger Observatory data. In: *arXiv preprint arXiv:0706.2960* (2007) (cit. on p. 13).

[43]   J Abraham, P Abreu, M Aglietta, C Aguirre, D Allard, I Allekotte, J Allen, P Allison, J Alvarez-Muniz, M Ambrosio, et al. Upper limit on the cosmic-ray photon flux above 1019 eV using the surface detector of the Pierre Auger Observatory. In: *Astroparticle Physics* 29.4 (2008), pp. 243–256 (cit. on p. 13).

[44]   A.M. Hillas. The Origin of Ultra-High-Energy Cosmic Rays. In: *Annual Review of Astronomy and Astrophysics* 22 (Jan. 1984), pp. 425–444 (cit. on p. 13).

[45]   Rafael Alves Batista, Jonathan Biteau, Mauricio Bustamante, Klaus Dolag, Ralph Engel, Ke Fang, Karl-Heinz Kampert, Dmitriy Kostunin, Miguel Mostafa, Kohta Murase, Foteini Oikonomou, Angela V. Olinto, Mikhail I. Panasyuk, Guenter Sigl, Andrew M. Taylor, and Michael Unger. Open Questions in Cosmic-Ray Research at Ultrahigh Energies. In: *Frontiers in Astronomy and Space Sciences* 6 (2019), p. 23. URL: https://www.frontiersin.org/article/10.3389/fspas.2019.00023 (cit. on p. 14).

[46]   Luis Alfredo Anchordoqui. Acceleration of ultrahigh-energy cosmic rays in starburst superwinds. In: *Physical Review D* 97.6 (2018), p. 063010 (cit. on p. 15).

[47]   Romero, G. E., Müller, A. L., and Roth, M. Particle acceleration in the superwinds of starburst galaxies. In: *A&A* 616 (2018), A57. URL: https://doi.org/10.1051/0004-6361/201832666 (cit. on p. 15).

[48]   Luis A Anchordoqui, Gustavo E Romero, and Jorge A Combi. Heavy nuclei at the end of the cosmic-ray spectrum? In: *Physical Review D* 60.10 (1999), p. 103001 (cit. on p. 15).

[49]   Mario Vietri. On the acceleration of ultra high energy cosmic rays in gamma ray bursts. In: *arXiv preprint astro-ph/9506081* (1995) (cit. on p. 15).

[50]   B Theodore Zhang, Kohta Murase, Shigeo S Kimura, Shunsaku Horiuchi, and Peter Mészáros. Low-luminosity gamma-ray bursts as the sources of ultrahigh-energy cosmic ray nuclei. In: *Physical Review D* 97.8 (2018), p. 083010 (cit. on p. 15).

[51]   Sayan Chakraborti, Alak Ray, AM Soderberg, Abraham Loeb, and Poonam Chandra. Ultra-high-energy cosmic ray acceleration in engine-driven relativistic supernovae. In: *Nature Communications* 2.1 (2011), pp. 1–6 (cit. on p. 15).

[52]   B Theodore Zhang and Kohta Murase. Ultrahigh-energy cosmic-ray nuclei and neutrinos from engine-driven supernovae. In: *Physical Review D* 100.10 (2019), p. 103004 (cit. on p. 15).

[53]   Matthew R George, Andrew C Fabian, Wayne H Baumgartner, Richard F Mushotzky, and Jack Tueller. On active galactic nuclei as sources of ultra-high energy cosmic rays. In: *Monthly Notices of the Royal Astronomical Society: Letters* 388.1 (2008), pp. L59–L63 (cit. on p. 15).

[54]   A Uryson. Identification of active galactic nuclei as possible sources of UHECR. In: *Astronomical and Astrophysical Transactions* 20.2 (2001), pp. 347–348 (cit. on p. 15).

[55]   Venya Berezinsky, AZ Gazizov, and SI Grigorieva. Signatures of AGN model for UHECR. In: *arXiv preprint astro-ph/0210095* (2002) (cit. on p. 15).

[56]   Claire Guépin, Kumiko Kotera, Enrico Barausse, Ke Fang, and Kohta Murase. Ultra-high-energy cosmic rays and neutrinos from tidal disruptions by massive black holes. In: *Astronomy & Astrophysics* 616 (2018), A179 (cit. on p. 15).

[57]   Eli Waxman. Cosmological gamma-ray bursts and the highest energy cosmic rays. In: *Physical Review Letters* 75.3 (1995), p. 386 (cit. on p. 15).

[58] B. Paczynski. Gamma-ray bursters at cosmological distances. In: 308 (Sept. 1986), pp. L43–L46 (cit. on p. 15).

[59] MJ Rees and P Mészáros. Relativistic fireballs: energy conversion and time-scales. In: *Monthly Notices of the Royal Astronomical Society* 258.1 (1992), 41P–43P (cit. on p. 15).

[60] Peter Meszaros. Gamma-ray bursts. In: *Reports on Progress in Physics* 69.8 (2006), p. 2259 (cit. on p. 15).

[61] OC De Jager. "Evidence for particle acceleration in a magnetized white dwarf from radio and gamma-ray observations". In: *International Astronomical Union Colloquium*. Vol. 142. Cambridge University Press. 1994, pp. 775–782 (cit. on p. 15).

[62] NR Ikhsanov and PL Biermann. High-energy emission of fast rotating white dwarfs. In: *Astronomy & Astrophysics* 445.1 (2006), pp. 305–312 (cit. on p. 15).

[63] Rafael Alves Batista and Joseph Silk. Ultrahigh-energy cosmic rays from tidally-ignited white dwarfs. In: *Physical Review D* 96.10 (2017), p. 103003 (cit. on p. 15).

[64] Jonathan Arons. Magnetars in the metagalaxy: an origin for ultra-high-energy cosmic rays in the nearby universe. In: *The Astrophysical Journal* 589.2 (2003), p. 871 (cit. on p. 15).

[65] Pasquale Blasi, Richard I Epstein, and Angela V Olinto. Ultra-high-energy cosmic rays from young neutron star winds. In: *The Astrophysical Journal Letters* 533.2 (2000), p. L123 (cit. on p. 15).

[66] Ke Fang, Kumiko Kotera, and Angela V Olinto. Newly born pulsars as sources of ultrahigh energy cosmic rays. In: *The Astrophysical Journal* 750.2 (2012), p. 118 (cit. on p. 15).

[67] Roger D Blandford and Roman L Znajek. Electromagnetic extraction of energy from Kerr black holes. In: *Monthly Notices of the Royal Astronomical Society* 179.3 (1977), pp. 433–456 (cit. on p. 15).

[68] Arman Tursunov, Zdeněk Stuchlík, Martin Kološ, Naresh Dadhich, and Bobomurat Ahmedov. Supermassive Black Holes as Possible Sources of Ultrahigh-energy Cosmic Rays. In: *The Astrophysical Journal* 895.1 (2020), p. 14 (cit. on p. 15).

[69] Leonel Morejon, Anatoli Fedynitch, Denise Boncioli, Daniel Biehl, and Walter Winter. Improved photomeson model for interactions of cosmic ray nuclei. In: *Journal of Cosmology and Astroparticle Physics* 2019.11 (2019), p. 007 (cit. on p. 15).

[70] R Alves Batista, M Erdmann, C Evoli, K-H Kampert, D Kuempel, G Müller, G Sigl, A van Vliet, D Walz, and T Winchen. "CRPropa: a public framework to propagate UHECRs in the universe". In: *EPJ Web of Conferences*. Vol. 99. EDP Sciences. 2015, p. 13004 (cit. on p. 16).

[71] Jörg Paul Rachen. "Interaction processes and statistical properties of the propagation of cosmic rays in photon backgrounds". PhD thesis. Rheinische Friedrich-Wilhelms-Universität Bonn, 1996 (cit. on p. 16).

[72] Christopher Heiter, Daniel Kuempel, David Walz, and Martin Erdmann. Production and propagation of ultra-high energy photons using CRPropa 3. In: *Astroparticle Physics* 102 (2018), pp. 39–50 (cit. on pp. 16, 17, 19).

[73] Denis Allard. Extragalactic propagation of ultrahigh energy cosmic-rays. In: *Astroparticle Physics* 39 (2012), pp. 33–43 (cit. on pp. 16, 17).

[74]   The Pierre Auger Collaboration. In: () (cit. on pp. 16, 48).

[75]   Ruth Durrer and Andrii Neronov. Cosmological magnetic fields: their generation, evolution and observation. In: *The Astronomy and Astrophysics Review* 21.1 (2013), p. 62 (cit. on p. 17).

[76]   Kandaswamy Subramanian. The origin, evolution and signatures of primordial magnetic fields. In: *Reports on Progress in Physics* 79.7 (2016), p. 076901 (cit. on p. 17).

[77]   Ronnie Jansson and Glennys R Farrar. A new model of the galactic magnetic field. In: *The Astrophysical Journal* 757.1 (2012), p. 14 (cit. on p. 17).

[78]   R Adam, PAR Ade, MIR Alves, M Ashdown, J Aumont, C Baccigalupi, AJ Banday, RB Barreiro, N Bartolo, E Battaner, et al. Planck intermediate results-XLII. Large-scale Galactic magnetic fields. In: *Astronomy Astrophysics* 596 (2016), A103 (cit. on p. 17).

[79]   Glennys R Farrar and Michael S Sutherland. Deflections of UHECRs in the Galactic magnetic field. In: *Journal of Cosmology and Astroparticle Physics* 2019.05 (2019), p. 004 (cit. on p. 17).

[80]   Olivier Deligny. Measurements and implications of cosmic ray anisotropies from TeV to trans-EeV energies. In: *Astropart. Phys.* 104 (2019), pp. 13–41. arXiv: 1808.03940 [astro-ph.HE] (cit. on p. 17).

[81]   Marek Sikora and Isaac Shlosman. Ultra–High-Energy Photons from Active Galactic Nuclei: Theory. In: *The Astrophysical Journal* 336 (Jan. 1989), p. 593 (cit. on p. 17).

[82]   P.L. Biermann and P.A. Strittmatter. Synchrotron Emission from Shock Waves in Active Galactic Nuclei. In: *The Astrophysical Journal* 322 (Nov. 1987), p. 643 (cit. on p. 17).

[83]   Arno A Penzias and Robert Woodrow Wilson. A measurement of excess antenna temperature at 4080 Mc/s. In: *The Astrophysical Journal* 142 (1965), pp. 419–421 (cit. on p. 18).

[84]   Robert H Dicke, P James E Peebles, Peter G Roll, and David T Wilkinson. Cosmic Black-Body Radiation. In: *The Astrophysical Journal* 142 (1965), pp. 414–419 (cit. on p. 18).

[85]   BISWAJIT Sarkar, KARL-HEINZ Kampert, JOERG Kulbartz, LUCA MACCIONE, NILS NIERSTENHOEFER, and GUENTER SIGL. "Ultra-high energy photon and neutrino fluxes in realistic astrophysical scenarios". In: *Proceedings of the 32nd Int. Cosmic Ray Conf.* Vol. 2. 2011, p. 198 (cit. on p. 20).

[86]   Graciela B Gelmini, Oleg E Kalashev, and Dmitry V Semikoz. GZK photons as ultra-high-energy cosmic rays. In: *Journal of Experimental and Theoretical Physics* 106.6 (2008), p. 1061 (cit. on p. 20).

[87]   R. U. Abbasi, T. Abu-Zayyad, M. Allen, J. F. Amman, G. Archbold, K. Belov, J. W. Belz, S. Y. Ben Zvi, D. R. Bergman, S. A. Blake, O. A. Brusova, G. W. Burt, C. Cannon, Z. Cao, B. C. Connolly, W. Deng, Y. Fedorova, C. B. Finley, R. C. Gray, W. F. Hanlon, C. M. Hoffman, M. H. Holzscheiter, G. Hughes, P. Hüntemeyer, B. F Jones, C. C. H. Jui, K. Kim, M. A. Kirn, E. C. Loh, M. M. Maestas, N. Manago, L. J. Marek, K. Martens, J. A. J. Matthews, J. N. Matthews, S. A. Moore, A. O'Neill, C. A. Painter, L. Perera, K. Reil, R. Riehle, M. Roberts, D. Rodriguez, N. Sasaki, S. R. Schnetzer, L. M. Scott, G. Sinnis, J. D. Smith, P. Sokolsky, C. Song, R. W. Springer, B. T. Stokes, S. B. Thomas, J. R. Thomas, G. B. Thomson, D. Tupa, S.

Westerhoff, L. R. Wiencke, X. Zhang, and A. Zech. First Observation of the Greisen-Zatsepin-Kuzmin Suppression. In: *Phys. Rev. Lett.* 100 (10 Mar. 2008), p. 101101. URL: https://link.aps.org/doi/10.1103/PhysRevLett.100.101101 (cit. on p. 20).

[88]   Erich Regener and Georg Pfotzer. Vertical intensity of cosmic rays by threefold coincidences in the stratosphere. In: *Nature* 136.3444 (1935), pp. 718–719 (cit. on p. 22).

[89]   Paul Adrien Maurice Dirac. The quantum theory of the emission and absorption of radiation. In: *Proceedings of the Royal Society of London. Series A, Containing Papers of a Mathematical and Physical Character* 114.767 (1927), pp. 243–265 (cit. on p. 22).

[90]   Enrico Fermi. Quantum theory of radiation. In: *Reviews of modern physics* 4.1 (1932), p. 87 (cit. on p. 22).

[91]   Hans Bethe and Walter Heitler. On the stopping of fast particles and on the creation of positive electrons. In: *Proceedings of the Royal Society of London. Series A, Containing Papers of a Mathematical and Physical Character* 146.856 (1934), pp. 83–112 (cit. on p. 22).

[92]   Walter Heitler. *The quantum theory of radiation.* Courier Corporation, 1984 (cit. on p. 22).

[93]   Homi J Bhabha and Walther Heitler. The passage of fast electrons and the theory of cosmic showers. In: *Proceedings of the Royal Society of London. Series A-Mathematical and Physical Sciences* 159.898 (1937), pp. 432–458 (cit. on p. 22).

[94]   JF Carlson and JR Oppenheimer. On multiplicative showers. In: *Physical Review* 51.4 (1937), p. 220 (cit. on p. 22).

[95]   James Matthews. A Heitler model of extensive air showers. In: *Astroparticle Physics* 22.5-6 (2005), pp. 387–397 (cit. on pp. 22–25).

[96]   Paul Sommers. Extensive air showers and measurement techniques. In: *Comptes Rendus Physique* 5.4 (2004), pp. 463–472 (cit. on p. 22).

[97]   Karl-Heinz Kampert and Michael Unger. Measurements of the cosmic ray composition with air shower experiments. In: *Astroparticle Physics* 35.10 (2012), pp. 660–678 (cit. on p. 24).

[98]   Dieter IKP Heck. KIT-CORSIKA-Documentation. In: (2020) (cit. on pp. 26, 27).

[99]   Ralph Engel, Dieter Heck, and Tanguy Pierog. Extensive air showers and hadronic interactions at high energy. In: *Annual review of nuclear and particle science* 61 (2011), pp. 467–489 (cit. on pp. 26, 29, 30).

[100] Walter Ralph Nelson, Hideo Hirayama, and David WO Rogers. *EGS4 code system.* Tech. rep. Stanford Linear Accelerator Center, Menlo Park, CA (USA), 1985 (cit. on p. 26).

[101] S. Agostinelli et al. Geant4—a simulation toolkit. In: *Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment* 506.3 (2003), pp. 250 –303. URL: http://www.sciencedirect.com/science/article/pii/S0168900203013688 (cit. on p. 26).

[102] Tanguy Pierog and Klaus Werner. EPOS model and ultra high energy cosmic rays. In: *Nuclear Physics B-Proceedings Supplements* 196 (2009), pp. 102–105 (cit. on p. 27).

[103] Tanguy Pierog, Benjamin Guiot, Iurii Karpenko, Gabriel Sophys, Maria Stefaniak, and Klaus Werner. "EPOS 3 and air showers". In: *EPJ Web of Conferences.* Vol. 210. EDP Sciences. 2019, p. 02008 (cit. on p. 27).

[104]   S Ostapchenko. QGSJET-II: towards reliable description of very high energy hadronic interactions. In: *Nuclear Physics B-Proceedings Supplements* 151.1 (2006), pp. 143–146 (cit. on p. 27).

[105]   Felix Riehn, Hans P Dembinski, Ralph Engel, Anatoli Fedynitch, Thomas K Gaisser, and Todor Stanev. The hadronic interaction model SIBYLL 2.3 c and Feynman scaling. In: *arXiv preprint arXiv:1709.07227* (2017) (cit. on p. 27).

[106]   Karl-Heinz Kampert and Michael Unger. Measurements of the cosmic ray composition with air shower experiments. In: *Astroparticle Physics* 35.10 (2012), pp. 660–678 (cit. on p. 27).

[107]   Lorenzo Cazon. "Probing High-Energy Hadronic Interactions with Extensive Air Showers". In: *Proceedings of 36th International Cosmic Ray Conference — PoS(ICRC2019)*. Vol. 358. 2019, p. 005 (cit. on pp. 27, 54).

[108]   Alexander Aab, P Abreu, MARCO Aglietta, EJ Ahn, Imen Al Samarai, IFM Albuquerque, I Allekotte, JD Allen, P Allison, A Almela, et al. Testing hadronic interactions at ultrahigh energies with air showers measured by the Pierre Auger Observatory. In: *Physical review letters* 117.19 (2016), p. 192001 (cit. on p. 27).

[109]   Giuseppe Battistoni, Till Boehlen, Francesco Cerutti, Pik Wai Chin, Luigi Salvatore Esposito, Alberto Fassò, Alfredo Ferrari, Anton Lechner, Anton Empl, Andrea Mairani, et al. Overview of the FLUKA code. In: *Annals of Nuclear Energy* 82 (2015), pp. 10–18 (cit. on p. 27).

[110]   H Fesefeldt. Report PITHA-85/02. In: *RWTH Aachen* 647 (1985) (cit. on p. 27).

[111]   Marcus Bleicher, E Zabrodin, Christian Spieles, Steffen A Bass, Christoph Ernst, Sven Soff, L Bravina, Mohamed Belkacem, Henning Weber, Horst Stöcker, et al. Relativistic hadron-hadron collisions in the ultra-relativistic quantum molecular dynamics model. In: *Journal of Physics G: Nuclear and Particle Physics* 25.9 (1999), p. 1859 (cit. on p. 27).

[112]   Dieter Heck, J Knapp, JN Capdevielle, G Schatz, T Thouw, et al. CORSIKA: A Monte Carlo code to simulate extensive air showers. In: *Report fzka* 6019.11 (1998) (cit. on p. 27).

[113]   SJ Sciutto. AIRES: A system for air shower simulations. In: *arXiv preprint astro-ph/9911331* (1999) (cit. on p. 27).

[114]   SJ Sciutto. The AIRES system for air shower simulations. An update. In: *arXiv preprint astro-ph/0106044* (2001) (cit. on p. 27).

[115]   Jeff Allen, Antonella Castellina, Ralph Engel, Katsuaki Kasahara, Stanislav Knurenko, Tanguy Pierog, Artem Sabourov, Benjamin T Stokes, Ralf Ulrich, Takashi Sako, et al. "Air shower simulation and hadronic interactions". In: *EPJ Web of Conferences*. Vol. 53. EDP Sciences. 2013, p. 01007 (cit. on p. 27).

[116]   Till Bergmann, Ralph Engel, Dieter Heck, NN Kalmykov, Sergey Ostapchenko, Tanguy Pierog, T Thouw, and Klaus Werner. One-dimensional hybrid approach to extensive air shower simulation. In: *Astroparticle Physics* 26.6 (2007), pp. 420–432 (cit. on p. 27).

[117]   Hans-Joachim Drescher and Glennys R Farrar. Air shower simulations in a hybrid approach using cascade equations. In: *Physical Review D* 67.11 (2003), p. 116001 (cit. on p. 27).

[118] Lorenzo Cazon. *Working Group Report on the Combined Analysis of Muon Density Measurements from Eight Air Shower Experiments.* 2020. arXiv: 2001.07508 [astro-ph.HE] (cit. on pp. 28, 55).

[119] Alexander Aab, P Abreu, MARCO Aglietta, EJ Ahn, I Al Samarai, IFM Albuquerque, I Allekotte, J Allen, P Allison, A Almela, et al. Muons in air showers at the Pierre Auger Observatory: Mean number in highly inclined events. In: *Physical Review D* 91.3 (2015), p. 032003 (cit. on p. 28).

[120] M. Risse, P. Homola, R. Engel, D. Góra, D. Heck, J. Pekala, B. Wilczyńska, and H. Wilczyński. Photon air showers at ultra-high energy and the photonuclear cross-section. In: *Czechoslovak Journal of Physics* 56.S1 (Sept. 2006), A327–A336. URL: http://dx.doi.org/10.1007/s10582-006-0166-7 (cit. on p. 28).

[121] Thomas Erber. High-energy electromagnetic conversion processes in intense magnetic fields. In: *Reviews of Modern Physics* 38.4 (1966), p. 626 (cit. on p. 29).

[122] B McBreen and CJ Lambert. Interactions of high-energy (E> 5× 10 19 eV) photons in the earth's magnetic field. In: *Physical Review D* 24.9 (1981), p. 2536 (cit. on p. 29).

[123] LD Landau and I Ya Pomeranchuk. Limits of the Applicability of the Theory of Electron Bremsstrahlung and Pair Production at High Energies Dokl. In: *Akad. Nauk SSSR* 92 (1953), p. 535 (cit. on p. 30).

[124] Arkady B Migdal. Bremsstrahlung and pair production in condensed media at high energies. In: *Physical Review* 103.6 (1956), p. 1811 (cit. on p. 30).

[125] AN Cillis, H Fanchiotti, CA Garcia Canal, and SJ Sciutto. Influence of the LPM effect and dielectric suppression on particle air showers. In: *Physical Review D* 59.11 (1999), p. 113012 (cit. on p. 30).

[126] X Bertou, P Billoir, and S Dagoret-Campagne. LPM effect and pair production in the geomagnetic field: a signature of ultra-high energy photons in the Pierre Auger Observatory. In: () (cit. on p. 30).

[127] Ralf Ulrich, Ralph Engel, and Michael Unger. Hadronic multiparticle production at ultrahigh energies and extensive air showers. In: *Physical Review D* 83.5 (2011), p. 054026 (cit. on pp. 31, 118).

[128] Hans-Joachim Drescher. Remnant break-up and muon production in cosmic ray air showers. In: *Physical Review D* 77.5 (2008), p. 056003 (cit. on p. 31).

[129] MA Lawrence, RJO Reid, and AA Watson. The cosmic ray energy spectrum above 4* 1017 eV as measured by the Haverah Park array. In: *Journal of Physics G: Nuclear and Particle Physics* 17.5 (1991), p. 733 (cit. on p. 34).

[130] J Abraham, M Aglietta, IC Aguirre, M Albrow, D Allard, I Allekotte, P Allison, J Alvarez Muniz, MG Do Amaral, M Ambrosio, et al. Properties and performance of the prototype instrument for the Pierre Auger Observatory. In: *Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment* 523.1-2 (2004), pp. 50–95 (cit. on p. 34).

[131] The Pierre Auger collaboration. *Reconstruction of Events Recorded with the Surface Detector of the Pierre Auger Observatory.* 2020. arXiv: 2007.09035 [astro-ph.IM] (cit. on p. 36).

[132] The Pierre Auger collaboration. The Pierre Auger Observatory scaler mode for the study of solar activity modulation of galactic cosmic rays. In: *Journal of Instrumentation* 6.01 (Jan. 2011), P01003–P01003. URL: https://doi.org/10.1088%2F1748-0221%2F6%2F01%2Fp01003 (cit. on p. 37).

[133] Pierre Auger Collaboration et al. The Pierre Auger cosmic ray observatory. In: *Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment* 798 (2015), pp. 172–213 (cit. on pp. 37, 42, 57).

[134] The Pierre Auger collaboration. Depth of Maximum of Air-Shower Profiles at the Pierre Auger Observatory: Measurements at Energies above $10^{17.8}$ eV. In: *Phys. Rev. D* 90.12 (2014), p. 122005. arXiv: 1409.4809 [astro-ph.HE] (cit. on p. 39).

[135] Alexander Aab, P Abreu, M Aglietta, IFM Albuquerque, JM Albury, I Allekotte, A Almela, J Alvarez Castillo, J Alvarez-Muñiz, GA Anastasi, et al. Data-driven estimation of the invisible energy of cosmic ray showers with the Pierre Auger Observatory. In: *Physical Review D* 100.8 (2019), p. 082003 (cit. on p. 40).

[136] Jorge Abraham, P Abreu, M Aglietta, C Aguirre, EJ Ahn, D Allard, I Allekotte, J Allen, P Allison, J Alvarez-Muñiz, et al. The fluorescence detector of the Pierre Auger Observatory. In: *Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment* 620.2-3 (2010), pp. 227–251 (cit. on p. 40).

[137] Fred James. *Minuit: Function minimization and error analysis reference manual.* Tech. rep. Cern, 1998 (cit. on p. 41).

[138] Koichi Kamata and Jun Nishimura. The lateral and the angular structure functions of electron showers. In: *Progress of Theoretical Physics Supplement* 6 (1958), pp. 93–155 (cit. on p. 42).

[139] K Greisen. The extensive air showers. In: *Progress in cosmic ray physics* 3.1 (1956) (cit. on p. 42).

[140] Kenneth Greisen. Cosmic ray showers. In: *Annual Review of Nuclear Science* 10.1 (1960), pp. 63–108 (cit. on p. 42).

[141] Thomas K Gaisser and A Michael Hillas. "Reliability of the method of constant intensity cuts for reconstructing the average development of vertical showers". In: *International Cosmic Ray Conference.* Vol. 8. 1977, pp. 353–357 (cit. on p. 43).

[142] Valerio Verzi. "Measurement of the energy spectrum of ultra-high energy cosmic rays using the Pierre Auger Observatory". In: *Proceedings of 36th International Cosmic Ray Conference — PoS(ICRC2019).* Vol. 358. 2019, p. 450 (cit. on pp. 43, 45, 46).

[143] J Abraham, P Abreu, M Aglietta, EJ Ahn, D Allard, I Allekotte, J Allen, J Alvarez-Muñiz, M Ambrosio, L Anchordoqui, et al. Trigger and aperture of the surface detector array of the Pierre Auger Observatory. In: *Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment* 613.1 (2010), pp. 29–39 (cit. on p. 44).

[144] Alexander Aab, P Abreu, M Aglietta, IFM Albuquerque, JM Albury, I Allekotte, A Almela, J Alvarez Castillo, J Alvarez-Muñiz, GA Anastasi, et al. The Pierre Auger Observatory: Contributions to the 36th International Cosmic Ray Conference (ICRC 2019). In: *arXiv preprint arXiv:1909.09073* (2019) (cit. on p. 45).

[145] The Pierre Auger Collaboration. *A measurement of the cosmic-ray energy spectrum above* 2.5×10^{18} *eV using the Pierre Auger Observatory.* 2020. arXiv: 2008.06486 [astro-ph.HE] (cit. on p. 45).

[146]   The Pierre Auger Collaboration. *Features of the energy spectrum of cosmic rays above* 2.5×10^{18} *eV using the Pierre Auger Observatory*. 2020. arXiv: 2008.06488 [astro-ph.HE] (cit. on p. 46).

[147]   Alexey Yushkov. "Mass Composition of Cosmic Rays with Energies above $10^{17.2}$ eV from the Hybrid Data of the Pierre Auger Observatory". In: *Proceedings of 36th International Cosmic Ray Conference — PoS(ICRC2019)*. Vol. 358. 2019, p. 482 (cit. on p. 47).

[148]   The Pierre Auger Collaboration. In: 2017.04 (Apr. 2017), pp. 038–038. URL: https://doi.org/10.1088%2F1475-7516%2F2017%2F04%2F038 (cit. on p. 47).

[149]   Antonella Castellina. "Highlights from the Pierre Auger Observatory". In: *Proceedings of 36th International Cosmic Ray Conference — PoS(ICRC2019)*. Vol. 358. 2019, p. 004 (cit. on p. 47).

[150]   The Pierre Auger Collaboration. In: 2019.11 (Nov. 2019), pp. 004–004. URL: https://doi.org/10.1088%2F1475-7516%2F2019%2F11%2F004 (cit. on pp. 48, 50).

[151]   Lorenzo Caccianiga. "Anisotropies of the Highest Energy Cosmic-ray Events Recorded by the Pierre Auger Observatory in 15 years of Operation". In: *Proceedings of 36th International Cosmic Ray Conference — PoS(ICRC2019)*. Vol. 358. 2019, p. 206 (cit. on p. 48).

[152]   IceCube Collaboration. The IceCube Neutrino Observatory: Instrumentation and Online Systems. In: *JINST* 12.03 (2017), P03012. arXiv: 1612.05093 [astro-ph.IM] (cit. on p. 49).

[153]   ANTARES Collaboration. ANTARES: The first undersea neutrino telescope. In: *Nuclear Instruments and Methods in Physics Research A* 656.1 (Nov. 2011), pp. 11–38. arXiv: 1104.1607 [astro-ph.IM] (cit. on p. 49).

[154]   S. Aiello et al. Sensitivity of the KM3NeT/ARCA neutrino telescope to point-like neutrino sources. In: *Astropart. Phys.* 111 (2019), pp. 100–110. arXiv: 1810.08499 [astro-ph.HE] (cit. on p. 49).

[155]   Pierre Auger Collaboration. Improved limit to the diffuse flux of ultrahigh energy neutrinos from the Pierre Auger Observatory. In: *Phys. Rev.* D91.9 (2015), p. 092008. arXiv: 1504.05397 [astro-ph.HE] (cit. on p. 50).

[156]   Pierre Auger Collaboration. Probing the origin of ultra-high-energy cosmic rays with neutrinos in the EeV energy range using the Pierre Auger Observatory. In: *JCAP* 10.10 (2019), p. 022. arXiv: 1906.07422 [astro-ph.HE] (cit. on p. 50).

[157]   Arnauld Albert, Michel André, Marco Anghinolfi, M Ardid, J-J Aubert, J Aublin, T Avgitas, B Baret, J Barrios-Martí, S Basa, et al. Search for high-energy neutrinos from binary neutron star merger GW170817 with ANTARES, IceCube, and the Pierre Auger Observatory. In: *arXiv preprint arXiv:1710.05839* (2017) (cit. on p. 51).

[158]   Alexander Aab et al. The Pierre Auger Observatory Upgrade - Preliminary Design Report. In: (2016). arXiv: 1604.03637 [astro-ph.IM] (cit. on pp. 51, 57, 60, 122).

[159]   RJ Protheroe. Origin and propagation of the highest energy cosmic rays. In: *arXiv preprint astro-ph/9612212* (1996) (cit. on p. 54).

[160]   R. Aloisio, V. Berezinsky, and P. Blasi. Ultra high energy cosmic rays: implications of Auger data for source spectra and chemical composition. In: *Journal of Cosmology and Astroparticle Physics* 2014.10 (Oct. 2014), pp. 020–020. URL: https://doi.org/10.1088%2F1475-7516%2F2014%2F10%2F020 (cit. on p. 54).

[161] Sarah Müller, Ralph Engel, Tanguy Pierog, and Markus Roth. Impact of muon detection thresholds on the separability of primary cosmic rays. In: *PoS* ICRC2017 (2018), p. 311 (cit. on p. 54).

[162] Alexander Aab et al. Prototype muon detectors for the AMIGA component of the Pierre Auger Observatory. In: *JINST* 11.02 (2016), P02012. arXiv: 1605.01625 [physics.ins-det] (cit. on p. 55).

[163] Alvaro Taboada. Preliminary results of the AMIGA engineering array at the Pierre Auger Observatory. In: *EPJ Web Conf.* 210 (2019), p. 02016. arXiv: 1905.04213 [physics.ins-det] (cit. on p. 55).

[164] DF Nitz. "New Electronics for the Surface Detectors of the Pierre Auger Observatory". In: *36th International Cosmic Ray Conference (ICRC2019)*. Vol. 36. 2019 (cit. on p. 56).

[165] Antonella Castellina. The dynamic range of the AugerPrime Surface Detector: technical solution and physics reach. In: *PoS* (2017), pp. 161–168 (cit. on p. 56).

[166] M Aglietta, PS Allison, R Assiro, James Beatty, K-H Becker, Corinne Berat, J Bouvier, DA Brandt, P Buchholz, CE Covault, et al. "New electronics for the surface detectors of the Pierre Auger Observatory". In: *The 34th International Cosmic Ray Conference*. Vol. 236. SISSA Medialab. 2016, p. 565 (cit. on p. 56).

[167] Bjarni Pont. "A Large Radio Detector at the Pierre Auger Observatory-Measuring the Properties of Cosmic Rays up to the Highest Energies". In: *36th International Cosmic Ray Conference*. Vol. 358. SISSA Medialab. 2019, p. 395 (cit. on p. 56).

[168] Alexander Aab et al. Observation of inclined EeV air showers with the radio detector of the Pierre Auger Observatory. In: *JCAP* 1810 (2018), p. 026. arXiv: 1806.05386 [astro-ph.IM] (cit. on p. 56).

[169] Marvin Gottowik. "Measurements of Inclined Air Showers with the Auger Engineering Radio Array at the Pierre Auger Observatory". In: *36th International Cosmic Ray Conference (ICRC2019)*. Vol. 36. 2019 (cit. on p. 56).

[170] N. Chiba, K. Hashimoto, N. Hayashida, K. Honda, M. Honda, N. Inoue, F. Kakimoto, K. Kamata, S. Kawaguchi, N. Kawasumi, Y. Matsubara, K. Murakami, M. Nagano, S. Ogio, H. Ohoka, To. Saito, Y. Sakuma, I. Tsushima, M. Teshima, T. Umezawa, S. Yoshida, and H. Yoshii. Akeno Giant Air Shower Array (AGASA) covering 100 km2 area. In: *Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment* 311.1 (1992), pp. 338 –349. URL: http://www.sciencedirect.com/science/article/pii/0168900292908825 (cit. on p. 56).

[171] K. H. Kampert et al. "First results from the KASCADE air shower experiment". In: *Very high-energy phenomena in the universe. Proceedings, 32nd Rencontres de Moriond, Les Arcs, France, January 18-25, 1997*. 1997, pp. 405–410. arXiv: astro-ph/9703182 [astro-ph] (cit. on p. 56).

[172] Yakutsk collaboration. The spectrum features of UHECRs below and surrounding GZK. In: *Nuclear Physics B - Proceedings Supplements* 136 (2004). CRIS 2004 Proceedings of the Cosmic Ray International Seminars: GZK and Surroundings, pp. 3 –11. URL: http://www.sciencedirect.com/science/article/pii/S092056320400444X (cit. on p. 56).

[173] Dmitriy Beznosko, Alan Bross, A. Dyshkant, and P.-D.V. Rykalin. "FNAL-NICADD extruded scintillator". In: vol. 2. Nov. 2004, 790 –793 Vol. 2 (cit. on p. 57).

[174] Kuraray. *Technical Sheets for the Wavelenght Shifting Fibers*. Tech. rep. 2020. URL: http://kuraraypsf.jp/psf/ws.html (cit. on p. 57).

[175] Hamamatsu. *Technical Sheet PMT Hamamatsu R9420*. Tech. rep. 2014. URL: https://www.hamamatsu.com/resources/pdf/etd/R9420_TPMH1296E.pdf (cit. on p. 57).

[176] Photonis. *Photomultiplier tubes: Principle and Applications*. 2002. URL: https://psec.uchicago.edu/library/photomultipliers/Photonis_PMT_basics.pdf (cit. on p. 58).

[177] M. Ave, R. Engel, M. Roth, and A. Schulz. A generalized description of the signal size in extensive air shower detectors and its applications. In: *Astroparticle physics* 87 (2017). 51.03.04; LK 01, pp. 23–39 (cit. on pp. 60, 61).

[178] J Pekala. "Production and Quality Control of the Scintillator Surface Detector for the AugerPrime Upgrade of the Pierre Auger Observatory". In: *36th International Cosmic Ray Conference (ICRC2019)*. Vol. 36. 2019 (cit. on p. 61).

[179] Breton D. and Maalmi J. *WaveCatcher Family User's Manual*. Jan. 2017. URL: http://www.hep.ucl.ac.uk/pbt/wikiData/manuals/WaveCatcher/WaveCatcherFamily_V1.2.pdf (cit. on p. 67).

[180] Breton D. and Maalmi J. *WaveCatcher Software User's Manual*. Jan. 2017. URL: http://www.hep.ucl.ac.uk/pbt/wikiData/manuals/WaveCatcher/WaveCatcherSoftware_V1.1.pdf (cit. on p. 68).

[181] Alexander Aab et al. Measurement of the average shape of longitudinal profiles of cosmic-ray air showers at the Pierre Auger Observatory. In: *JCAP* 03 (2019), p. 018. arXiv: 1811.04660 [astro-ph.HE] (cit. on p. 78).

[182] AA Watson and JG Wilson. Fluctuation studies of large air showers: the composition of primary cosmic ray particles of energy $E_p > 10^{18} eV$. In: *Journal of Physics A: Mathematical, Nuclear and General* 7.10 (1974), p. 1199 (cit. on p. 79).

[183] Alexander Aab, P Abreu, MARCO Aglietta, EJ Ahn, I Al Samarai, IFM Albuquerque, I Allekotte, P Allison, A Almela, J Alvarez Castillo, et al. Azimuthal asymmetry in the risetime of the surface detector signals of the Pierre Auger Observatory. In: *Physical Review D* 93.7 (2016), p. 072006 (cit. on pp. 79, 81).

[184] Alexander Aab, Pedro Abreu, Marco Aglietta, Imen Al Samarai, IFM Albuquerque, Ingomar Allekotte, Alejandro Almela, J Alvarez Castillo, Jaime Alvarez-Muniz, Gioacchino Alex Anastasi, et al. Inferences on mass composition and tests of hadronic interactions from 0.3 to 100 EeV using the water-Cherenkov detectors of the Pierre Auger Observatory. In: *Physical Review D* 96.12 (2017), p. 122003 (cit. on pp. 81, 82, 110).

[185] Nicole Krohm, Carla Bleve, and Rautenberg Julian Kampert Karl-Heinz. "GAP 2012-059" (cit. on p. 83).

[186] G. Ros, A.D. Supanitsky, G.A. Medina-Tanco, L. del Peral, J.C. D'Olivo, and M.D. Rodríguez Frías. A new composition-sensitive parameter for ultra-high energy cosmic rays. In: *Astroparticle Physics* 35.3 (Oct. 2011), 140–151. URL: http://dx.doi.org/10.1016/j.astropartphys.2011.06.011 (cit. on p. 83).

[187] Pierre Auger Collaboration. Search for photons with energies above 10^{18} eV using the hybrid detector of the Pierre Auger Observatory. In: *JCAP* 1704.04 (2017), p. 009. arXiv: 1612.01517 [astro-ph.HE] (cit. on pp. 83, 90).

[188] Alexander Aab, Pedro Abreu, Marco Aglietta, Eun-Joo Ahn, I Al Samarai, IFM Albuquerque, Ingomar Allekotte, Patrick Allison, Alejandro Almela, J Alvarez Castillo, et al. Improved limit to the diffuse flux of ultrahigh energy neutrinos from the Pierre Auger Observatory. In: *Physical Review D* 91.9 (2015), p. 092008 (cit. on p. 83).

[189] M Ave, JA Hinton, RA Vazquez, AA Watson, and E Zas. Constraints on the ultrahigh-energy photon flux using inclined showers from the Haverah Park array. In: *Physical Review D* 65.6 (2002), p. 063007 (cit. on p. 84).

[190] Markus Risse, P Homola, R Engel, D Góra, D Heck, J Pękala, B Wilczyńska, and H Wilczyński. Upper limit on the photon fraction in highest-energy cosmic rays from AGASA data. In: *Physical review letters* 95.17 (2005), p. 171102 (cit. on pp. 84, 85).

[191] AV Glushkov, IT Makarov, MI Pravdin, IE Sleptsov, DS Gorbunov, GI Rubtsov, and SV Troitsky. Constraints on the flux of primary cosmic-ray photons at energies E> 10 18 eV from Yakutsk muon data. In: *Physical Review D* 82.4 (2010), p. 041101 (cit. on pp. 85, 86).

[192] RU Abbasi, M Abe, T Abu-Zayyad, M Allen, R Azuma, E Barcikowski, JW Belz, DR Bergman, SA Blake, R Cady, et al. Constraints on the diffuse photon flux with energies above 1018 eV using the surface detector of the Telescope Array experiment. In: *Astroparticle physics* 110 (2019), pp. 8–14 (cit. on pp. 86, 87, 91).

[193] Pierre Auger Collaboration. Upper limit on the cosmic-ray photon flux above 10^{19} eV using the surface detector of the Pierre Auger Observatory. In: *Astropart. Phys.* 29 (2008), pp. 243–256. arXiv: 0712.1147 [astro-ph] (cit. on p. 87).

[194] Carla Bleve. "Updates on the neutrino and photon limits from the Pierre Auger Observatory". In: *The 34th International Cosmic Ray Conference*. Vol. 236. SISSA Medialab. 2016, p. 1103 (cit. on pp. 88–90).

[195] Julian Rautenberg. "Limits on ultra-high energy photons with the Pierre Auger Observatory". In: *Proceedings of 36th International Cosmic Ray Conference — PoS(ICRC2019)*. Vol. 358. 2019, p. 398 (cit. on pp. 88, 91).

[196] A. Aab et al. Data-driven estimation of the invisible energy of cosmic ray showers with the Pierre Auger Observatory. In: *Phys. Rev. D* 100 (8 Oct. 2019), p. 082003. URL: https://link.aps.org/doi/10.1103/PhysRevD.100.082003 (cit. on p. 90).

[197] A. Aab et al. Search for photons with energies above 1018eV using the hybrid detector of the Pierre Auger Observatory. In: *Journal of Cosmology and Astroparticle Physics* 2017.04 (Apr. 2017), pp. 009–009. URL: https://doi.org/10.1088%2F1475-7516%2F2017%2F04%2F009 (cit. on p. 91).

[198] Yu A Fomin, NN Kalmykov, IS Karpikov, GV Kulikov, M Yu Kuznetsov, GI Rubtsov, VP Sulakov, and SV Troitsky. Constraints on the flux of $(10^{16} - 10^{17.5})$ eV cosmic photons from the EAS–MSU muon data. In: *Physical Review D* 95.12 (2017), p. 123011 (cit. on p. 91).

[199] WD Apel, JC Arteaga-Velázquez, K Bekk, M Bertaina, J Blümer, H Bozdog, IM Brancus, E Cantoni, A Chiavassa, F Cossavella, et al. KASCADE-Grande limits on the isotropic diffuse gamma-ray flux between 100 TeV and 1 EeV. In: *The Astrophysical Journal* 848.1 (2017), p. 1 (cit. on p. 91).

[200] P Billoir and M Settimo. "The muonicity observable for photon discrimination using SD traces". GAP2014-112. 2014 (cit. on p. 92).

[201] A Castellina and G Navara. "The muonic/em ratio and the SD energy scale". GAP2007-126. 2007 (cit. on p. 94).

[202] A Castellina and L Collica. "Estimation of the Electromagnetic and Muon shower components from the Surface Detector traces." GAP2012-133. 2012 (cit. on p. 94).

[203] F. Pedregosa, G. Varoquaux, A. Gramfort, V. Michel, B. Thirion, O. Grisel, M. Blondel, P. Prettenhofer, R. Weiss, V. Dubourg, J. Vanderplas, A. Passos, D. Cournapeau, M. Brucher, M. Perrot, and E. Duchesnay. Scikit-learn: Machine Learning in Python. In: *Journal of Machine Learning Research* 12 (2011), pp. 2825–2830 (cit. on p. 103).

[204] Tanguy Pierog and Klaus Werner. Muon production in extended air shower simulations. In: *Physical review letters* 101.17 (2008), p. 171101 (cit. on pp. 110, 122).

[205] Alexander Aab, P Abreu, M Aglietta, Imen Al Samarai, IFM Albuquerque, I Allekotte, A Almela, J Alvarez Castillo, J Alvarez-Muñiz, GA Anastasi, et al. Combined fit of spectrum and composition data as measured by the Pierre Auger Observatory. In: *Journal of Cosmology and Astroparticle Physics* 2017.04 (2017), p. 038 (cit. on p. 113).

[206] Lu Lu. "A search for photons of energy above 6×10^{18} eV using data from the Water-Cherenkov detectors of the Pierre Auger Observatory". PhD thesis. University of Leeds, 2014 (cit. on p. 118).

Chapter 9

Traductions françaises

Conformément aux règles mises en place par l'école doctorale de physique de Grenoble, les traductions de l'introduction et de la conclusion de ce manuscrit sont incluses ici en langue française.

9.1 Introduction

La Terre est continuellement bombardée par des atomes ionisés appelés rayons cosmiques (CR). Au dessus de 10^{18} eV (soit 0.16 joules), les rayons cosmiques d'ultra-haute énergie (RCUHE) sont des messagers des évènements astrophysiques extrêmes. En effet dans l'univers, les conditions nécessaires à l'accélération des RCUHE sont difficiles à atteindre et observer un flux significatif de RCUHE en provenance d'un objet astrophysique permettrait de contraindre les propriétés de ce dernier. Durant ces dernières années, le potentiel de l'astronomie via l'étude des RCUHE a été mis en évidence à travers les campagnes multi-messagers entreprises à la suite des détections d'ondes gravitationnelles émises dans des événements de fusions d'objets astrophysiques. Dans de tels événements, il est crucial d'observer tous les types d'émissions, non seulement les sondes "classiques" de l'astronomie mais également les rayons cosmiques, les neutrinos, les ondes gravitationnelles et les photons d'ultra-haute énergie (photons UHE). L'astronomie multi-messagers présente de très belles perspectives dans le futur, par les observations combinées de multiples expériences.

L'observation de particules cosmiques neutres (neutrinos et photons) d'ultra-haute énergie ouvre une nouvelle fenêtre sur l'astronomie aux plus hautes énergies. En effet, ces messagers neutres sont particulièrement intéressants puisqu'ils pointent directement vers leurs lieux d'émissions, n'étant pas déviés par les champs magnétiques galactiques et extragalactiques. Les particules neutres d'ultra-haute énergie peuvent être produites lorsque des RCUHEs interagissent avec le milieu interstellaire et intergalactique, tel le fond diffus cosmologique (CMB) ou autres champs de photons.

L'estimation du flux de photons UHE peut permettre de poser des contraintes fortes sur certains modèles astrophysiques. Dans le cas de sources de photons de haute énergie, la recherche de photons UHE peut aider à mettre des limites sur l'énergie d'accélération maximale disponible au sein de la source ou pour contraindre la forme de leur spectre d'émission; le flux de photons UHE est également une observable primordiale pour comprendre l'origine de la suppression du flux observée aux plus hautes énergies, au delà de 5×10^{19} eV. En effet, les deux principaux scénarios expliquent la suppression soit par une limitation du potentiel d'accélération des sources aux plus hautes énergies soit par le scénario GZK. Dans ce deuxième cas, la suppression du flux d'RCUHEs est expliquée comme la conséquence de la photo-production résonnante de pion par l'interaction d'un rayon cosmique avec un photon du CMB. En effet, au dessus de $\sim 6.9 \times 10^{19}$ eV, les protons peuvent interagir avec les photons du CMB de façon résonnante, augmentant ainsi considérablement la probabilité de la photo-production de pion. Il est intéressant de noter qu'un tel processus produit des pions neutres, qui se désintègrent en photons. Ainsi, une augmentation du flux de photons en dessous de l'énergie limite requise pour le processus de photo-production résonnant serait un indicateur du processus GZK.

Jusqu'ici, aucun photon n'a été détecté au dessus de quelques TeV en énergie. Toutefois, des limites supérieures sur le flux de photons UHE ont été mesurées par des observatoires de rayons cosmiques. Ces limites ont permis de rejeter certains modèles *top-down* proposés pour expliquer l'origine des RCUHEs. Actuellement, les limites supérieures les plus fortes sur le flux de de photons UHE ont été mesurées par l'Observatoire Pierre Auger (PAO).

Le PAO est actuellement l'expérience phare de la détection d'RCUHEs. Il est déployé au sol et il mesure les gerbes atmosphériques (EAS) induites par les CRs lorsqu'ils interagissent avec l'atmosphère, afin de mesurer les caractéristiques des rayons cosmiques. Le PAO a enregistré le plus grand ensemble de gerbes atmosphériques et a obtenu de nombreux résultats sur les RCUHEs, notamment sur leur spectre en énergie, leur composition en masse et leur direction d'arrivée. Des analyses sont conçues dans la collaboration Pierre Auger spécifiquement pour

identifier des photons UHE dans le flux des RCUHEs collectés par l'Observatoire. Certaines caractéristiques des gerbes atmosphériques sont utilisées pour différencier des cascades issues de primaires photons, principalement électromagnétiques, de celles induites par des noyaux atomiques, ayant une composante hadronique. Néanmoins, les recherches de photons UHE peuvent être améliorées en utilisant de nouvelles observables dans des analyses multivariées dédiées. Ceci est l'objectif principal de mon travail de thèse. J'ai conçu une nouvelle analyse multivarié (MVA) pour identifier des photons UHE dans le flux de rayons cosmiques détectés par l'Observatoire. Cette analyse a pour objectif de construire des observables discriminantes à partir des signaux des détecteurs cherenkov à eau (WCD) afin d'estimer la composante muonique des EAS. Les observables sont ensuite combinées dans une MVA afin de séparer les photons des hadrons dans les données de l'Observatoire Pierre Auger.

Malgré les progrès effectués dans le domaine de la physique des RCUHEs grâce aux résultats obtenus par la collaboration Pierre Auger, des questions restent en suspens. Aussi, un projet d'amélioration de la détection appelé AugerPrime est en cours, afin d'améliorer les capacités de l'Observatoire. L'élément principal de cette amélioration est l'ajout de détecteurs plastiques à scintillation (SSD) au dessus de chaque détecteur cherenkov composant le détecteur de surface (SD). Durant ma thèse, j'ai contribué à l'intégration d'une partie des SSDs. L'assemblage des détecteurs à scintillation était partagé entre 6 instituts membres de la collaboration Pierre Auger. Le LPSC a eu la responsabilité de 90 SSDs. Une procédure d'assemblage a été conçue pour garantir une construction efficace durant toute la durée de la production. Des tests ont été effectués afin de garantir le bon fonctionnement de chaque détecteur et de caractériser la réponse de chaque SSD au passage de muons atmosphériques.

Ce manuscrit est divisé en 6 chapitres. Le premier est dédié aux rayons cosmiques et plus particulièrement à ceux d'ultra-hautes énergies et parmi eux les photons. Ensuite, le second chapitre est consacré aux gerbes atmosphériques dont je décris les principales caractéristiques avant de souligner les différences entre cascades hadroniques et électromagnétiques, puisque ces différences sont utilisées pour séparer les deux types de particules primaires dans les analyses de recherche de photons. Dans le troisième chapitre, je présente l'Observatoire Pierre Auger, en insistant plus particulièrement sur le détecteur de surface, étant donné que l'analyse que j'ai conçue utilise les informations provenant de ce détecteur. Après avoir exposé les résultats majeurs de l'Observatoire Pierre Auger, je présente le projet AugerPrime dans le quatrième chapitre. L'accent est mis sur les Détecteurs de Surface à Scintillation, puisque le LPSC à assemblé certains de ces détecteurs avant de les envoyer à l'Observatoire. Je décris également dans le détail les procédures de test et de caractérisations qui ont été utilisées au LPSC et auxquelles j'ai contribué. Dans le cinquième chapitre, les recherches de photons UHE menées dans différents observatoires de rayons cosmiques sont présentées, ainsi que les principales observables utilisées pour différencier les EAS issues de photons de celles issues de primaires hadroniques. Ce chapitre est conclu par une description des observables discriminantes choisies pour mon analyse et une présentation des outils et méthodes utilisés pour combiner ces observables au sein d'une analyse multi-variée. Enfin, les performances de l'analyse multivariée en terme d'identification de photons sont présentées dans le dernier chapitre.

9.2   Conclusion

Ce travail de thèse se concentre sur la recherche de photons d'ultra-haute énergie dans le contexte expérimental fourni par l'Observatoire Pierre Auger (PAO), qui est actuellement l'instrument le plus performant dans l'étude des rayons cosmiques d'ultra-haute énergie (RCUHE). Étant conçu comme un détecteur hybride, il est composé d'un détecteur à fluorescence (FD) et d'un détecteur de surface (SD). Ce dernier est un réseau de 1660 détecteurs cherenkov à eau (WCD) déployé sur $3000\,km^2$ qui échantillonne au sol les particules secondaires des gerbes atmosphériques (EAS) produites lorsqu'un RCUHE interagit avec l'atmosphère.

La majeure partie de mon travail de thèse a été de concevoir une nouvelle analyse multivarié (MVA) pour identifier les photons d'ultra-haute énergie dans le flux de RCUHEs au delà de 3×10^{18} eV détectés avec le SD du PAO. Cette MVA vise à améliorer le pouvoir de discrimination entre les EAS produites par des hadrons de celles générées par des photons UHE, à partir de l'estimation de la fraction du signal due à la composante muonique des EAS. A cette fin, des variables corrélées à la fraction muonique du signal enregistré par le SD sont calculées au niveau des détecteurs individuels, avant d'utiliser des algorithmes de régression par machine-learning (ML) afin de construire une observable discriminante pour chaque EAS. Habituellement, des fonctions analytiques sont utilisées pour construire des benchmarks reproduisant la valeur des variables au niveau des détecteurs individuels en fonction de certaines caractéristiques de ce détecteur et de la gerbe. Pour cette étude, des algorithmes de ML ont été utilisés plutôt que des fonctions analytiques pour trois raisons principales : premièrement, le nombre de critères de sélection à appliquer aux données est réduit; deuxièmement, cette méthode permet une prédiction continue sur tout l'espace des paramètres des détecteurs au lieu de valeurs discrètes pour chaque benchmark; troisièmement, la méthode ML et la façon dont elle a été implémentée est plus flexible (toute caractéristique spécifique aux photons peut être utilisée sans la nécessité de construire une fonction analytique dédiée). L'analyse a été conçue et ses performances ont été évaluée avec des cascades simulées.

La recherche de photon UHE à été effectuée sur 2% de toutes les données collectées par le SD entre le 01/01/2004 et le 31/12/2017, cela correspond à l'échantillon de données utilisé par la collaboration pour tester les nouvelles analyses. À une efficacité de sélection (S.E.) de 50%, le pouvoir de réjection du bruit de fond atteint par la MVA est de 99.87%. Sur 2441 évènements, 1 candidat photon a été identifié, mais a pu être exclu grâce à une observable complémentaire. Les limites supérieures à 95% de niveau de confiance (C.L.) sur le flux de photons au delà de $E > E_0$ avec $\log(E_0)$= (18.5, 19.0, 19.5) ont été obtenue en supposant un bruit de fond nul (hypothèse la plus conservative) et sont respectivement: $\phi^{95\%}(E > E_0) = 1.17 \times 10^{-2}, 7.07 \times 10^{-3}, 7.07 \times 10^{-3}\,km^{-2}.sr^{-1}.y^{-1}$.

Les performances de cette analyse MVA peuvent être améliorées. Premièrement en augmentant l'exposition et également en augmentant le nombre de simulations de cascades utilisées dans l'entraînement des modèles. Deuxièmement, en ajoutant de nouvelles observables. En effet j'ai montré, en ajoutant l'observable RiseTime à la MVA, que toute caractéristique discriminante au niveau des détecteurs individuels du SD peut être utilisée pour créer une nouvelle observable discriminante en utilisant l'architecture développée pour cette analyse.

La capacité du PAO à chercher des photons UHE va également être améliorée par l'amélioration de l'observatoire en cours, le projet AugerPrime. Dans ce manuscrit, j'ai présenté mes contributions à ce projet, sous la forme de la mise en place d'une procédure d'assemblage ainsi que d'une procédure de test pour les détecteurs de surface à scintillation (SSD) au LPSC. Le principal critère de qualité sélectionné pour comparer les détecteurs est

le *single photoelectron* (SPE) par *vertical minimum ionizing particle* (VMIP). Cette valeur représente la quantité de lumière produite par un SSD lorsqu'il est traversé verticalement par un muon atmosphérique. Les valeurs de SPE par VMIP peuvent être comparées entre instituts car elles sont peu dépendantes du système de test utilisé. Les résultats des tests sont partagés par les instituts ayant pris part à la construction des SSDs avec toute la collaboration et permettront un meilleur suivi et un meilleur contrôle des détecteurs lorsqu'ils seront installés à l'Observatoire. Tout les 90 SSDs construits au LPSC ont été validés par les tests de qualité mis en place par la collaboration. Ils ont été expédiés à l'Observatoire Pierre Auger et sont en train d'être déployés.

Les informations fournis par les SSDs vont améliorer l'estimation de la composante muonique des EAS faite par le SD. En effet, cela ajoutera une nouvelle mesure indépendante des particules secondaires des EAS au sol à celle effectuée avec les WCDs du SD. En particulier, de nouvelles variables discriminantes pourraient être ajoutées à la MVA présentée dans ce manuscrit. Globalement, l'amélioration AugerPrime va augmenter la capacité du détecteur à effectuer l'identification de la masse du rayon cosmique primaire, gerbe par gerbe. Dans les années à venir, l'exploitation de cette amélioration va augmenter l'exposition disponible pour les recherches de photons UHE (et les analyses RCUHEs en général).

L'amélioration des analyses de recherche de photons UHE, combinée à la plus grande exposition va mener à un de ces deux scénarios : soit un photon UHE est détecté, soit les limites supérieures sur le flux de photons aux UHE seront abaissées. La détection de photon UHE ouvrirait une nouvelle fenêtre de l'astronomie. L'observation de ces particules neutres nous permettrait de localiser les sources d'RCUHEs. Cela mettrait également en lumière les limites du pouvoir d'accélération des sources. Quelque soit le scénario, une analyse de recherche de photons UHE plus performante nous permettra d'appliquer de plus fortes contraintes sur des modèles astrophysiques, tels que: les modèles de propagation d'RCUHEs, les modèles de sources d'RCUHEs et de photons aux TeV, ainsi que les modèles de matière noire super massive. Enfin, cela pourra encore augmenter le rôle des observatoires d'RCUHEs dans les campagnes multi-messagers.

Remerciements

Comment conclure ces 3 années de thèse? Ces remerciements que je m'apprête à rédiger ont la double tâche de mettre un point final à l'écriture de ce manuscrit et de tenter d'exprimer ma gratitude aux personnes qui m'ont accompagné, aidé, soutenu durant ces trois années et durant ma formation scientifique.

Tout d'abord j'aimerais remercier les membres de mon jury, Laurent Derome, Frédéric Piron, Mariangela Settimo, qui ont accepté de prendre de leur temps pour lire mon manuscrit et assister à ma soutenance, tous en présentiel (un exploit en octobre 2020 !). Un remerciement particulier à Isabelle Lhenry-Yvon et Cristina Carloganu pour les discussions sur mon manuscrit en amont de la soutenance.

Here I want to adress a huge thank you to all of the Auger collaboration, it has been a pleasure to work in such an open-minded environment and the moments shared in Malargüe will remain unforgetable memories for me.

Bien évidemment, je dois m'étendre quelque peu sur le groupe Auger du LPSC. Merci Murielle, Jean-Pierre et Patrick, vous côtoyer fût un réel plaisir, tant sur le plan professionnel que sur le plan humain. Merci François pour avoir partagé avec moi ton expérience scientifique, humaine.

J'ai eu le plaisir de réaliser ma thèse sous la direction de Corinne Berat. Merci Corinne de m'avoir offert l'opportunité d'effectuer cette thèse, pour la bienvieillance de ton encadrement et pour t'être toujours rendue disponible, travailler avec toi durant ces trois années fût un réel plaisir.

Lors de ma formation scientifique j'ai eu la chance de rencontrer des scientifiques qui m'ont fait aimer la physique, sans qui je n'en serais pas ici aujourd'hui et que je tiens à remercier ici : Julien Masbou, Loic Rolland et Cécile Renault.

Les Michels ! Manu, Hugo, David, merci et continuons à boire des coups dès que l'occasion se présente.

Merci aux doctorants du BIDUL pour les pauses cafés du matin, les pauses cafés du midi et les pauses cafés de l'après-midi ! Cette camaraderie entre les doctorants est une vraie force du LPSC.

Merci aussi aux doctorants partis trop tôt (en post-doc !), Ju (le bogoss), Kiki (même si il s'accroche au labo), Toto (Boboule), Lolo (Raton). J'aurais aimé que l'on débute nos doctorats en même temps ! Flora, Marine, Seb, merci d'avoir été mes camarades durant ces trois années. Ce fût un réel plaisir de partager bouffes, discussions (à haut volume sonore), cafés (thés pour Marine) et bien sûr bières avec vous !

Merci Ana, petit panda roux, de m'avoir supporté durant cette dernière année.

Merci à mes parents de m'avoir toujours aimé d'un amour inconditionnel, vous m'avez donné le goût d'apprendre, de comprendre. Jérôme merci d'être un moteur pour moi.

Pour finir, si jamais vous tombez sur ces remerciements la Communauté... Merci de votre amitié, merci d'être des frères pour moi, ça doit faire 10 ans au moment où j'écris ces lignes que j'ai la chance de vous avoir comme amis, plus longtemps encore pour certains d'entre vous et cette amitié s'enrichit d'année en année. Merci.

Title : Search for cosmogenic photons in the data flux of the Pierre Auger Observatory

Abstract: This thesis manuscript present two contributions to the study of ultra-high energy cosmic rays with the Pierre Auger Observatory: on one hand the assembly and characterization of scintillation detectors and on the other, the development of a new multivariate analysis for the search of ultra-high energy photons.

The scintillation detectors are part of the AugerPrime project, to enhance the performances of the Pierre Auger Observatory. Scintillation detectors are being installed on top of the water cherenkov detectors making up the surface detector of the observatory. This new system will perform a complementary measurement of the secondary particles inside the atmospheric showers, thereby increasing the capacity of the observatory to identify the nature of the cosmic rays. The LPSC has had the responsibility of assembling 90 detectors. A detailed procedure was written to ensure a continuity throughout the assembly process. Once assembled, the integrity of the detectors is controlled and the performances of each detector are evaluated. This is done by measuring the response of the detectors to a vertical through-going muon in units of photoelectrons. The measurements performed on the detectors assembled at the LPSC are presented: they are compatible with the quality criteria required by the collaboration. They are also compared to the measurements performed in the other assembly sites.

The main contribution presented in this manuscript concerns the search for ultra-high energy photons. Indeed, identifying photons in the flux of ultra-high energy cosmic rays would allow a direct localization of cosmic rays sources in the nearby universe. The upper-limits on the flux of such photons, set by the previous analyses have already constrained ultra-high energy cosmic rays production scenarios, such as the so-called "top-down" scenarios and could constrain the astrophysical scenarios of production and propagation (GZK effect). The analysis presented in this manuscript uses the signals of the water cherenkov detectors making up the surface detector of the Pierre Auger Observatory in order to construct discriminant observables. These oervables take advantage of the differences between hadronic and electromagnetic showers in terms of their muonic component ratio, and in terms of the thickness of their shower fronts. The observables are reconstructed using machine-learning models, trained on simulated events in a biased way in order to differentiate the photons from other types of cosmic rays. Finally, the discriminant observables are combined in a photon search analysis, applied to a sample of data from the Pierre Auger Observatory.

The performances of such an analysis are discussed, as well as the perspectives in the framework of AugerPrime.

Keywords : Cosmic rays, Pierre Auger Observatory, AugerPrime, UHE photons, Scintillation detectors, Multivariate analysis, Machine learning

Titre : Recherche de photons cosmogéniques dans le flux de données de l'observatoire Pierre Auger

Résumé : Ce manuscrit de thèse présente deux contributions à l'étude des rayons cosmiques d'ultra-haute énergie avec l'observatoire Pierre Auger : d'une part l'assemblage et la caractérisation de détecteurs à scintillation et d'autre part le développement d'une nouvelle analyse multivariée pour la recherche de photons d'ultra haute énergie.

Les détecteurs à scintillation font partie du projet AugerPrime pour améliorer les performances de l'Observatoire Pierre Auger. Ils doivent compléter le réseau de surface de l'Observatoire actuellement composé de détecteurs cherenkov à eau. Ce nouveau système permettra, en effectuant une mesure complémentaire des particules secondaires des gerbes atmosphériques, d'améliorer la capacité de l'observatoire à identifier la nature des rayons cosmiques. Le LPSC a eu la charge d'assembler 90 détecteurs. Afin d'uniformiser le processus d'assemblage tout au long de la production, une procédure a été rédigée. Une fois assemblés, l'intégrité du fonctionnement des détecteurs est contrôlée et les performances de chaque détecteur sont évaluées. Pour celà, la réponse des détecteurs à un muon vertical traversant est mesurée en nombre de photoélectrons. Les mesures effectuées sur les détecteurs assemblés au LPSC sont présentées et sont conformes aux critères de qualité requis par la collaboration. Elles sont également comparées aux mesures effectuées dans les autres sites d'assemblage.

La principale contribution présentée dans ce manuscrit concerne la recherche de photons d'ultra-haute énergie. En effet, l'identification de photons dans le flux des rayons cosmiques d'ultra-haute énergie permettrait de localiser directement des sources de rayons cosmiques dans l'univers proche. Les limites supérieures sur le flux de tels photons obtenues par les précédentes analyses permettent d'ores et déjà de contraindre les modèles de production des rayons cosmiques d'ultra haute énergie, tels que les scénarios dits "top-down" et constituent un moyen de contraindre les scenarios astrophysiques de production et de propagation (effet GZK). L'analyse présentée dans ce manuscrit utilise les signaux des détecteurs cherenkov à eau composant le détecteur de surface de l'observatoire Pierre Auger afin de construire des observables discriminantes. Celles-ci exploitent les différences entre gerbes hadroniques et gerbes électromagnétiques en terme de proportion de la composante muonique, ainsi qu'en terme d'épaisseur du front de développement des gerbes. Les observables sont reconstruites au moyen de modèles de machine-learning, entrainés sur des évènements simulés de manière biaisée afin de différencier les photons des autres types de rayons cosmiques. Enfin, les observables discriminantes sont combinées dans une analyse de recherche de photons, appliquée à un échantillon des données de l'observatoire Pierre Auger.

Les performances d'une telle analyse sont discutées, ainsi que les perspectives dans le cadre d'AugerPrime

Mots clés : Rayons cosmiques, Observatoire Pierre Auger, AugerPrime, Photons UHE, Détecteurs à scintillation, Analyse multivariée, Machine learning