Pontificia Universidad Católica del Perú

TESIS PARA OPTAR EL GRADO DE MAGISTER EN FISICA

Javier Rengifo Gonzáles

Disentangling atmospheric cascades started by gamma rays from cosmic rays with CORSIKA

Escuela de Posgrado

Asesor de Tesis: Dr. José Luis Bazo Alba

Jurado: Dr. Alberto Martín Gago Medina
Dr. Joel Jones Pérez

Perú 2017

I would like to thank Cienciactiva-CONCYTEC. I also thank my advisor, my professors, my family, my friends of the PUCP and my girlfriend for helping me in this process.

ABSTRACT

In this work we search for a method to differentiate between particle showers produced by cosmic rays and by gamma rays at TeV energies, using CORSIKA simulations. This method tries to solve the dominant hadron flux background problem when looking for gamma-ray signals measured by different experiments. The results of this work can be applied to the study of Gamma-Ray Bursts (GRBs). GRBs emit very energetic photons, which after interacting in the Earth's atmosphere, produce a large detectable electromagnetic cascade of secondary particles.

The procedure will be to simulate events produced by photons, the signal, and protons, the most abundant cosmic-ray background. We extract several parameters from fitting particle air-shower longitudinal profiles, characterizing the simulated showers. Some of the most important fit parameters are the shower maximum (X_{max}), the width of the shower $FWHM$, the asymmetry parameter, the maximum number of particles N_{max} and the shower start X_{start}. There are different experiments using water Cherenkov tanks and fluorescence detectors which can measure these shower parameters.

We tested two methods. The first relies on simple cuts, while the second is based on a multivariate analysis using the TMVA package, which improves individual cuts. The first method was applied to single simulated energies of 10^2, 10^3, 10^4 and 10^5 GeV to find adequate cuts. We found that X_{max}, $FWHM$, X_{start} and N_{max} depend on the energy. Later we applied these energy-dependent cuts and other fixed cuts to a realistic sample, which consists of 10^4 signal events (photons) and 10^6 background events (protons) covering an energy range from 10^2 to 10^5 GeV with different spectra. Moreover, we introduced an energy smearing to simulate a detector energy reconstruction efficiency. The obtained result leaves 54% signal events and 12% background events. Applying the multivariate analysis TMVA, we found that the Boosted Decision Trees (BDT) method was the best for distinguishing signal from background. The result for a similar signal efficiency was 0.7% of background events.

Finally using tighter cuts on the BDT to improve the significance results in 1 photon event for every 1000 protons. Given the initial flux proportion, it means a 10^3 background rejection capability. Thus the feasibility of gamma/hadron separation requires further improvement.

Keywords: Cosmic-Rays, Gamma-Rays, Longitudinal Profile, X_{max}, CORSIKA

RESUMEN

En este trabajo buscamos un método para diferenciar entre lluvias de partículas producidas por rayos cósmicos y por rayos gamma a energías de TeV, utilizando simulaciones CORSIKA. Este método intenta resolver el problema que existe en la búsqueda de señales de rayos gamma medidos por diversos experimentos frente a un fondo de flujo dominante de hadrones. Los resultados de este trabajo pueden aplicarse al estudio de Explosiones de Rayos Gamma (GRBs). Los GRBs emiten fotones muy energéticos, que al interactuar con la atmósfera terrestre, producen una gran cascada electromagnética de partículas secundarias, las cuales son detectables.

El procedimiento será simular eventos producidos por fotones, la señal, y protones, el fondo, que son las partículas más abundantes de los rayos cósmicos. Extraemos varios parámetros de los perfiles longitudinales de las lluvias de partículas, caracterizando las lluvias simuladas. Algunos de los parámetros de ajuste más importantes son el máximo de lluvia (X_{max}), el ancho de la lluvia $FWHM$, el parámetro de asimetría, el número máximo de partículas N_{max} y el comienzo de lluvia X_{Start}. Existen diferentes experimentos utilizando tanques Cherenkov de agua y detectores de fluorescencia que pueden medir estos parámetros de las lluvias.

Hemos probado dos métodos. El primero se basa en cortes simples, mientras que el segundo se basa en un análisis multivariado utilizando el paquete TMVA, que mejora los cortes individuales. El primer método se aplicó a las energías simuladas separadas de 10^2, 10^3, 10^4 y 10^5 GeV para encontrar cortes adecuados. Encontramos que X_{max}, $FWHM$, X_{start} y N_{max} dependen de la energía. Posteriormente aplicamos estos cortes dependientes de la energía y otros cortes fijos a una muestra realista, que consiste en 10^4 eventos de señales (fotones) y 10^6 eventos de fondo (protones) que cubren un rango de energía de 10^2 a 10^5 GeV con diferentes espectros. Además, se introdujo un error en la energía simulada para simular la eficiencia de reconstrucción de energía de un detector. El resultado obtenido deja 54% eventos de señal y 12% eventos de fondo. Aplicando el análisis multivariado TMVA, encontramos que el método Boosted Decision Trees (BDT) era el mejor para distinguir la señal del fondo. El resultado para una eficiencia de señal similar fue 0.7% de eventos de fondo.

Por último, utilizando cortes más estrictos en la BDT para mejorar la significancia, el resultado fue 1 evento de fotón por cada 1000 eventos de protón. Dada la proporción de flujo inicial, significa una capacidad de rechazo de fondo de 10^3. Por lo tanto, la viabilidad de la separación gamma/hadrón requiere una mejora adicional.

Palabras Claves: Rayos Cósmicos, Rayos Gamma, Perfil longitudinal, X_{max}, CORSIKA

Contents

Chapter 1

Introduction

Cosmic rays, accelerated in astrophysical sources, are charged particles and ionized nuclei (e.g. electrons, protons and helium, as well as carbon, oxygen, iron, and other nuclei synthesized in stars) that arrive to Earth from not yet identified sources, given their deviation due to galactic and extragalactic magnetic fields(Alania et al., 2009). Cosmic rays arriving at the Earth's atmosphere are called primary cosmic rays. Most of them are mainly protons (Olive et al., 2014). When the primary cosmic rays reach the atmosphere, they interact and produce a shower of other particles called secondaries.

Cosmic rays are a very important topic because of their relation with the universe. These particles carry information from different astrophysical sources, their history and evolution. On the other hand, other messengers from the cosmos are gamma rays. One source of gamma rays is a very energetic cosmic event, called Gamma Ray Burst (GRB), that can be detected at Earth. These photons also carry important information about the universe. Moreover, cosmic rays (Dorman, 2012) and gamma rays (Thomas et al., 2005) have a relation with the space and Earth weather.

The problem that we try to solve in this work arises because there is a dominant background of showers induced by hadrons, mainly protons, in the studies that look for signals of gamma rays. To solve this, we apply two methods. While the first method relies on simple cuts, the second method uses a multivariate analysis based on the TMVA package.

To better understand this problem this work will be developed in nine chapters. After a brief introduction in the present chapter, the second chapter elaborates on the cosmic ray spectrum and its properties. It also describes the types of extensive air showers. Showers initiated by different particles contain distinct features that can be used to distinguish electromagnetic showers originated by photons from hadronic showers produced by protons.

The third chapter contains the definition and some properties of gamma rays and GRBs. It includes the flux of these particles, since it is very important for comparing the expected number of photons and protons that arrive to Earth.

Chapter four gives succinct information about three detectors that motivated this work. The detectors are Pierre Auger, which has surface detectors and fluorescence detectors. These two kinds of detectors could measure some parameters that we will study in this work. We will also mention HAWC and LAGO, which work in gamma rays research with Cherenkov tanks, similar detectors to the

surface detectors of Pierre Auger.

In chapter five the simulation package CORSIKA used in this work will be described. In addition we show how to simulate showers of cosmic rays and the chosen input values for this.

The sixth chapter introduces the parameters that we later use to differentiate the showers. These parameters are features of the showers that can be extracted by fitting the longitudinal profile (number of particles vs. atmospheric depth).

Chapter seven contains the first method used to distinguish photons from protons. This method is simple, but useful to understand more complex methods and serves as benchmark.

The eight chapter describes the multivariate analysis, the second method used to solve the main question of this work. There we will show that the disentangling capabilities of this method have a higher efficiency.

In the last chapter we present the conclusions achieved in this work and some topics that could be develop in a future work for improving the results.

Chapter 2

Cosmic Rays and Extensive Air Showers

In this chapter we will describe the cosmic ray spectrum and the types of extensive air shower that we will study. The cosmic ray spectrum is important because it describes many properties about the cosmic rays. In the spectrum we can see the flux in the atmosphere and the energy range. Also it is necessary for us to understand the types of showers because we want to distinguish between photons and protons that produce showers with different properties that we will later study as parameters.

2.1 Cosmic rays

Primary cosmic rays cover a wide energy range. Their energy spectrum is shown in Fig.2.1 in the range from 10^{13} to 10^{20} eV.

The intensity of primary nucleons in the sub-range up to $\approx$ 100TeV is given approximately by:

$$I_N(E) \approx 1.8 \times 10^4 (E/GeV)^{-\alpha} \frac{\text{nucleons}}{\text{m}^2 \text{ s sr GeV}} \tag{2.1}$$

where E is the energy-per-nucleon (including rest mass energy) and $\alpha \approx 2.7$ is the differential spectral index of the cosmic ray flux. About 79% of the primary nucleons are free protons and about 21% of the rest are nucleons bound in helium nuclei (Olive et al., 2014).

The spectral index change with the spectrum at differences energies. Figure 2.1 shows the all-particle spectrum, 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 observe. The flux of cosmic rays is influenced at energies lower than $\approx$10 GeV by the solar wind, with the solar cycle having a period of 11 years (Potgieter, 2013).

At higher energies, at 3×10^{15} eV, is the *knee* where the spectral index falls from 2.7 to 3. This behavior can be related to the maximum acceleration in supernova remnants inside our galaxy (Gaisser, 2006). At 4×10^{17} eV another downturn is observed called the *second knee* where the slope of the spectrum goes from 3 to 3.3. At this energy the upper end of CR galactic sources could be reached (Bergman and Belz, 2007).

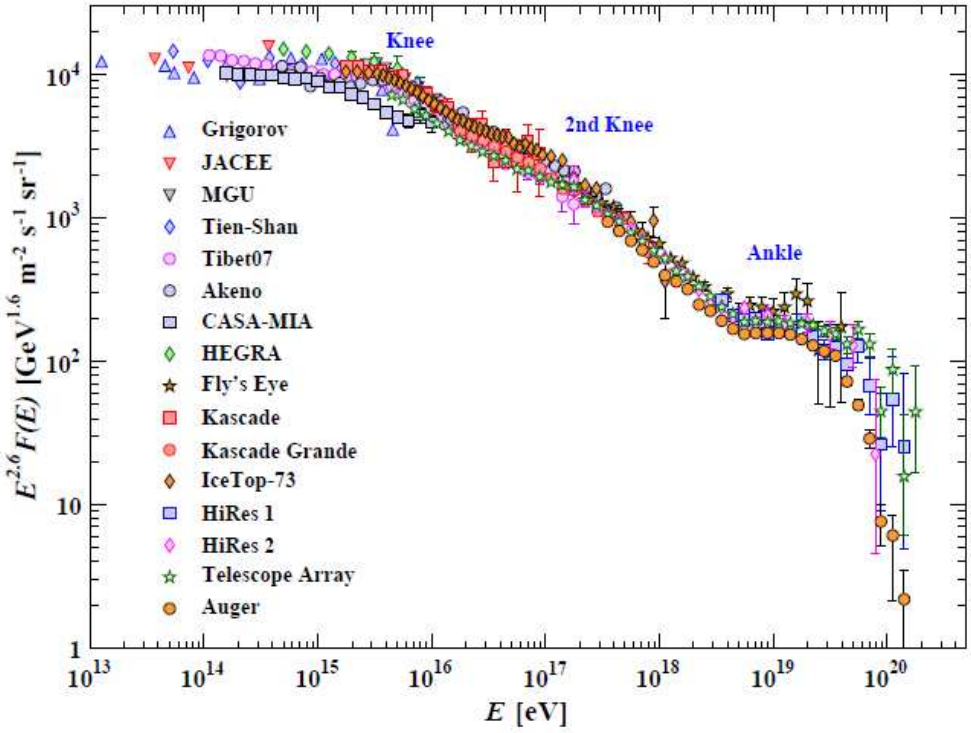

Figure 2.1: All-particle cosmic ray spectrum from experiments (Olive et al., 2014)

Afterwards at approximately $10^{18.5}$ eV is the *ankle* region where the spectral index reaches 4.2. This flattening of the CR energy spectrum (Younk, 2007), can be interpreted as a change from the galactic component to a component of extragalactic origin. For particles with energies in excess of $\approx 10^{19}$ eV the intensities are extremely low (Asorey, 2012). At these energies there is a cut called GZK (Greisen, Zatsepin, and Kuzmin) produced because protons of higher energies would interact with the CMB (i.e. Cosmic Microwave Background, the black body thermal radiation that escaped at the time of recombination).

2.2 Extensive Air Showers

Primary nuclei will interact with the Earth's atmosphere predominantly via processes of strong interactions and will produce Extensive Air Showers (EAS). EAS are showers of secondary cosmic rays produced by transport and interactions of primary cosmic rays that arrive to Earth and collide with atoms and molecules, mainly oxygen and nitrogen. In this context, it is necessary to understand the atmosphere.

The column density (Fig. 2.2) of the whole atmosphere amounts to approximately 1000 g/cm^2. The altitude for flight of scientific balloons is of 35 or 40 km and corresponds to approximately ten g/cm^2. The amount of matter traversed for the interactions of particles in the atmosphere is called radiation length (Beringer et al., 2012)[1] and for photons and electrons in air is $\lambda_r = 36.66\ g/cm^2$ (Grupen,

[1]The radiation length is defined at high energy for electrons as the length over which they have left over $1/e$ of their original energy, by bremsstrahlung, and for photons as 7/9 of the

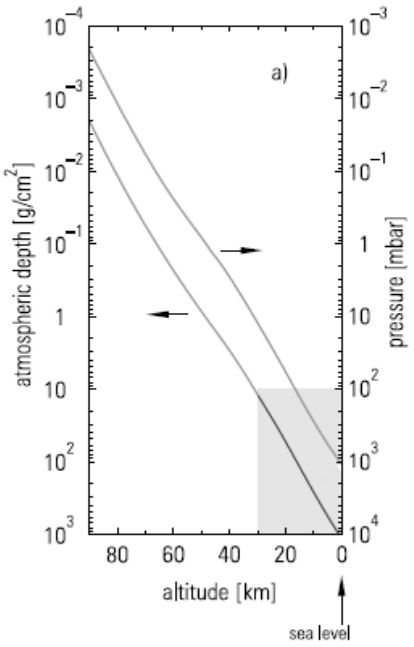

Figure 2.2: Relation between atmospheric depth (column density) and pressure (Grupen, 2005)

2005) and for the whole atmosphere it corresponds to a depth of 27 radiation lengths. The mean free path or interaction length for hadrons in air is $\lambda_I = 120.0$ g/cm^2 (Weidenhaupt, 2014), corresponding to 11 interaction lengths for the whole atmosphere. Already at altitudes of 15 to 20 km primary cosmic rays interact with air atomic nuclei and initiate, depending on energy and particle species, electromagnetic and/or hadronic cascades (Grupen, 2005).

As mentioned before, protons are the most abundant primary cosmic rays. The mostly produced secondary particles are pions. Kaons are only produced with a probability of 10% compared to pions. Neutral pions initiate via their decay ($\pi^0 \rightarrow \gamma + \gamma$) electromagnetic cascades, whose development is characterized by the shorter radiation length ($\lambda_r = \frac{1}{3}\lambda_I$ in air)(Grupen, 2005).

The leptonic decays of pions and kaons produce the penetrating muon and neutrino components ($\pi^+ \rightarrow \mu^+ + \nu_\mu$; $\pi^- \rightarrow \mu^- + \bar{\nu_\mu}$; $K^+ \rightarrow \mu^+ + \nu_\mu$; $K^- \rightarrow \mu^- + \bar{\nu_\mu}$).Muons can also decay $\mu^+ \rightarrow e^+ + \nu_e + \bar{\nu_\mu}$ and $\mu^- \rightarrow e^- + \bar{\nu_e} + \nu_\mu$. Of all charged particles that arrive at sea level 80% are muons (Grupen, 2005). The part of the shower formed by muons is called muons cascade. The showers with secondary mesons and baryons that can also survive down to sea level are called hadronic cascades.

Showers can be classified as electromagnetic, muonic or hadronic. This work is focused in the study of electromagnetic and hadronic showers since we want to differentiate showers initiated by photons from those by protons.

Then we show the Heitler model (Heitler, 1954) of showers and derive basic shower properties. Hadronic showers are described by an extension of the electromagnetic model in the subsequent section (Weidenhaupt, 2014).

mean free path for pair production

2.2.1 Electromagnetic shower

Heitler's model (Heitler, 1954) described the evolution of electromagnetic cascades as a perfect binary tree (See Fig. 2.3, left). This description starts with an electron or a photon with energy E_0. Electrons emit a single photon via Bremsstrahlung and survive with half of their initial energy. Photons interact by pair production, creating an electron/positron of equal energy. In this approach all the processes cross sections are independent of energy and equal.

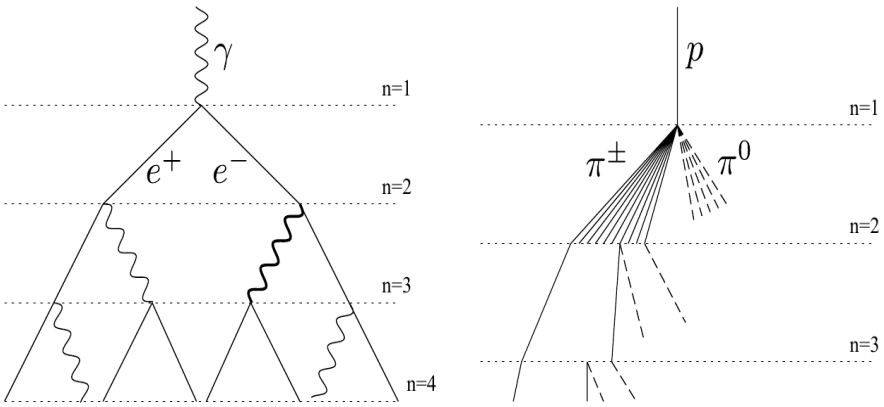

Figure 2.3: Schematic views of an electromagnetic cascade (left) and a hadronic cascade (right). (Weidenhaupt, 2014)

The splitting length d is $d = \lambda_r ln(2)$. After n steps the particle number is $N_n = 2^n$ and their individual energy is $E = E_0/N_n$. The shower continues until the individual energy of the shower particles drops below a critical value E_c where the rate of energy loss by electrons via Bremsstrahlung is equal to ionization. This energy is $E_c^\gamma = 85$ MeV in air (Letessier-Selvon and Stanev, 2011). This model shows that the maximum number of particles is:

$$N_{max} = \frac{E_0}{E_c^\gamma} \tag{2.2}$$

The maximum depth of the shower (measured in g/cm^2) increases logarithmically with energy:

$$X_{max} = X_0 + \lambda_r ln \frac{E_0}{E_c^\gamma} \tag{2.3}$$

where X_0 is the starting point of the shower.

Heitler's model (Heitler, 1954) predicts a ratio of electrons to photons of 2 while simulations and direct cascade measurements in air show a ratio of the order of 1/6th. This is due to the fact that multiple photons are emitted during Bremsstrahlung and that electrons lose energy much faster than photons (Letessier-Selvon and Stanev, 2011).

2.2.2 Hadronic shower

Heitler's model can be adapted to describe hadronic showers (Heitler, 1954). The hadronic shower is modeled as a tree as depicted in Figure 2.3, right.

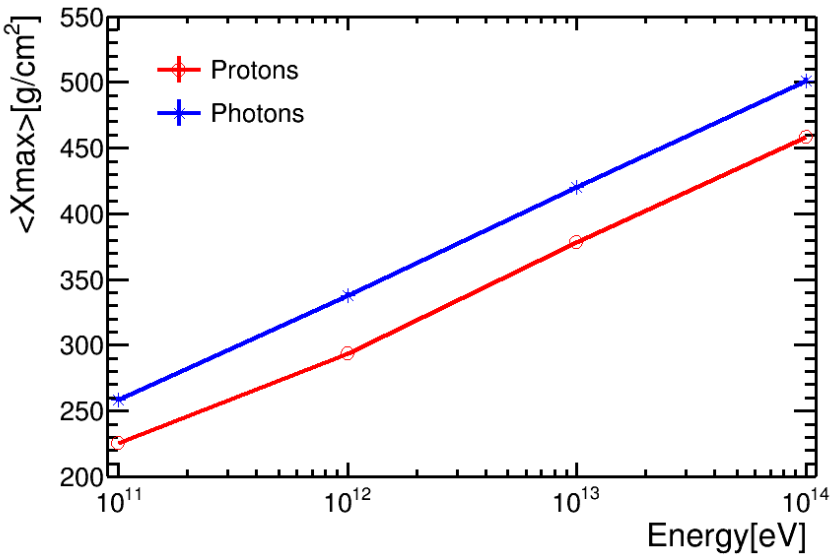

Figure 2.4: X_{max} vs Energy for proton and photon. Results of simulation of 1000 showers using model QJSJET-II04 for low energies from this thesis.

The shower starts with a proton with energy E_0. The splitting length d is $d = \lambda_I ln(2)$, which is assumed to be constant. At every node, a hadronic interaction produces N_{ch} charged pions and $1/2$ N_{ch} neutral ones of equal energy (Weidenhaupt, 2014). As explained in Sec. 2.2, neutral pions decay and extend the electromagnetic cascade. The hadronic cascade decreases with every step as:

$$E^{\pi}_{tot} = \frac{2}{3}^{n} E_0 \tag{2.4}$$

The critical energy for pions is $E_c = 20$ GeV in air, below which they will decay into muons. Therefore the number of muons at this energy is:

$$N_{\mu} = (2N_{\pi})^{n_c} \tag{2.5}$$

where:

$$n_c = \frac{ln(E_0/E^{\pi}_c)}{ln3N_{\pi}} \tag{2.6}$$

is the number of steps for the pions to get E_c (Letessier-Selvon and Stanev, 2011).

The muon number for a given initial energy depends on the mass of the primary cosmic ray. Showers of heavier primaries produce more muons (Weidenhaupt, 2014). The experiments use the muon number to estimate the composition.

The maximum of the air shower (X_{max}) is given by the maximum of the electromagnetic cascade, since in air-showers the electromagnetic component is dominant. The position of shower maximum is more complex in the case of hadronic showers because the larger cross section and the larger multiplicity at each step will reduce the value of X_{max}, while the energy evolution of those quantities will modify the rate of change of X_{max} with energy (Letessier-Selvon and Stanev, 2011).

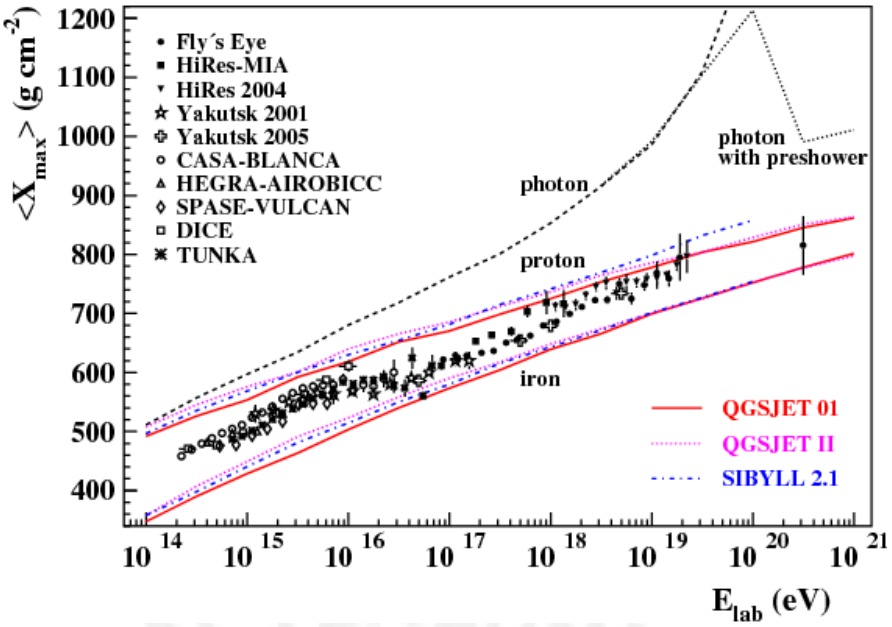

Figure 2.5: X_{max} vs Energy for iron, proton and photon. Results from different hadronic interaction models and experimental data (Matthiae, 2008).

However after doing some simple assumptions, showers induced by nuclei with atomic number A will develop higher in the atmosphere (Letessier-Selvon and Stanev, 2011). The offset with respect to proton showers is:

$$X^A_{max} = X^p_{max} - \lambda_r ln A \tag{2.7}$$

The depth of the shower maximum is used as composition estimator for detectors that are able to measure the shower development in the atmosphere. The X_{max} vs energy plot shown in the Fig. 2.4 from this work and Fig. 2.5 from simulated and experimental data. These plots show how the shower from photons and protons could be distinguished at low and high energies, although we see that it is more difficult at lower energies. We also observe that our results are in agreement with the experimental data at low energies.

Chapter 3

Gamma Rays and GRB's

In this chapter we introduce the concept of gamma rays, their sources and production mechanisms. Then we describe the features of Gamma Ray Bursts or GRBs, their fluxes and spectra. We also show other sources of gamma rays in order to compare them with the flux of cosmic rays. In this way, we can contrast the number of photons and protons arriving to Earth.

3.1 Gamma Rays

Gamma-rays are a form of electromagnetic (EM) radiation similar to visible light. The light has both particle and wave-like properties and according to the Plank-Einstein equation its energy is $E = hf$. Under the wave model, light is characterized by a frequency f and wavelength λ which are related by $v = f\lambda$, where v is the speed of the wave. The speed of light in vacuum is 3 x 10^8 m/s, but when light travels in different mediums, its speed is less than c and depends on the refractive index n of the medium $v_{phase} = \frac{c}{n}$ ($n = 1$ in vacuum and $n > 1$ in medium).

The wavelength of visible light is in the range of 400 to 700 nm and wavelengths of gamma-rays are from 0.001 nm to much smaller lengths, consequently they have the highest energy from 100 keV to over 10 TeV approximately.

Gamma-rays do not penetrate the Earth's atmosphere like visible light does. Many gamma-ray detectors are on board satellites to compensate for this fact. However, satellites only observe a limited energy range of gammas from 1 GeV to 300 GeV (Grupen, 2005). Understanding atmospheric interactions is key to studying higher energy gamma rays on Earth. They travel in a straight trajectory due to their neutral charge and their interaction length depends on energy. Gamma-rays can carry information about our universe from distances up to 10 Mpc. When gamma-rays interact with the Earth's atmosphere, they produce an electromagnetic shower as described in Sec. 2.2.1.

3.1.1 Sources and production mechanisms for gamma rays

Possible sources for cosmic rays and γ rays are supernovae and their remnants, pulsars and neutron stars, active galactic nuclei, and matter-accreting black holes. In these sources γ rays can be produced by different mechanisms like (Grupen, 2005):

Synchrotron radiation

In the presence of magnetic fields the charged particles are deflected and get an accelerated motion. Therefore they radiate electromagnetic waves called synchrotron radiation, which are produced in cosmic magnetic fields and are predominantly emitted by electrons.

Bremsstrahlung

In the presence of a Coulomb field produced by a charge (e.g. atomic nucleus), a charged particle is deflected and emits Bremsstrahlung photons. This mechanism is similar to synchrotron radiation.

Inverse Compton scattering

Inverse Compton scattering happens when energetic photons can transfer part of their energy to free electrons in collisions and they lose a certain amount of energy.

π^0 Decay

Protons accelerated in the sources can produce charged and neutral pions in proton-proton or proton-nucleus interactions.

$$p + nucleus \rightarrow p' + nucleus' + \pi^+ + \pi^- + \pi^0 \tag{3.1}$$

Charged pions decay in muons and neutrinos, while neutral pions decay rapidly in two γ rays.

$$\pi^0 \rightarrow \gamma + \gamma \tag{3.2}$$

Others

Photons from Matter-Antimatter Annihilation: Charged particles can annihilate with their antiparticles similar to pair production. Dominant sources are electron-positron and proton-antiproton:

$$e^+ + e^- \rightarrow \gamma + \gamma \tag{3.3}$$

Photons from Nuclear Transformations: In supernova explosions heavy elements decay in radioisotopes, which emit, mostly γ-ray as a consequence of a beta decay:

$$^{60}Co \rightarrow ^{60} Ni^{**} + e^- + \bar{\nu}_e \rightarrow ^{60} Ni^* + \gamma(1.17\text{MeV}) \rightarrow ^{60} Ni + \gamma(1.33\text{MeV}) \tag{3.4}$$

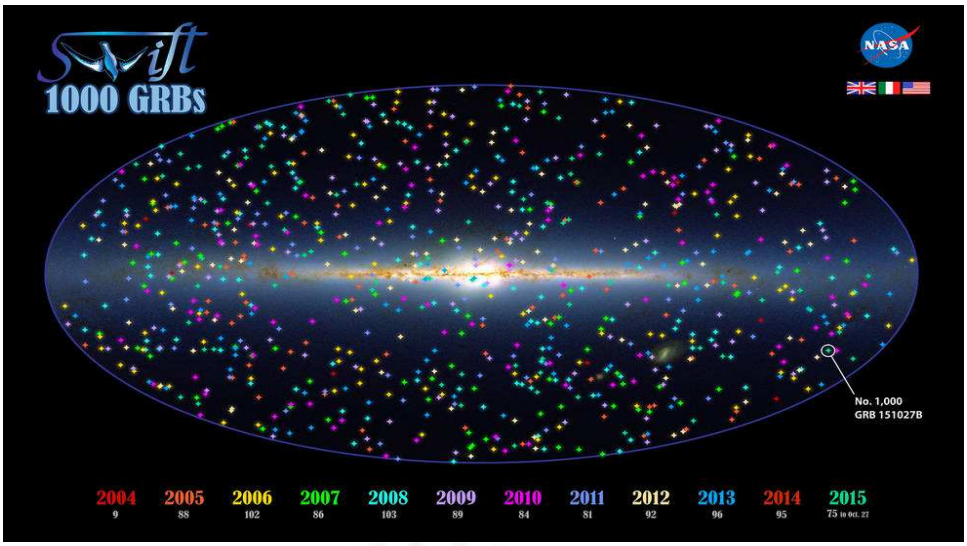

Figure 3.1: Positions of 1000 Swift GRBs in galactic coordinates. Bursts are color coded by year. The total annual number of detected GRBs is quoted below each year. Background: An infrared view from the Two Micron All-Sky Survey. Credits: NASA's Goddard Space Flight Center and 2MASS/ J. Carpenter, T. H. Jarrett, and R. Hurt

3.2 Gamma ray burst

Gamma Ray Bursts or GRBs are very energetic cosmic events that can be detected at Earth. They occur suddenly and unpredictably with a rate of approximately one burst per day (Grupen, 2005). Their duration is very short from fractions of a second up to 100 seconds.

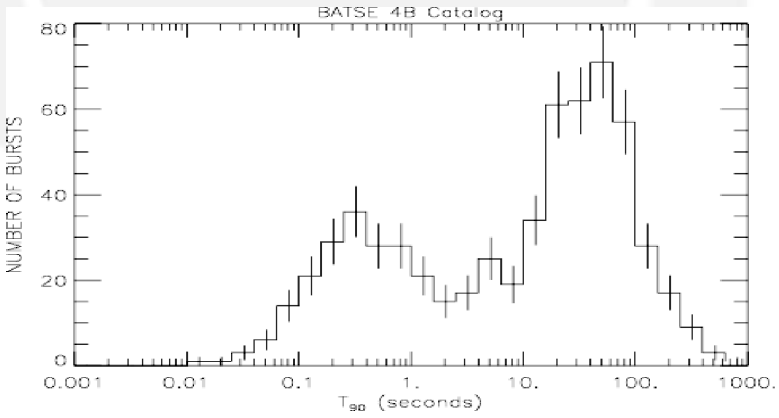

Figure 3.2: The distribution of GRB durations from the bursts observed by the BATSE experiment showing the short and long bursts. Credits: NASA, G.J. Fishman and the Compton Gamma-Ray Observatory (CGRO) Science Team. (Longair, 2011))

Burst and Transient Source Experiment (BATSE) [*gammaray.nsstc.nasa.gov/*] was a detector that in $1990's$ detected and located GRBs and established their main properties (Longair, 2011): GRBs are uniformly distributed over the sky (Fig. 3.1), the distribution of the sources is cosmological and there are bursts

which last less than 2 seconds being termed 'short' and those of greater duration 'long', as is shown in Fig. 3.2.

It is believed that the 'long' bursts can be caused by violent supernova explosions (e.g., hypernova explosions with a collapse into a rotating black hole). This has been confirmed by the *GRB* 030329 and its source, supernova explosion *SN* 2003*dh* (Grupen, 2005). On the other hand, the possible sources for 'short' burst are collisions of neutron stars, collisions of neutron stars with black holes, coalescence of two neutron stars forming a black hole, asteroid impacts on neutron stars, or exploding primordial mini black holes.

3.2.1 GRB Spectrum

There is much information in the literature about this topic, but in Longair (2011) a GRB is described as:

A *relativistic fireball*, which is a relativistic sphere expanding and heating its surrounding gas, moving with a relativistic shock wave. If the sphere is optically thick, a thermal spectrum is observed. However, GRB's spectra observed are non-thermal with power-law spectra of the form $N(\varepsilon) \propto \varepsilon^{\alpha}$, where $\alpha \sim 2-3$ at different energies. Therefore the radiation starts in optically thin regions. The relativistic fireball drives an external shock and there is a contact discontinuity between the material of the fireball and the shocked gas (Longair, 2011), as shown in the Fig. 3.3, where the radiation goes out from black hole and forms two regions: the prompt emission and afterglows at different energies.

In general, according to the results of BATSE, many models were intended for GRB's with energy spectra analysis at low energy band (keV to MeV). Most part of GRBs energy spectra is usually described by a two-component Band function where the first component is proportional to power law with index α while second one is proportional to power law with index β (Arkhangelskaja, 2015). This model is called Band Model and is described with four free parameters: one that describes the spectral break, other for the flux normalization and two more for the behavior at high and low energy. The values typically are $\alpha < 2$ for low energy and $\beta > 2$ for high energy (Nava et al., 2011).

On the other hand, there are some instruments that have measured the properties of GRBs at difference energies and their results vary. The HESS Collaboration (HESS-Collaboration et al., 2014) studied the *GRB* 100621*A*, showing that for low energy the index is ~ 1.69 and for high-energy the index is ~ 2.46. This variations of the spectral index have influence on the fluence at lower energies. For E^2 the flux corresponds to 1×10^{-5} erg cm^{-2} for 10 keV and 10 GeV or 2×10^{-6} erg cm^{-2} for more than 100 MeV, in contrast for $E^{1.5}$ it changes to 2×10^{-7} erg cm^{-2} and 4×10^{-8} erg cm^{-2} (HESS-Collaboration et al., 2014).

Also Merck et al. (2016) showed the properties of five GRB's detected with *EGRET*. Two of them, *GRB* 930131 at 1200 MeV has spectral index 2.0 ± 0.4 and duration 25s and *GRB* 940217 at 18000 MeV has spectral index 2.5 ± 0.1 and duration 180s. However, in Ryan et al. (1993), they show the photon spectrum with a data-fitted power-law equal to $7.5 \pm 3.1 \times$ (E/1MeV)$^{(-1.8 \pm 0.4)}$ cm^{-2} s^{-1} MeV^{-1}.

In (Gilmore et al., 2009b), a high-energy spectral index of -1.95 for prompt phase photons is assumed, consistent with EGRET results, and a harder spectral

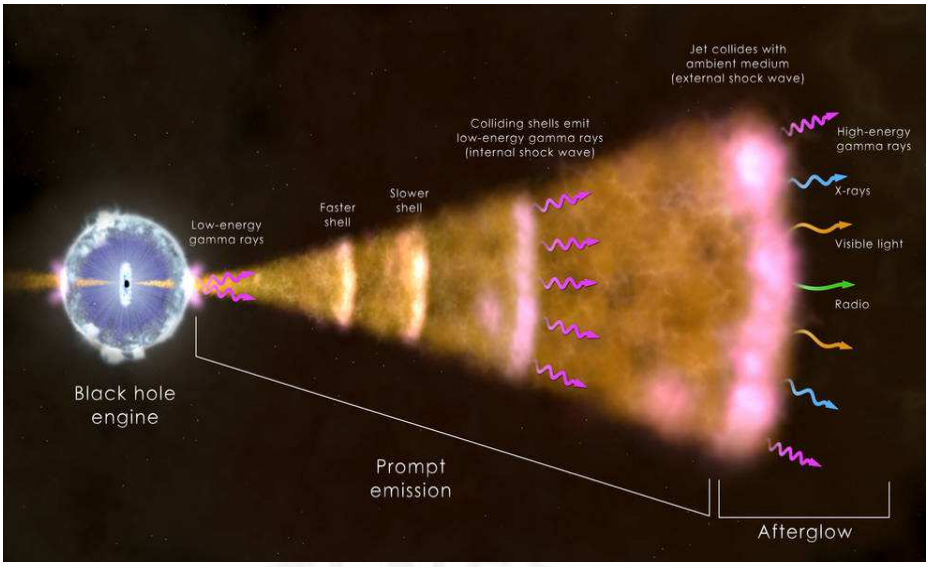

Figure 3.3: Most common type of gamma-ray burst. The core of a massive star (left) has collapsed, forming a black hole that sends a jet moving through the collapsing star and out into space at near the speed of light. Credits: NASA's Goddard Space Flight Center

index of -1.5 for the afterglow component.

Finally, in this work, for shower simulation of cosmic and gamma rays, we choose the energy spectral index for the primaries according to the energies under study. Thus for cosmic rays the index will be 2.7 and for gamma rays 2, as a compromise of the data presented before, for a generic GRB.

3.2.2 Other sources of gamma rays

There are many other sources of gamma rays. We introduce some to compare their fluxes with the flux of cosmic rays. A source of gamma rays is PG 1553+113 a distant BL Lac, a type of Active Galactic Nuclei (AGN) with a power-law fit for its energy spectra with a spectral index of 4.0. In this source of gamma-rays, the flux above 200 GeV corresponds roughly to 2% of Crab Nebula flux; since high-energy gamma rays are more strongly absorbed. Surviving gamma rays for very distant sources are expected to show a flux which rapidly decreases towards high energy (Aharonian and et. al. , H.E.S.S. Colaboration).

The energy spectra of other sources of gamma rays have a power law fit with a spectral index of 2.1. One of these sources is supernova remnant $RX\ J0852.0-4622$ which is a strong source of extended TeV emission spatially correlated with X-rays. The flux of gamma-ray of RX $J0852.0-4622$ is 7×10^{-11} erg cm^{-2} between 1 and 10 TeV (HESS-Collaboration et al., 2005).

We also use the energy spectra of blazar Mrk 421, another type of AGN, with data taken between January 19, 2009 and June 1, 2009 by different experiments (Krawczynski and Treister, 2013).

3.3 Flux of cosmic and gamma rays

In this work we will compare the fluxes of cosmic rays and gamma rays. We extract data manually from a compilation of cosmic ray spectra, updated with respect to the compilation by T.K.Gaisser. Cited experimental data for protons are: AMS, BESS, ATIC, JACEE, KASCADE(SIBYLL), TibetIII(SIBYLL); all-particle, Tibet(SIBYLL), KASCADE(SIBYLL), Akeno, GAMMA, TUNKA, Yakutsk, Auger, AGASA, HiRes; and for γ:, EGRET extracted from Hun (2009).

We also include energy spectra of gamma rays from HESS $J1616-508$, which is one of the brightest emitters in the TeV sky. In Landi et al. (2007) they conclude that there is good evidence to assume that the HESS $J1616-508$ source is driven by the pulsar PSR $J1617-5055$ in which a combination of synchrotron and inverse Compton processes create the observed morphology of a broad-band emitter from keV to TeV energies. HESS $J1616-508$ has a spectra with slope 2.3 with an exponential cutoff at 40 TeV. In Landi et al. (2007) they present a 1-yr HAWC data simulation with cutoff and without to demonstrate that HAWC could distinguish the spectrum.

Now, having the gamma ray fluxes of various sources and the cosmic ray flux, we compare the spectra to be able to deduce the difference between the fluxes of gamma rays and cosmic rays and thus be able to make more realistic considerations to distinguish photons from protons. In Fig.3.4 the comparison of the fluxes is shown. The expected photon flux from single sources compared to the cosmic ray flux is much lower: 10^{-6} at 10^{13} eV approximately. Taking this into account we will later simulate 10^4 photon showers as signal and 10^6 showers of protons as background. More background simulations would have been needed to account for a realistic and statistically sound ratio between expected protons and photons, but due to computational resources and time constrains, only a smaller sample was produced.

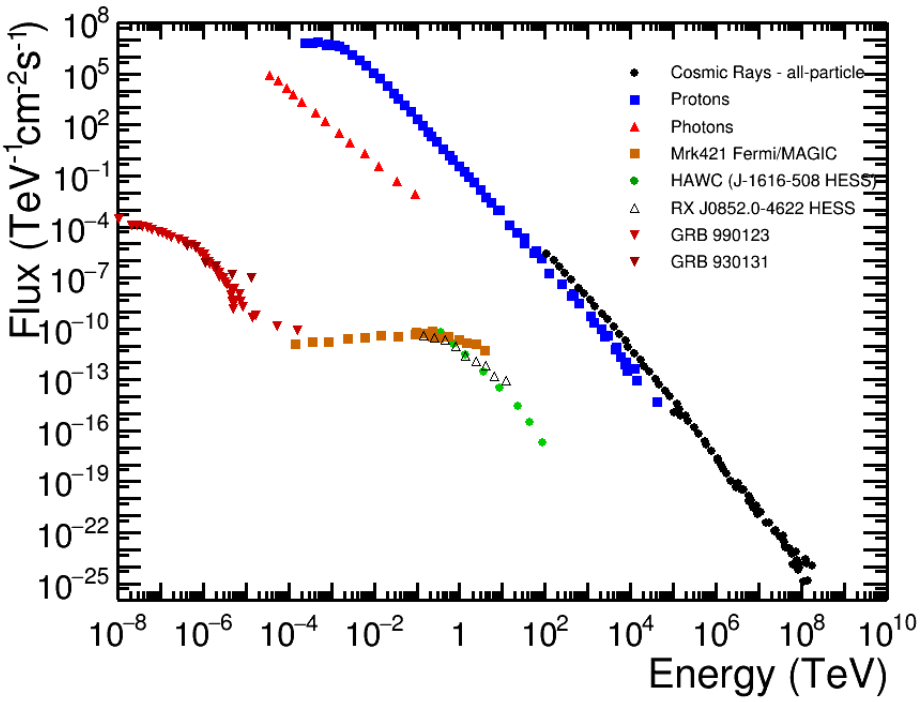

Figure 3.4: Comparison of cosmic ray (Hun, 2009), proton and gamma-ray fluxes from different sources from Landi et al. (2007), Krawczynski and Treister (2013), HESS-Collaboration et al. (2005), Aharonian and et. al. (H.E.S.S. Colaboration), Gilmore et al. (2009b), Ryan et al. (1993), Merck et al. (2016) and HESS-Collaboration et al. (2014).

Chapter 4

Cosmic and Gamma Rays Detectors

The detection of astroparticles at high energies is mostly indirect. There are effects that occur when these astroparticles arrive at the Earth's atmosphere that help us to identify them indirectly. Two of these effects are going to be mentioned in this work: the Cherenkov effect and the fluorescence effect.

The Cherenkov effect occurs when the speed of a charged particle in a medium (e.g. water, air or ice) is greater than the speed of light in it ($c_n = c/n$, where n is the refraction index). Then radiation, called *Cherenkov* light, (see Fig. 4.1) will be emitted. The main characteristic of Cherenkov light is the critic angle of the Cherenkov cone of light emitted (Grupen, 2005), which is:

$$\theta_c = arccos(1/n\beta) \tag{4.1}$$

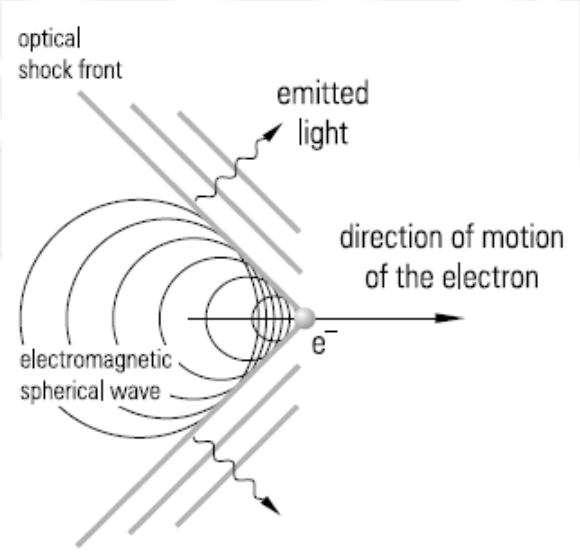

Figure 4.1: Emission of Cherenkov radiation by an electron traversing a medium (Grupen, 2005)

This radiation can be recorded, via the photoelectric effect, by photomultipliers (PMTs) tubes. This is then used to reconstruct the energy and the incidence

direction of the muon, which is a particle produced from hadronic, electromagnetic or muonic showers.

The fluorescence effect occurs when charged particles produced in atmospheric showers interact with the atmosphere, exciting and ionizing the air molecules. In this process atmospheric nitrogen emits UV-radiation isotropically, called *fluorescence* light, whose emission spectrum is concentrated in the band of 300 to 400 nm (Melo, 2007). A characteristic of fluorescence in the atmosphere is that it is isotropical and its intensity depend of the particle's number. The number of fluorescence photons, N_γ, produced by a number of charged particles, N_e, moving a distance dl is (Melo, 2007):

$$dN_\gamma = Y_f N_e dl \tag{4.2}$$

where Y_f is called fluorescence yield, which is the number of fluorescence photons produced by single charged particles by unit of longitude.

Fluorescence light is recorded by detectors as seen in Fig. 4.2). The detection depends on the development of the electromagnetic cascade with energy, related to N_{max} (maximum number of particles), and the composition, related to X_{max} (atmospheric depth when the number of particles is maximum).

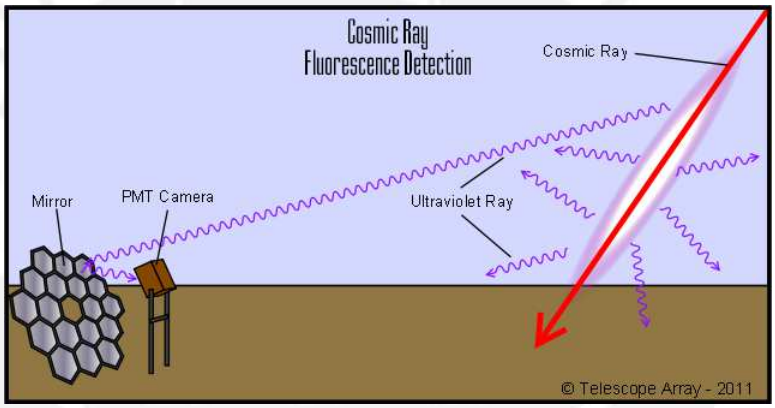

Figure 4.2: Schema of a fluorescence air shower detector extracted from http://www.telescopearray.org/

Detectors on the ground sample the shower at one particular depth, but contain information about the lateral distribution or longitudinal profile of particle's showers. There are many experiments that use detectors with the effects describe before. In this work three experiments are described: HAWC, Pierre Auger and LAGO.

4.1 HAWC

The High-Altitude Water Cherenkov Gamma-Ray Observatory or HAWC [*www.hawc−observatory.org*]uses detectors with the water-Cherenkov technique. It was designed to detect gamma rays and cosmic rays at TeV energies.

HAWC is constructed on the slopes of the Sierra Negra volcano, close to Puebla, Mexico at 4100 m.a.s.l. HAWC was proposed for surveying and monitoring the high energy gamma-ray sky. Water Cherenkov detectors (WCDs)like

Milagro and HAWC function as wide field monitors of photons and cosmic rays of energies from 100 GeV to 100 TeV, operating continuously without interruptions due to weather conditions (H. Salazar, 2009). The observatory consists of an array of 300 WCDs, as shown in Fig. 4.3, covering an area of 22000 m^2. Each WCD consists of a tank 7.3 m in diameter and 4.5 m in depth.

Figure 4.3: Cherenkov tanks of HAWC Observatory. Extracted from *http* : *//www.inaoep.mx/ hawc/*

The data recorded by HAWC works with two acquisition systems (DAQs), the main DAQ and the scaler DAQ. The main DAQ measures the arrival time and time over threshold of PMTs pulses, giving information to reconstruct the shower core, direction and lateral distributions, which helps to distinguish primary particles and their energies. The scaler DAQ, operates in a PMT pulse and is sensitive to gamma rays, GRBs in particular, in the GeV and TeV range using the single particle or scaler technique. You can see a signal with respect to atmospheric showers or noise.

Fig. 4.4 shows the minimum flux for 5 σ detection of a source with a zenith angle of 20° and a duration of 10^{-2} to 5 x 10^3 s for the two systems DAQs. Also shown is the sensitivity of Fermi LAT for 10 GeV photon detection and the effects of EBL (modeled by Gilmore et al. (2009a)) at various values of redshift (Abeysekara et al., 2011).

Last year, members of the HAWC collaboration presented a new gamma-ray sky map at the annual meeting of the American Society of Physics in Salt Lake City, Utah. The new map shows many objects along the Milky Way plane, some of these had not been previously observed (*www.hawc−observatory.org/news/*).

4.2 Pierre Auger

The Pierre Auger Observatory (PAO) [*www.auger.org*] is in Malargue, Argentina at 1400 m.a.s.l. and was designed to resolve the enigmatic ultra high energetic cosmic rays (UHCR) that means the energies above 10^{18} eV. This observatory can determine the energy, arrival direction and the nature of UHCR. This observatory is mentioned as an example because its sensitivity starts at much higher energies than those covered in this work.

The technique of PAO to detect cosmic rays is 'hybrid'. This technique consist

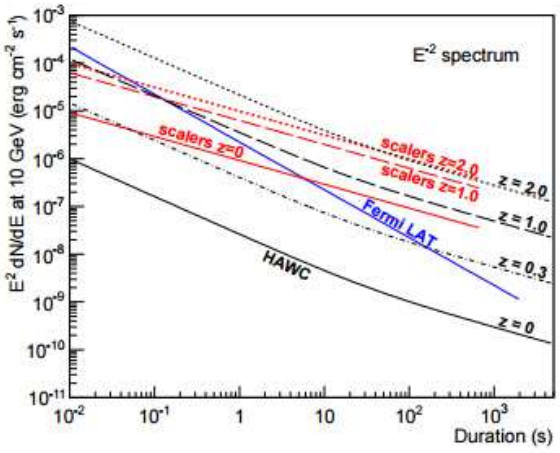

Figure 4.4: Sensitivity of HAWC for the source spectrum $E^{-2.0}$. The Gilmore model is used to obtain the sensitivity curves for different redshifts. Also a marker is inserted for GRB 090510 (Abeysekara et al., 2011)

.

of an array of Cherenkov surface detectors and a system of atmospheric fluorescence telescopes to observe the showers of secondary particles.

4.2.1 Surface detectors

Surface's detectors are 1660 water surface detector tanks that cover about 3000 km^2 of the Pampa and serve as particle detectors. Each 12000 liter tank, separated from each of its neighbors by 1.5 km, is completely dark inside, except when particles from a cosmic ray air shower pass through it(Auger-Collaboration, 2015). These detectors use the Cherenkov effect and measure the Cherenkov light with photomultiplier tubes mounted on the tanks. In Fig. 4.5 we show the Cherenkov tank of the PAO.

Figure 4.5: Cherenkov Tank of Pierre Auger Observatory (Auger-Collaboration, 2015).

4.2.2 Fluorescence detectors

Fluorescence detectors are telescopes (Auger-Collaboration, 2015). This system of detectors is constituted of 24 telescopes within 4 buildings. Its range is greater than 30 km for showers of 10^{20} eV. Each telescope have spherical surface mirrors of $3,6$ m $\times$ $3,6$ m. These telescopes have a camera too. Their cameras have 440 photomultiplier tubes each (see Fig. 4.6).

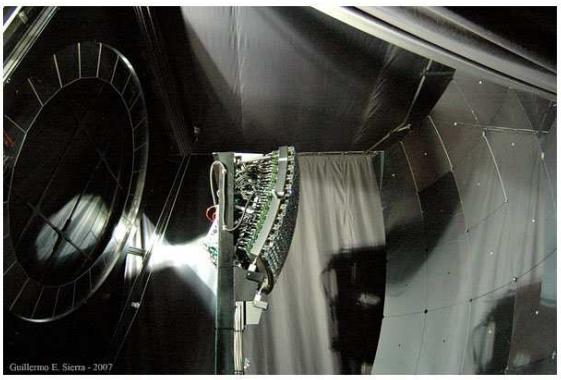

Figure 4.6: Fluorescence Telescope of Pierre Auger Observatory (Auger-Collaboration, 2015).

4.2.3 Pierre Auger and Gamma rays

Although PAO was designed to study high energy cosmic rays, there are studies Allard et al. (2005) that look into the capability of PAO to detect the high energy emission of GRBs with the single particle technique Aglietta et al. (1996). According to the detector response to photon showers around one GeV, and making reasonable assumptions about the high energy emission of GRBs, they show that PAO is a competitive instrument for this technique.

The Single Particle Technique allows some observatories to research gamma rays bursts at GeV. This technique consists in detecting small photon showers arriving isolated at ground level, but with a intense emission such that could be significant compared with the background fluctuations (Vernetto, 2005).

4.3 LAGO

The Latin American Giant Observatory (LAGO) project [*lagoproject.org*] is a collaboration that is installing WCD in high altitude sites Otiniano et al. (2015). Actually LAGO is in Argentina, Bolivia, Colombia, Ecuador, Guatemala, Mexico, Peru, Venezuela and Brazil. LAGO was created to observe the high energy component of Gamma Ray Bursts (GRBs) in arrays of Water Cherenkov Detectors (WCD) at high mountain sites at 10 GeV using the single particle technique (Aglietta et al., 1996). There have been several tests with Cherenkov tanks near to Huancayo city, as shown in the Fig. 4.7

Part of the motivation of this work started because the LAGO project searched to establish a detector of gamma rays in our country, Perú. Actually the LAGO

Figure 4.7: Testing Cherenkov tanks in Huancayo, Peru for LAGO project.

project in Peru is planning the installation of an array of WCDs at high altitude (4600 m a.s.l.)in the central highlands of Peru Otiniano et al. (2015).

Chapter 5

Air Shower Simulations with CORSIKA

5.1 CORSIKA

CORSIKA (COsmic Ray SImulation for KAscade) is a program for simulating extensive air showers initiated by cosmic ray particles. It was developed for simulations for the KASCADE experiment at Karlsruhe, Germany. It is a Monte Carlo program with FORTRAN routines (Heck et al., 1998). It gives the type, energy, location, direction and arrival times of all secondary particles that are created in an air shower and pass a selected observation level.

Particles as protons, light nuclei up to iron and photons may be treated as primary particles from 10^{11} eV up to some 10^{20} eV.

CORSIKA operates with input and output and consists basically of four parts (Heck et al., 1998):

- Decay and tracking of the particles taking into account ionization energy loss and deflection by multiple scattering and the Earth's magnetic field.
- Hadronic interactions at higher energies.
- Hadronic interactions at lower energies.
- Transport and interaction of electrons, positrons, and photons.

5.2 Interaction models

CORSIKA can simulate high energy hadronic interactions with different models (Heck et al., 1998). For this work the QGSJET II-04 model (Quark Gluon String model with JETs) was chosen, because it has the latest LHC data. Low energy hadronic interactions are simulated with GHEISHA 2002*d*, because this model also includes the latest LHC data.

The electromagnetic interactions can be treated following each particle with EGS4, or using the analytic NKG formula to obtain electron densities and the total number of electrons at selected locations (Heck et al., 1998).

The installation of CORSIKA will be detailed in the appendix A of this work.

5.3 Input file

The input file controls the parameters of the simulation that can be modified according to the configuration that we want to simulate. In first place, the primary type of the cosmic ray should be defined. For this work protons (14) and photons (1) have been used (the numbers corresponding to each different particle are presented in the CORSIKA manual (Heck et al., 1998)). Also the primary particle energy with a given spectrum slope can be defined. We can fix the value or choose a particle energy range.

As a first approach, we chose four fixed energies 10^2 GeV, 10^3 GeV, 10^4 GeV and 10^5 GeV. Later, considering a more realistic scenario, we chose an energy range from 10^2 GeV to 10^5 GeV with 2.7 of slope for protons and 2 of slope for photons, as described in Sec. 3.2.1.

The primary angle of the cosmic ray can be predefined. The particle flux arriving at Earth can be assumed with constant intensity from all directions of the sky. However, at the surface it will have an angular dependence given the different atmosphere path that they traverse. For a simplified study we only consider protons and photons arriving in the vertical direction.

Different atmospheric depths can be defined as the observation levels, we chose the altitude of Pierre Auger: 1400 m.a.s.l. The geomagnetic field can be defined too, for the ground site of interest so that the trajectory deflection for the charged particles can be taken into account. We kept the other parameters as by default as we show in the section Input of the appendix A.3.

5.4 Output file

The output files of the simulation depend on the installation and some options defined in the input file.

In this work, since we decided to analyze the data with ROOT (Guide, 2015), we chose as output file a root file, usually named as DAT000###.root. Also we chose the option to plot the shower development with output files named DAT000###.track.all.ppm, which we then convert to jpg.

These kind of graphs can be seen in Fig. 5.1 for two primary particles: a proton (left) and a photon (right) respectively at 10^3 GeV of energy. The different colors represent the kind of showers according to particle type. The red lines are for photons, electrons and positrons, the green lines are for muons and antimuons and finally the blue lines are for hadrons. The form of the shower development is caused by secondary particles loosing energy while penetrating the atmosphere. The maximum width of the shower corresponds to the maximum number of particles. After this point the number of particles decreases. More detail about this can be found in Sec. 2.2.

The CORSIKA scripts are run in a condor computer cluster, which distributes the jobs in the farm, parallelizing the execution of the scripts and thus reducing the total time of these intensive computing calculations. Details about job submission can be found in appendix B.

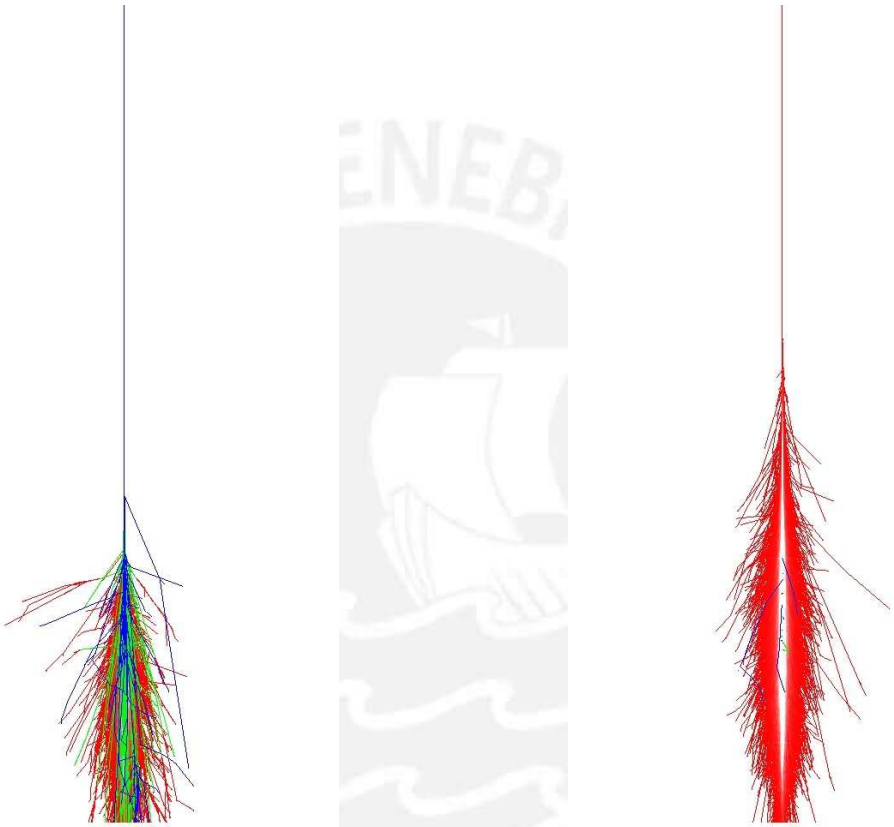

Figure 5.1: Shower development for (Left) a proton and (Right) a photon at 10^3 GeV.

Chapter 6

Extensive Air Showers' Parameters

Particle showers in the atmosphere are detected by different experiments, as described in Chapter 4. These detectors obtain information about the showers, which is related to some parameters that we study in this Chapter. First we describe the longitudinal profile. Then we introduce the fitting of this longitudinal profile. Finally we present the principal parameters that we will use to distinguish showers initiated by a photon or by a proton. The use of these parameters will be presented in the next chapters.

6.1 Longitudinal profile

The longitudinal profile or number of particles vs. atmospheric depth is a smooth curve with a well-defined maximum called maximum depth or X_{max}. With this longitudinal profile we can study the characteristics of the particle showers produced by some primary particle. In Fig. 6.1 we show a longitudinal profile for ten showers with a proton and a photon as primary particle.

The measurement of the longitudinal profile allows to reconstruct the energy of the primary particle and to obtain information about the type of particle (Prado and De Souza, 2013). In our case it helps us to distinguish between photons and protons. One form to obtain this information is through X_{max}, because X_{max} is the most basic parameter of a particle shower and increases with the energy of the primary (Pryke, 2013) particle, as shown in Fig. 2.4.

In order to reconstruct the longitudinal profile, we have to assume a functional form of the longitudinal development and fit this profile to the measured points. Then, the X_{max} can be determined directly from the fitted function (Prado and De Souza, 2013). Also with this fitted function we can extract other parameters related to the longitudinal profile.

6.2 Longitudinal profile fitting

There are several functions proposed to fit a longitudinal profile. The Gaisser-Hillas function (Gaisser and Hillas, 1977), the Greisen function and the Gaussian in shower age Abu-Zayyad et al. (2001) have been traditionally used. Here we

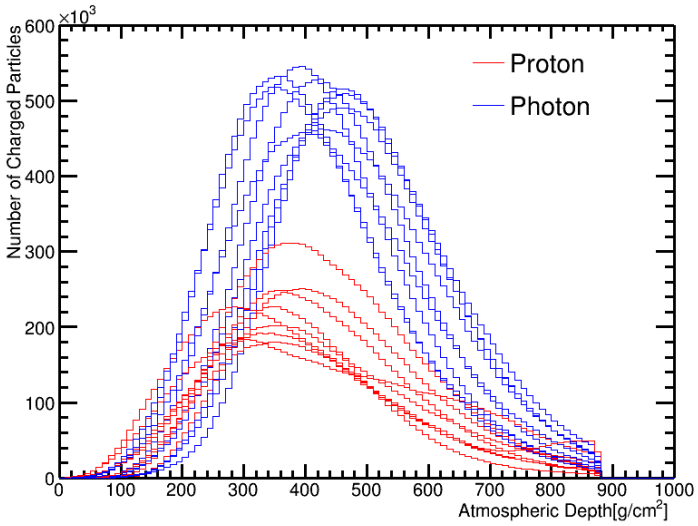

Figure 6.1: Number of charged particles vs. atmospheric depth generated by CORSIKA. Red lines are for protons and blue lines for photons. Each different line represents a shower.

present the Gaisser-Hillas formula with four parameters, which gives the approximate number of charged particles as a function of atmospheric depth along the shower axis (Al-Rubaiee and Jumaah, 2013):

$$N(X) = N_{max}\left(\frac{X - X_0}{X_{max} - X_0}\right)^{\frac{X_{max}-X_0}{\lambda}} e^{(X_{max}-X)/\lambda} \tag{6.1}$$

where X is the atmospheric depth (in g/cm^2); N_{max} is the number of charged particles; X_{max} is the depth of shower maximum; X_0 is the depth of the point of first interaction and λ is a characteristic length parameter according to Al-Rubaiee and Jumaah (2013). In Rahman et al. (2001), they mention that λ is the proton interaction mean free path in air (70 g/cm^2) and for Abu-Zayyad et al. (2001) λ is the shower decay length.

X_0 depends on the collision cross section and therefore on the energy and mass composition of the particle. X_{max} depends on the position of X_0, the shower energy and the composition (Al-Rubaiee and Jumaah, 2013).

X_{max} varies from event to event, for this reason one can define the shower 'age' that depends on X as:

$$s = \frac{3X}{X + 2X_{max}} \tag{6.2}$$

The shower 'age' (s) varies from 0 (initial position of the shower) to 1 (shower maximum) and decays from 1 to 3 (infinite depth) (Abu-Zayyad et al., 2001). For the physical shower it varies from 1 to 2. Then introducing s in Eq. 6.1 and using a normalized shower size $n = N/N_{max}$ (Al-Rubaiee and Jumaah, 2013):

$$n(s) = \left(1 - \frac{1-s}{3-s}\frac{3T_m}{T_m - T_0}\right)^{T_m - T_0} e^{3T_m(1-s)/(3-s)} \tag{6.3}$$

where $T_m = X_m/\lambda$ and $T_0 = X_0/\lambda$ are two remaining parameters.

The second function to fit the longitudinal profile is the Greisen function, which describes the development of the electromagnetic shower (Abu-Zayyad et al., 2001). The depth is expressed as a function of radiation length, $L_0 = 36.66$ g/cm^2. This function has a single parameter $y = X_m/L_0$ and has the form:

$$N(T) = (\frac{0.31}{\sqrt{y}})e^{T(1-\frac{3}{2}ln(s))} \tag{6.4}$$

where $T = X/L_0$ is the atmospheric depth in radiation lengths in the air. Also Greisen shows that $y = ln(E_0/E_c)$, with E_c, is the electron critical energy (Abu-Zayyad et al., 2001). In Abu-Zayyad et al. (2001), they propose a new Gaussian function that reduces the number of parameters. This function has the form:

$$n(X) = exp[-\frac{2}{\sigma^2}(\frac{X - X_m}{X + 2X_m})^2] \tag{6.5}$$

where the single parameters are X_{max} and σ to indicate the shower width and it is $scale - free$ (Abu-Zayyad et al., 2001).

In this work, in order to choose the best function, we have compared these functions using the reduced χ^2 (χ^2_{red}). When χ^2_{red} is near 1, then it is a better fit. In Fig. 6.2 and 6.3 we show the fit of the longitudinal profile of a photon and a proton using these three functions. In the legend we can see the χ^2_{red} for these functions. The conclusion of this part is that the best function is Gaisser-Hillas, which we then use to fit the longitudinal profile of the simulated showers.

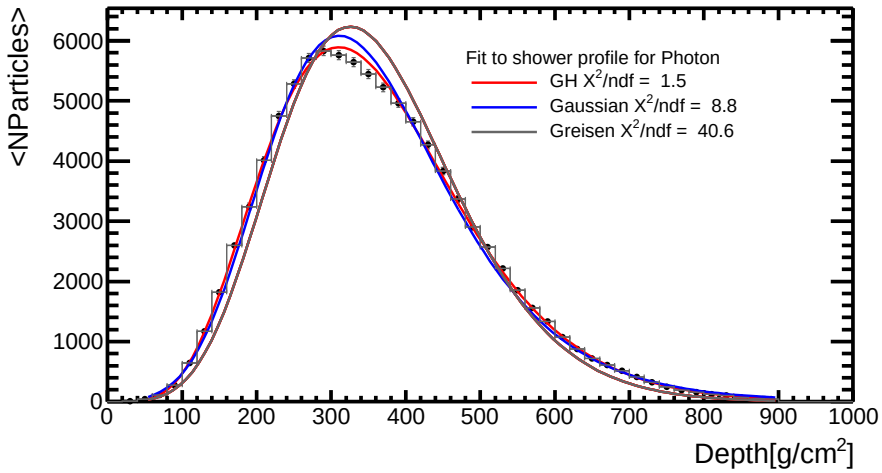

Figure 6.2: Shower development for a 10^2 GeV photon.

6.3 Longitudinal profile parameters

After fitting the longitudinal profile of the showers we extract the parameters associated to the fitted functions, which are:

- the energy or E.
- the depth of shower maximum or X_{max}.
- the parameter related to depth of the point of first interaction or X_0.

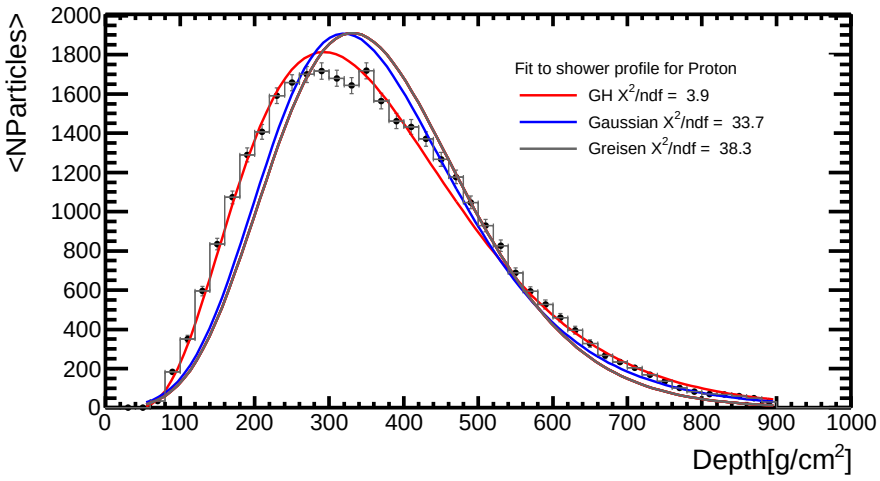

Figure 6.3: Shower development for a 10^2 GeV proton.

- the number of charged particles or N_{max}.
- the shower decay length or λ.
- the reduced Chi-squared of the fit χ^2_{red}.

6.4 Other parameters extracted from the longitudinal profile

There are other parameters that we have calculated. From the longitudinal profile we measured the point at which the shower begins, which we call X_{start}. This is related to the fitted parameter X_0. We also mapped the root mean square or RMS, which is defined as the square root of mean square (the arithmetic mean of the squares of a set of numbers)(Dictionary, 2009). In our case the RMS equals $\sqrt{x^2+y^2}$. Where x is the total value in the x-axis and y is the total value in the y-axis.

Later, motivated by the paper of Matthews et al. (2009), new parameters are extracted from the longitudinal profile. These parameters are the shower full width at half-maximum, $FWHM$, and shower asymmetry parameter, f. The shower full width at half-maximum is defined as $FWHM = L + R$ and the shower asymmetry parameter as $f = L/(L+R)$. Where L is the left part of the longitudinal profile and R is the right part of the longitudinal profile assuming that the longitudinal profile are not symmetric. We show in Fig. 6.4 an example of the extraction of these parameters after fitting the longitudinal profile.

In summary these parameters are:

- the point of shower start or X_{start}.
- the root mean square or RMS.
- the shower full width at half-maximum or $FWHM$.
- the shower asymmetry parameter or f.

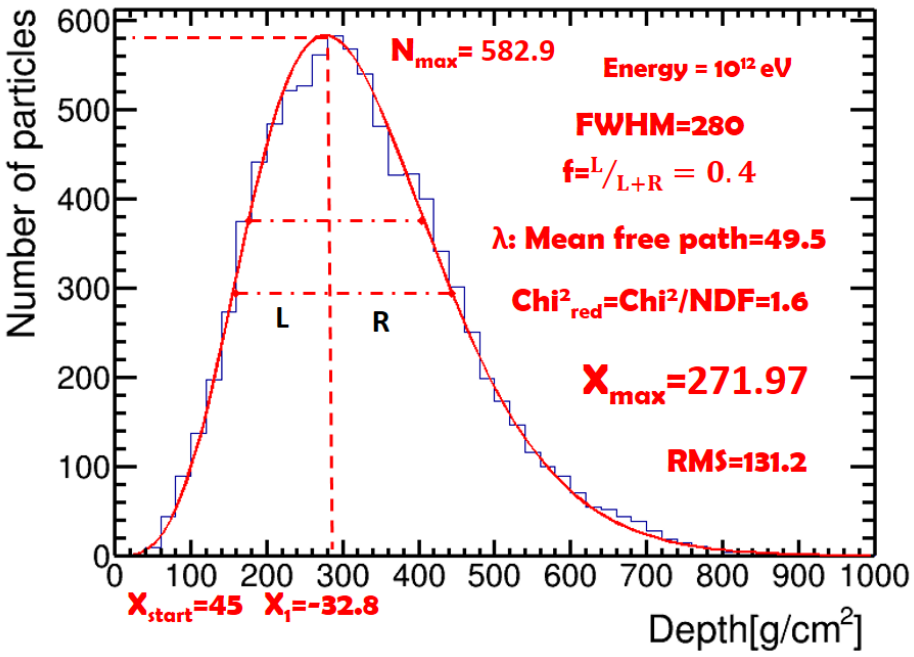

Figure 6.4: Example of parameters extracted from the longitudinal profile of the shower after fitting with the Gaisser-Hillas function

6.5 Energy parameters and gaussian smearing

Another important parameter is the energy, since the shower characteristics depend strongly on it. In this work, we have first simulated four fixed energies 10^2 GeV, 10^3 GeV, 10^4 GeV and 10^5 GeV, as mentioned in Sec. 5.3. For each separate set we will apply the cuts of the next chapter 7 in order to find a relation between some important cuts parameters and the energy.

In addition, we considered a more realistic simulation with energies ranging from 10^2 GeV to 10^5 GeV that followed a different spectral index for protons and photons. The cut-value for each single energy mentioned above was used as a benchmark for the whole energy range, finding an energy dependence for the cut parameters in order to better separate signal from background events. Since we also wanted to include the energy as a further parameter in the analysis, we applied a Gaussian smearing to the energy to introduce a realistic detector energy resolution. We used a Gaussian with $\sigma = 0.5$ of the energy, taking as an example HAWC's energy resolution, which at 10 TeV is 50% of the real energy approximately. Particles observed beyond this energy in HAWC will reconstruct the energy inside of 50% of the real energy (info extracted from $http://www.hawc-observatory.org$). Fig. 6.5 shows the simulated energy vs. the smeared energy used in this work.

6.6 Other parameters

There are other parameters that we could extract from the particle shower. These parameters are related to the time and the spatial distribution.

The time distribution is extracted from the simulation with CORSIKA, we

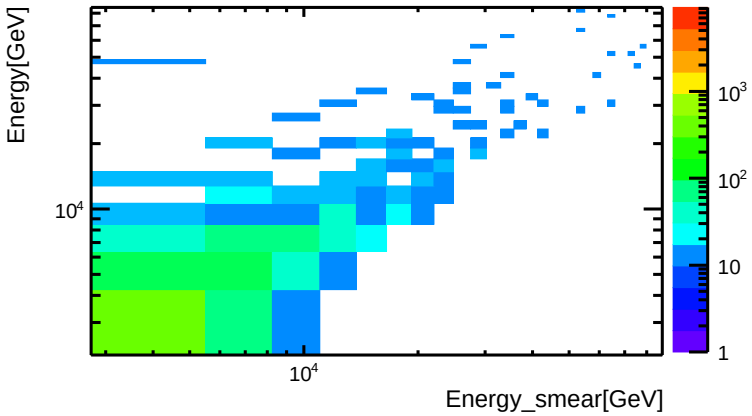

Figure 6.5: Simulated versus smeared energy.

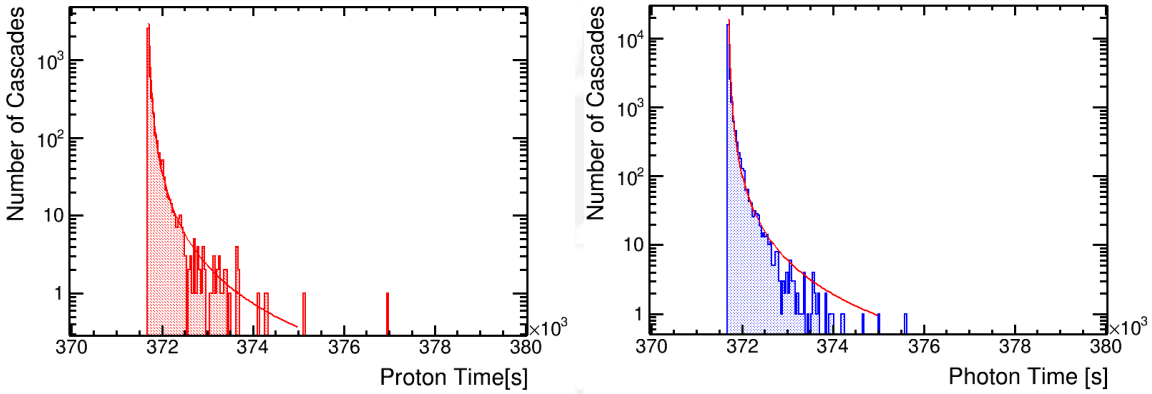

Figure 6.6: Time distribution of a shower from CORSIKA simulations for a proton (left) and a photon (right) at 10^3 GeV.

have fitted the distribution with a Landau function for a photon and a proton at energies 10^2 GeV, 10^3 GeV, 10^4 GeV and 10^5 GeV. In Fig. 6.6 we show an example of the time distribution that we have fitted.

From the spatial distribution it is also possible to extract a numerical parameter but given the time constrains we only show in Fig. 6.7 the form of this distribution.

However, the parameters from fitting the time distribution and the radius of spacial distribution have not been used in this work since there were a few problems related to low energies and processing time.

In CORSIKA the clock and coordinate system are reset with each air shower. Therefore the showers are uncorrelated in space/time. However the shower particles are distributed relative to shower start time and start position. The solution to this is to distribute the primary showers in space/time based on measurements of the intensity of primary nucleons Bass (2015), a problem that we have not addressed.

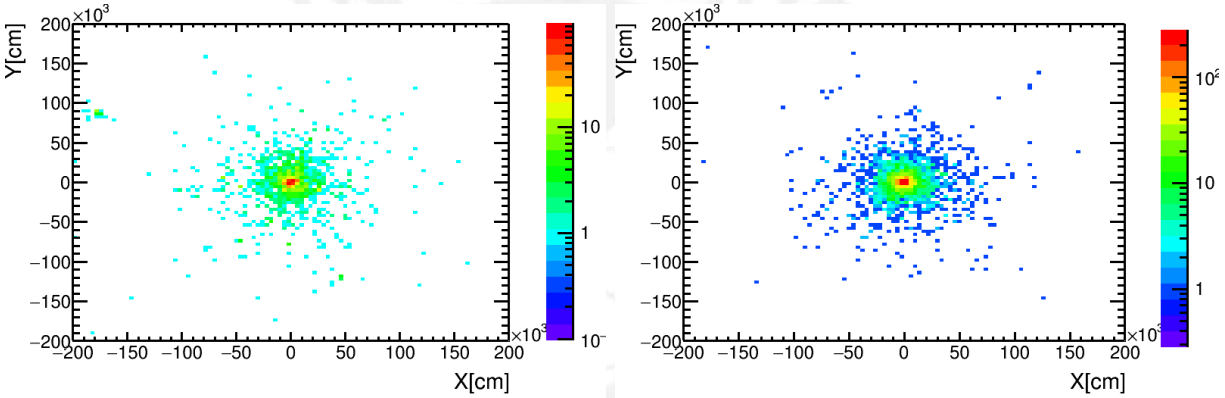

Figure 6.7: Spacial distribution of a shower from CORSIKA simulations for a proton (left) and a photon (right) at 10^4 GeV.

Chapter 7

Simple Cuts Method

In this Chapter a straightforward cut analysis for separating protons from photons will be introduced. It uses the parameters described previously. This will represent a benchmark for the more advanced multivariate method presented in the next Chapter.

The simple cuts method consists in finding the optimal numerical cut values (greater or smaller) for each of the different variables (parameters) that reduce background events preserving signal events. First these cuts will be chosen according to physical considerations. Later we will choose different variable/value pairs in order to find the best cuts based on an efficiency plot, which will also depend on the energy.

7.1 Initial cuts

The first cuts that we apply to the parameters distributions are: $X_{max} < 900$ and $\chi^2_{red} < 200$, which can be seen in Table 7.1. These cuts were apply in all parameters at all energies as we describe in the next section 7.2

The cut applied to X_{max} was used because $900[g/cm^2]$ is the atmosphere depth at the observation level chosen before. A value higher than this would not be possible for the longitudinal development. The cut applied to χ^2_{red} was

Cuts	**Signal**	**Background**	$S/\sqrt{S+B}$
$N_{max} > 40000$	46.1%	17.8%	5.77
$\chi^2_{red} < 10$	55.1%	44.9%	5.51
$X_0 < 0[g/cm^2]$	69.5%	37.3%	6.73
$X_{start} < 100[g/cm^2]$	81.4%	65.1%	6.73
$f > 0.25$	69.1%	35.7%	6.75
$FWHM < 375[g/cm^2]$	83.8%	69.2%	6.77
$X_{max} > 250[g/cm^2]$	93.5%	85.3%	6.99
$RMS < 150$	77.5%	44.9%	7.01
$\lambda < 65$	93.4%	59.2%	7.56

Table 7.1: Efficiency for different cuts for protons and photons after initial cuts. The third column is the significance.

done because there are many longitudinal profiles with non-gaussian distributions, having χ^2_{red} much larger than one. The value 200 corresponds to a conservative number that do not decrease much the signal according to our results.

In Fig. 7.1 we show the X_{max} distributions after applying these cuts to photons and protons. The efficiency of the cuts was 93.2% for photons and 86.6% for protons including all energies. In general we see that the X_{max} distribution depends on the energy.

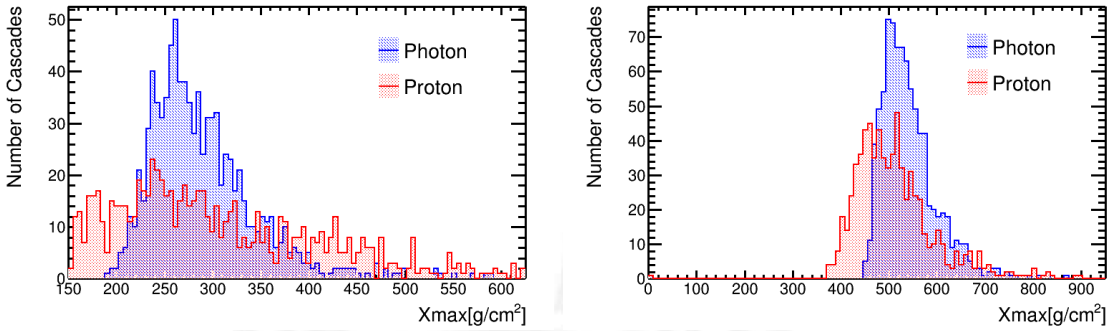

Figure 7.1: Distribution of X_{max} for photon (blue) and proton (red) for 10^2 GeV (left) and 10^5 GeV (right). In these plots the following initial cuts are applied: $X_{max} < 900$ and $\chi^2_{red} < 200$.

7.2 Cuts in all parameters at all energies

A simple method was used to understand the multivariate analysis. This method was called 'Simple Cuts Method'. In first place, we use the distributions of all parameters at different energies. Later, we determine for each parameter, simple cuts, which split signal from background. In our case the signal refers to photons and the background to protons.

In Fig. 7.2 we show the efficiency plot comparing signal and background at the same time.

Each point corresponds to the fraction of events left after applying the cut at the value given in the x-axis, where the direction of the cut (greater or lower than) depends on each variable. In the case of X_{max} it keeps the events with a numerical value greater than the cut value. Thus, we determine the best cut that splits most efficiently the signal from the background, i.e. where the difference in signal and background efficiency is the largest, while keeping most of the signal.

The results of applying the cuts, chosen in the way described above, are shown in Table 7.1 including all discrete energies. We can see that the cut that have less background is N_{max}, which splits the signal with 46.1% from background with 17.8%. But according to the significance ($S/\sqrt{S+B}$) the best cut is λ, since its significance is higher. The significance is an indicator of the best cut, where the signal is the highest possible and the background is the lowest. However, we need to evaluate how these cuts vary with the energy.

We repeat this method for individual energies to see the dependence of the parameters with the energy. Then we obtain that the best cut is the N_{max}

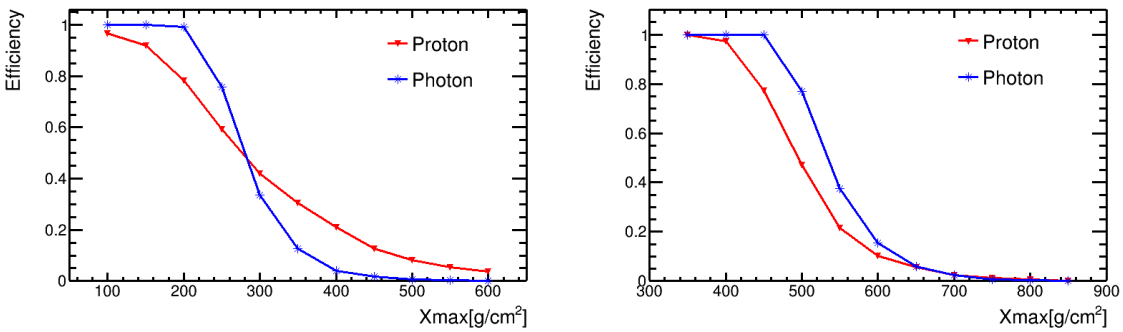

Figure 7.2: Efficiency for the distributions of X_{max} for photon (blue) and proton (red) at 10^2 GeV (left) and 10^5 GeV (right). The initial cuts ($X_{max} < 900$ and $\chi^2_{red} < 200$) are already included.

Cuts for N_{max}	Signal	Background
$> 5.0 \times 10^2$ at 10^2 GeV	98%	6.5%
$> 5.0 \times 10^3$ at 10^3 GeV	85%	0.01%
$> 4.0 \times 10^4$ at 10^4 GeV	96%	0.08%
$> 3.6 \times 10^5$ at 10^5 GeV	98%	1.9%

Table 7.2: Efficiencies for best N_{max} cut for protons and photons.

parameter (see Table 7.2 and Fig. 7.3 and 7.4). These cuts were modified because the Nmax varies with energy and the first cut was made in an area where there was more signal than background.

Finally, we evaluate all cuts found and apply them to each energy. These cuts and their results are shown in Table 7.3

7.3 Applying cuts under realistic considerations

We show in Fig. 7.5 the relation between the best cut values of some parameters with the energy. This gives an idea about the variation of the cuts with energy, which will be next used for more realistic considerations.

The fit function for X_{max} was a straight line $a + b * x$, but with the logarithm of the energy, where a is -1.3 ± 20.6 and b is 88.3 ± 5.62. For $FWHM$ was also a straight line $a + b * x$ with the logarithm of the energy, where a is 268.5 ± 3.18 and b is 26.5 ± 0.86. For N_{max} was a exponential $a * exp(b * x)$, where a is 1.24 ± 0.025 and b is 2.31 ± 0.005. Finally for X_{start} was a polynomial of degree two $a + b * x + c * x^2$, where a is -44.25 ± 19.25, b is 59.25 ± 11.84 and c is -11.25 ± 1.68. For simplicity we use the logarithm for the energy axis.

Now, we apply these energy-dependent cuts and other fixed cuts to our sample with a realistic spectra. These cuts are show in Table 7.4. The results achieved are shown using the energy with and without smearing. There are 54% signal events left and 12% background events when the energy smearing is considered. These results will be used as benchmark for the multivariate analysis using the

Cuts/Energy	10^2 **GeV**	10^3 **GeV**	10^4 **GeV**	10^5 **GeV**
X_{max}	> 185	> 250	450	> 446
	< 600	< 600	< 600	–
N_{max}	> 400	> 3735	> 35000	> 352000
$FWHM$	–	< 470	< 480	< 421
	> 175	–	–	–
f	> 0	> 0	> 0.4	> 0
	< 1.22	< 2	< 2	< 2
RMS	> 100	> 118	> 106	–
	–	–	< 180	–
X_{start}	> 30	> 30	> 15	–
	< 140	< 140	< 140	–
X_0	< 250	< 250	< 300	< 200
λ	< 100	< 220	< 100	< 100
χ^2_{red}	–	< 120	–	–
Events	$S : 89\%$	$S : 98\%$	$S : 97\%$	$S : 70\%$
remaining	$B : 12\%$	$B : 3\%$	$B : 0\%$	$B : 1\%$

Table 7.3: Cuts specific for each energy. In the first column are the parameters and in the next columns the cut values for each energy. The last rows refer to the percentage of remaining events after all the cuts, where S is signal and B is background. Where no useful cut value for a given energy has been found, the entry remains empty.

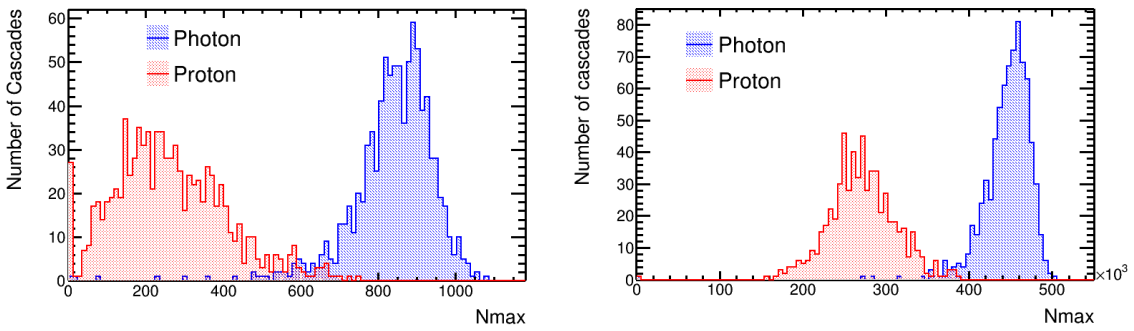

Figure 7.3: Distributions of N_{max} for photons (blue) and protons (red) at 10^2 GeV (left) and 10^5 GeV (right), including the initial cuts.

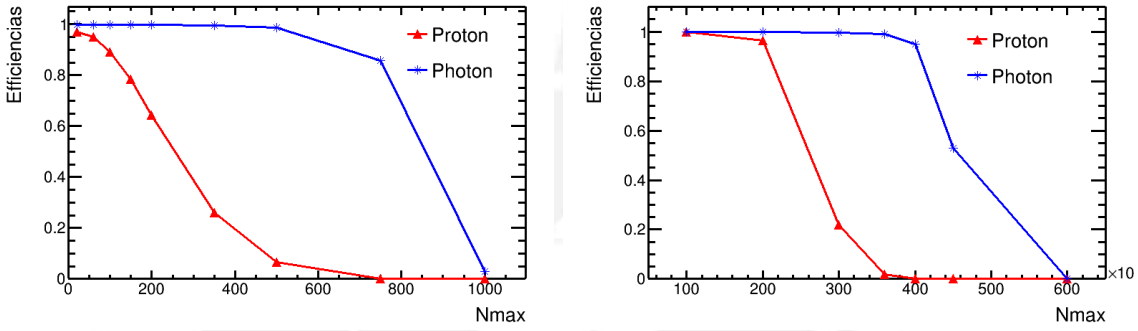

Figure 7.4: Efficiency as a function of N_{max} for photons (blue) and protons (red) at 10^2 GeV (left) and 10^5 GeV (right), including the initial cuts.

TMVA package that will be described in the next Chapter.

It is important to note that we include some fixed cuts for X_{max}, $FWHM$ and X_{start} although we also apply other energy-dependent cuts. The reason is because these cuts are upper and lower limits for the distributions of signal events. This can be seen for example in Fig. 7.6 for the distribution of the X_{max} where the number of signal events is much lower than the number of background events.

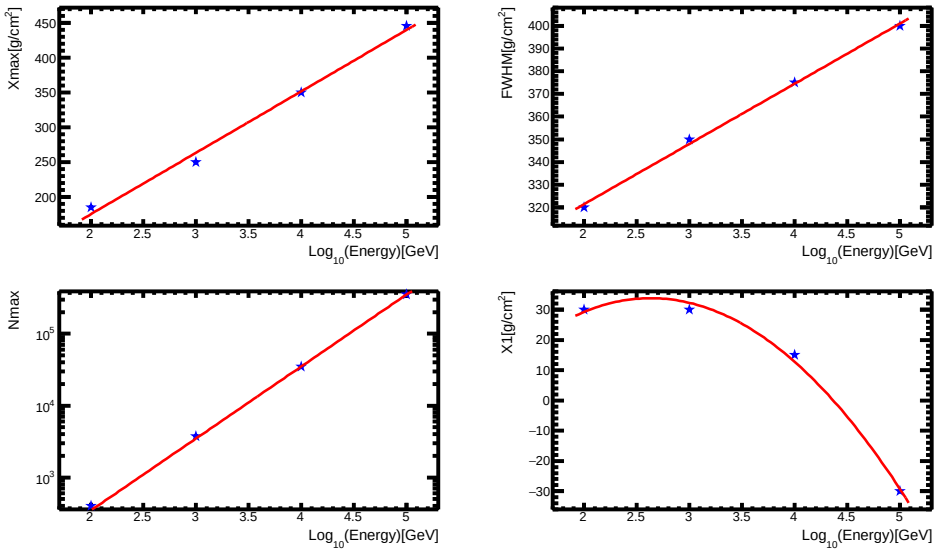

Figure 7.5: Best cuts for some important parameters for four single energies and their fitted functions.

Parameters	Cut1	Cut2
X_{max}	> 220	< 415
$FWHM$	> 197.5	< 362
f	> 0.25	< 0.59
RMS	> 112	< 150
X_{start}	$-$	< 188.2
X_0	> -102	< 74
N_{max}		< 150000
λ	> 33.5	< 61
χ^2_{red}	> 0.63	< 6.8
Energy not smearing	$S: 56\%$	$B: 13\%$
Energy smearing	$S: 54\%$	$B: 12\%$

Table 7.4: Cuts for the sample with realistic spectra. The last row shows the results of applying these cuts, where S is the signal and B is the background.

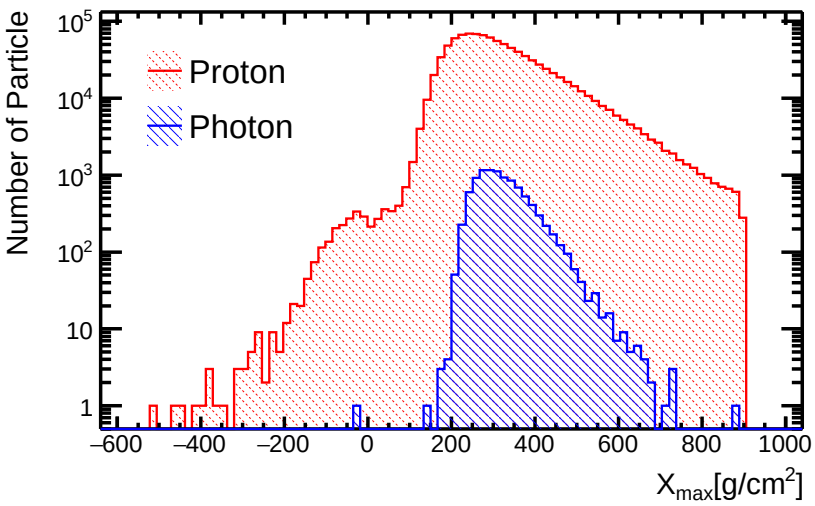

Figure 7.6: Distribution of signal and background events for Xmax.

Chapter 8

Multivariate Analysis - TMVA

This Chapter explains how photons are identified in a background of primaries protons using a multivariate analysis. This statistical technique outperforms the Simple Cuts Method presented in the last chapter 7. For this we use TMVA (Hoecker et al., 2009), the Toolkit for Multivariate Data Analysis, a package from ROOT. TMVA is more complex and effective because it has several methods for signal/background classification. In addition, TMVA has been designed specifically for high-energy physics, but it could have other applications.

TMVA methods are based on machine learning techniques, which means to extract patterns from training data. The analysis procedure can be divided into four steps: the first is the classification between signal and background, the second is to determine the correct method and its parameters, the third is the training and test of the data and the fourth is the evaluation (Hoecker et al., 2009).

The photon/proton separation is performed using several multivariate analysis methods. We will present only the three main methods that give the best results according to TMVA: Linear discriminant analysis, Artificial Neural Networks (ANN) and Boosted Decision Tree (BDT) (Hoecker et al., 2009). We start by explaining briefly the two less important ones for our data in the next sections. Then the BDT is treated with more detail, because this method gives the best results for our data. We also show the BDT response and some important TMVA plots for understanding the results of this toolkit.

When we found the energy dependent cuts in the last Chapter, we worked with the following CORSIKA simulations: 1000 events for signal and 1000 events for background at specific energies 10^2 GeV, 10^3 GeV, 10^4 GeV and 10^5 GeV. The results were applied to a more realistic sample considering different energy spectra and number of events. The latter sample is the one we feed into TMVA. The energies were simulated in a range from 10^2 GeV to 10^5 GeV. We also modified the number of events to 10000 for the signal and 1000000 for the background in order to have enough statistics, even if ideally it should have followed the proportion between the flux of protons and photons presented in Sec. 3.3. The latter would have meant to increase the number of background simulations by a factor of at least ten thousand, which was in practice not feasible. In addition, we have applied a Gaussian smearing mentioned in Sec. 6.5, similar to a resolution or manually introduced detector energy reconstruction error.

8.1 Linear Discriminant Analysis (LDA)

This method classifies the data using a linear model. LDA uses a linear discriminant function in the parameter β. This function has the form:

$$y(x) = \beta x + \beta_0 \tag{8.1}$$

where β_0 is adjusted by $y(x) \geq 0$ for the signal and $y(x) < 0$ for the background (Hoecker et al., 2009). Moreover, this method focuses in maximizing the separability among known categories.

For example, LDA uses the information of the parameters and creates a new linear function and projects the data in this new linear function to maximize the separation of the two categories (signal and background). This new linear function is created according to two criteria: 1) maximize the distance between means (mean of the signal and mean of the background) and 2) minimize the variation within each category. Fig. 8.1 shows the concept of this method.

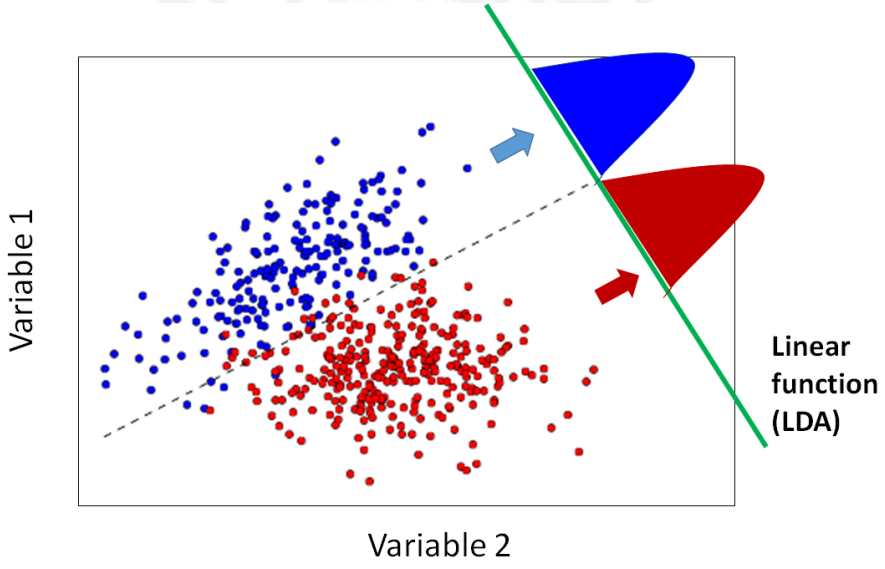

Figure 8.1: Schematic view of the Linear Discriminant Analysis. The green line is the linear function where the data are projected.

8.2 Multi-Layer Perceptron- Artificial Neural Networks (MLP-ANN)

The Multilayer Perceptron (MLP) is an example of an artificial neural network (ANN) that is used for the solution of a number of different problems, including pattern recognition and interpolation (Nieminen, 2005).

An artificial neuron is a computational model which combines its input signals, for example, by a weighted sum. The output is one numerical value, calculated by an activation function, modelling approximately a firing as a biological neuron that pass it signal via synapses. For example, $O = f(b + \sum_{i=1}^{n} w_i(a)_i)$, where O is the output, $(a)_i$ are n inputs, w_i are summation weights, and f is the activation

function; b is a bias term that determines how much activation induces firing. This activation function can be zero, linear or non-linear (Noriega, 2010).

Now, we have many artificial neurons next to each other forming a layer of multiple neurons. Each neuron will have its own weight coefficients, therefore an activation function is operating on each neuron. The input and the output are both numerical vectors. The output will be $(O^l)_j = f_j^l(b_j^l + \sum_{i=1}^{n_j-1} w_{j,i}^l (O^{l1})_i)$, where (O^l) is the output vector, $(O^{l1})_i$ are n_{l1} inputs, $w_{j,i}^l$ is the weight of the $i-th$ input in the $j-th$ neuron, f_j^l is the activation function and b_j^l the bias of the $j-th$ neuron. The indices $l1$ for the input vector and l for the output vector anticipate the next step, where layers are interconnected (Noriega, 2010).

Combining layers we get an MLP. But now, the neurons or input signals will be organized in layers. The first layer in the MLP is called input layer and the last is called output layer. The intermediate layers are called hidden layers (Hoecker et al., 2009). The output of each layer is computed using the same mechanism. The output from layer l is given to layer $l+1$. The formulation can be written in a layer-wise matrix form. Each layer will have a coefficient matrix and a function matrix (Noriega, 2010).

In summary, first this method gets the input variable, then associates an activation function to all variables. Then, it applies a pattern to MLP and calculates the activation of input signals. It then propagates forward the activations, step by step, and finally, reads the network output from all output neurons (Riedmiller, 2005). Fig. 8.2 gives a picture of this method.

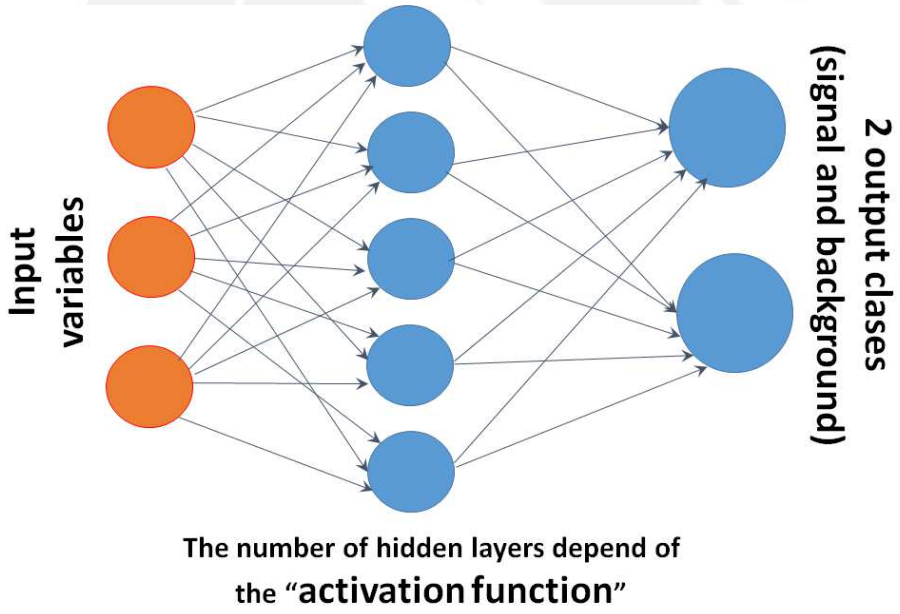

Figure 8.2: Schematic view of a Multi-layer Perceptron.

8.3 Photon/Proton separation with BDT

The Boosted Decision Tree (BDT) method was the most effective for separating signal from background cascades according to the results of TMVA for this

case. A schematic view of a Decision Tree (DT) is shown in Fig. 8.3, with variables x_i and cuts c_a. At the end, one output parameter is obtained, with signal and background separated.

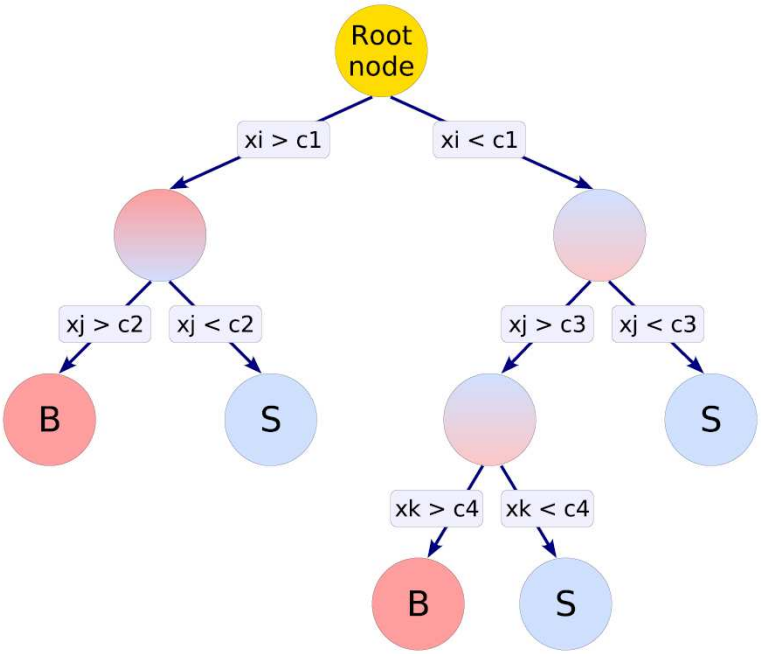

Figure 8.3: Schematic view of a decision tree (Hoecker et al., 2009). At each node there is a decision, where S is signal and B is background.

In the BDT method, the selection criteria is the result of many decision trees from the same training sample weighting differently events during training.

8.3.1 BDT training

The training is the process to define the 'cut criteria' for each node. The training starts with the root node. One takes the full training event sample and selects a variable and cut value that better splits the signal from background. With this cut the sample is split in two sub-samples, a signal-like and a background-like sample forming a tree. Other two nodes are created. The same mechanism is stopped with a minimum number of events or a minimum or maximum signal purity. These final nodes are called 'signal' or 'background' if they contain more signal than background events from the training sample (Hoecker et al., 2009).

8.3.2 BDT boosting

The boosting of a decision (regression) tree extends this concept from one tree to several trees which form a forest. The boosting can be applied several times (100-500 times). The main idea is that signal events from the training sample, that are still in a background node are given a larger weight than events that are in the correct leave node. This results in a re-weighed training event sample with a new decision tree developed. Boosting stabilizes the response of

the decision trees with respect to fluctuations in the training sample and enhances the performance of a single tree (Hoecker et al., 2009).

We used the default boosting algorithm of TMVA: the adaptive boost (AdaBoost). This algorithm gives to misclassified events during the training phase a weight α in the training of the following tree. α is defined in Eq. 8.2, where err is the misclassification rate of the previous tree. Later these weights are normalized so that the sum of all events remains constant (Haffner, 2010).

$$\alpha = \frac{1 - err}{err} \tag{8.2}$$

8.3.3 Training parameters

For a good photon/proton separation we need to train the BDT with several shower parameters. We will use the parameters described Sec. 6. These parameters are:

- the energy E.
- the depth of shower maximum X_{max}.
- the parameter related to depth of the point of first interaction X_0.
- the number of charged particles N_{max}.
- the shower decay length λ.
- the reduced Chi-squared of the fit χ^2_{red}.
- the starting point of the shower X_{start}.
- the root mean square RMS.
- the shower full width at half-maximum $FWHM$.

Fig. 8.4 shows the distributions of the selected variables or parameters, which were extracted or calculated fitting the longitudinal profile. We see that the following parameters: maximum depth, first shower, width, first interaction, maximum number particles and RMS could separate signal from background with some simple cuts. However, other parameters as the smeared energy, the mean free path and the reduced χ^2 are not straightforward to be used. We also note that the maximum number particles is the parameter that could better separate signal from background.

8.3.4 BDT settings

The advantage of the BDT method is that we consider no-linear correlations between input parameters, but the problem is a possible overtraining. Overtraining occurs when a machine learning problem has few degrees of freedom, because too many model parameters of an algorithm were adjusted to few data points (Hoecker et al., 2009). For this reason it is necessary to understand the different BDT settings. However the used values for the BDT are those set by default. In a future work, we could test the results with different values in order to look for a better efficiency.

The number of trees for the forest was set to 850. The minimum node size is 2.5%, which is the minimum percentage of training events that need a leaf node ($http://tmva.sourceforge.net/$). The maximum depth of each tree is 3. A possible overtraining of the tree could be reduced by choosing another maximum depth of the forest. The Gini Index was chosen as splitting criterion. The Gini

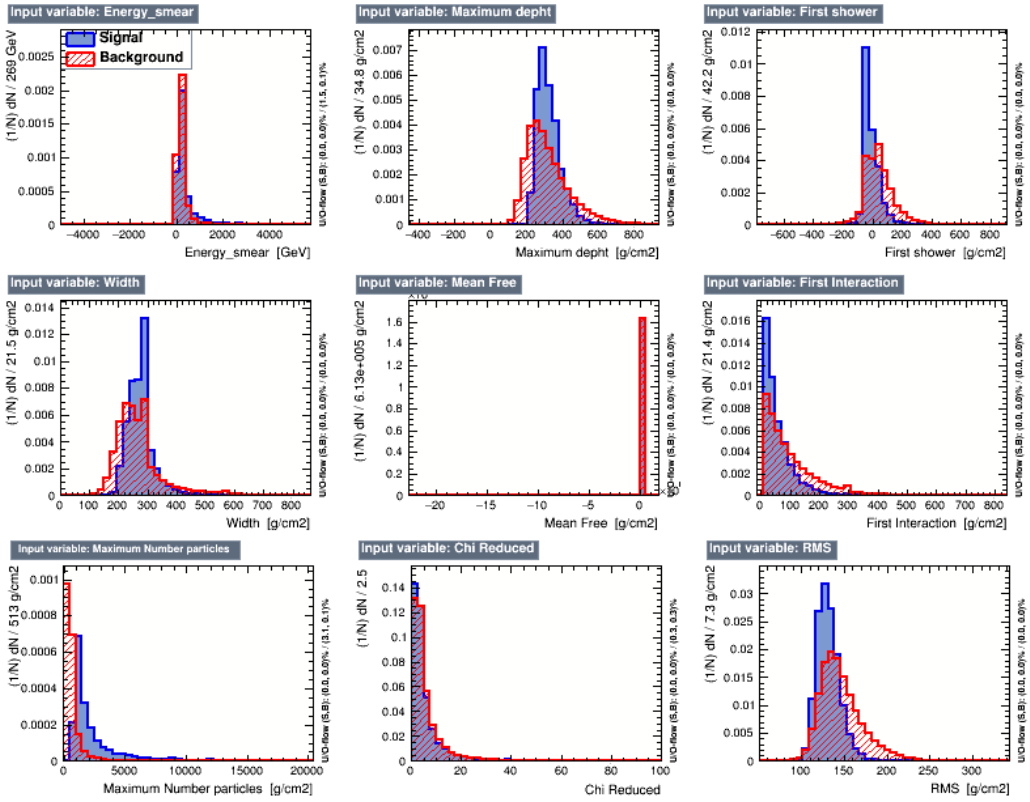

Figure 8.4: Distribution of input parameters for the Boosted Decision Tree training. In blue the distributions of photons as signal and in red the protons as background.

Index is $p(1-p)$, where p is the purity of the signal, which is $p = \frac{S}{(S+B)}$, where S is the number of signal events and B is the number of background events. The number of grid points evaluated in a variable range for finding the optimal cut in each node splitting was 20.

8.3.5 Classification and evaluation

Classification in TMVA is performed from the input variables (observables), where signal or background events have values close to 1 or 0. The mapping to the signal and background class is done defining cuts for signal and for background. For each cut value TMVA provides the signal efficiency, purity and background rejection (Speckmayer et al., 2010).

TMVA provides several evaluation outputs which help to decide which is the best classifier method. There are outputs for each trained method. The output shows signal, background and their efficiencies, which are calculated with cuts as discussed in Chapter 7. Events larger than the cut are signal and events below are background. The plot of TMVA that shows the efficiencies is called ROC. The Reciever Operating Characteristics (ROC) plot shows the sets of signal efficiencies and background rejection defined by the cuts(see Fig.8.5). We can use the ROC plot to decide which is the best method by choosing the one that keeps more signal and less background. This point in the efficiency plot is the cut criteria that better split the signal-like from background-like distributions. The result of

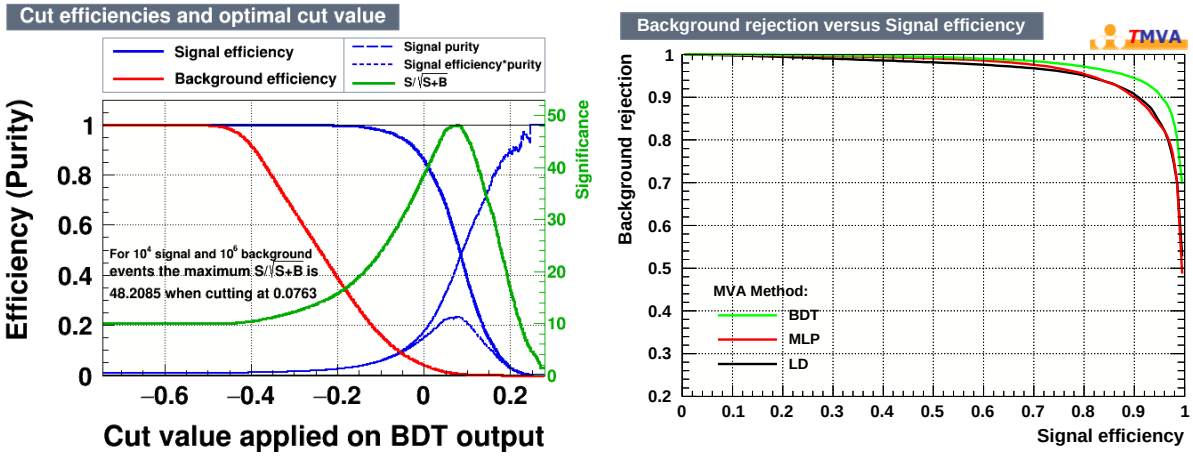

Figure 8.5: (Left) Significance, purity and efficiencies for signal and background, as well as the signal efficiency multiplied by the purity for the BDT output. (Right) ROC curve showing the relationship between signal efficiency and background rejection.

different tests of these cuts are shown in Table 8.1.

The best separation of signal and background can be achieved with the method that has the largest area below the ROC curve (Speckmayer et al., 2010). We can see in Fig.8.5 that the best method is the BDT. On the other hand, if we find signal, then the best cut is where $\frac{S}{\sqrt{(S+B)}}$ is maximum. (Speckmayer et al., 2010).

Finally, TMVA calculates significances, purities and efficiencies and proposes an optimal cut value to get the best performance (Speckmayer et al., 2010). In Fig. 8.5 we show the output of the tool with those values as a function of the cut for the BDT method.

8.3.6 BDT response and performance

The BDT response (output) assigns a numerical value to signal and background events. The BDT training was applied to signal (photon) and background (proton) events for energies between 10^2 GeV and 10^5 GeV. The initial sample was divided in two separate samples: 50% for test and 50% for training. Fig. 8.6 shows the BDT output distributions of the training and test samples superimposed, in order to check the BDT classification for overtraining. When overtraining occurs, the BDT response of training and test samples differ strongly. In this case, the training and test distributions match very well, so there is no overtraining.

In Fig. 8.6 we can see that photon/proton separation is possible. We also see that the best cut is 0.07. Different results from several tests are shown in Table 8.1. We tested the BDT response considering the energy with and without smearing to evaluate if a better energy resolution would improve the efficiency. We conclude that the improvement is almost negligible. We also looked for possible improvements related with different number of tress in the BDT, 400 and 850.

Energy	BDT Cut	Signal	Background	$S/\sqrt{S+B}$
Not smearing(NT=850)	0.07	56.9%	0.7%	7.49
	−0.05	94.5%	8.7%	9.30
Smearing (NT=850)	0.07	56.6%	0.8%	7.47
	0.0095	83.1%	3.4%	8.93
Smearing (NT=400)	0.12	56.8%	1.0%	7.47
	0.032	83.2%	4.3%	8.89

Table 8.1: Cuts for BDT response. This table shows different cuts for energies with and without smearing. The number of trees (NT) is also given.

The results do not vary much in these cases either.

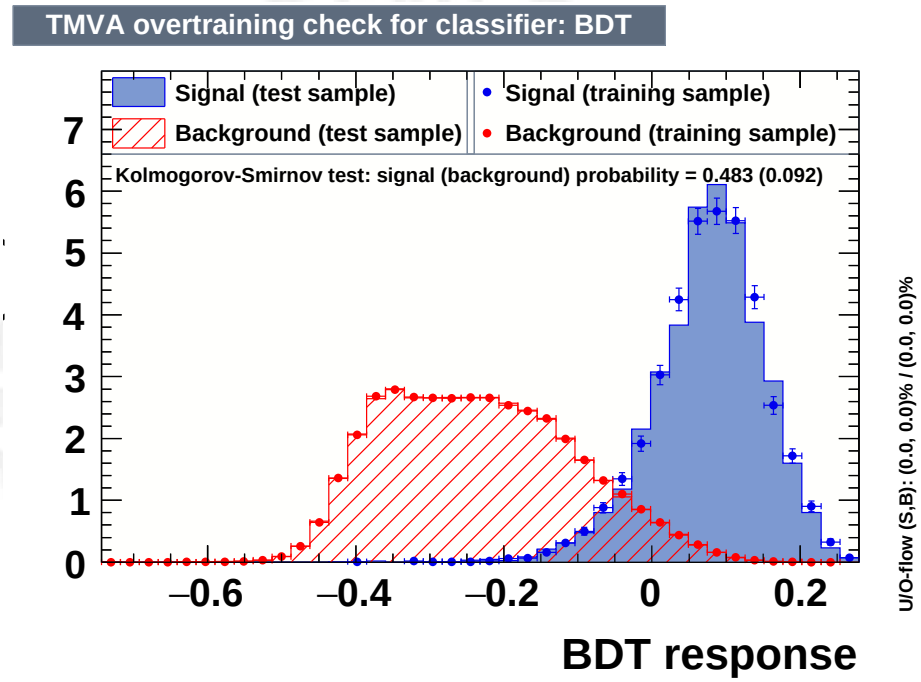

Figure 8.6: Distributions of the Boosted Decision Tree response for an independent test sample for energies between 10^2 GeV to 10^5 GeV.

Chapter 9

Discussion and Conclusions

The main aim of this work was to find a method and the corresponding parameters that could help us to distinguish photons (signal) from protons (background) air-showers. For this purpose, we simulated with CORSIKA the showers initiated by photons and protons in the Earth's atmosphere. A useful way to study them is by looking at their longitudinal profile, i.e. the number of particles in the shower versus the atmospheric depth.

On the other hand, there are detectors that can measure some parameters described in this work. For example fluorescence detectors can measure the X_{max} and Cherenkov tanks can measure the direction of the primary particle and find a pattern that allows to discriminate between gamma-rays from hadrons.

We studied the showers of hadronic and electromagnetic origin, as well as the cosmic-ray and gamma-ray point-source fluxes. From this we observe that the flux of protons is much higher than the gamma-ray flux. For instance, at 10^{13} eV the relation between the fluxes was $\Phi_\gamma/\Phi_p = 10^{-6}$ approximately. This number will be used to compare the expected number of photon and proton events.

The showers that we studied were simulated with CORSIKA. We chose a proton (hadron) and a photon (gamma) as primary particles in the energy range from 10^{12} to 10^{15} eV at the same zenith angle (vertical). After simulating the showers, we have extracted some features (parameters) from fitting the longitudinal development. The Gaisser-Hillas function was the best to fit the longitudinal profiles of both showers. From the fit we can extract the shower parameters that are then used to differentiate photon from proton showers.

The obtained results led to the conclusion that there are differences between showers produced by protons and photons. Furthermore, the maximum atmospheric depth and the maximum number of particles are important variables for distinguishing the showers.

We first applied two general cuts on X_{max} and χ^2_{red}. Then we applied the simple cuts method which was very useful to learn the contribution of the different parameters in the multivariate analysis. With this simple method we could also find adequate cuts for our realistic sample and distinguish signal from background, although with little efficiency.

To distinguish better the signal from the background we used the TMVA toolkit which is very efficient in combining several variables, finding cuts and distinguishing signal from background. Using this tool we found that the BDT method was the best to solve the classification problem. Our best result was

$54.4 \pm 0.01\%$ of the signal and $0.7 \pm 0.0001\%$ of background. We conclude that it is a much better method comparing it to the simple method where we obtain $54.1 \pm 0.007\%$ signal and $12.5 \pm 0.0003\%$ background.

The sample training achieved a good classification of photons from the background. However, the expected photon flux from a point-source compared to the observed proton background is much lower. If we cut in the BDT response trained with 850 trees at 0.2 in the sample including energy smearing, the signal that remains is $3 \pm 0.002\%$and the background is $0.003 \pm 0.000008\%$.

The cut could be tighter, thus improving this results, if the simulated event statistics would be higher. We could apply cuts to completely eliminate the background, but it would not be possible to estimate the error, since the sample is not large enough. If we compare the real flux of photons and protons with this efficiencies, we obtain that for every 1000 protons events we would find 1 photon event.

Considering a possible improvement in the detector energy estimation, we applied the same cut including the energy without smearing. The results improve slightly.

Given the initial flux proportion, the reached background rejection capability is 10^3. Thus the feasibility of gamma/hadron separation requires further improvement. These results could be improved by studying other variables that had been neglected in this work, like the space an temporal profiles. Moreover, we could apply other parameters of BDT as parameter transformations, different depths of the trees, etc. Furthermore, the analysis could be applied to a specific detector.

Bibliography

J.A. Abeysekara, A.U. Aguilar, S. Aguilar, and et. al. for HAWC Collaboration. On the Sensitivity of the HAWC Observatory to Gamma-Ray Bursts. *arXiv:1108.6034v2 [astro-ph.HE]*, 35:641–650, 2011.

T. Abu-Zayyad, K. Belov, and D.J. for the HiRes Collaboration Bird. A measurement of the Average Longitudinal Development Profile of Cosmic Ray Air Showers between 10^{17} eV and 10^{18} eV. *Astroparticle Physical. astro-ph/0008206*, 16:1–11, 2001.

M. Aglietta, B. Alessandro, P. Antonioli, and et al. for EAS-TOP Collaboration. Search for Gamma-Ray Burst at Photon Energies $e \geq 10$ GeV and $e \geq 80$ TeV. *Astrophysical Journal*, 469:305–310, 1996.

F. Aharonian and et. al. (H.E.S.S. Colaboration). Evidence for VHE Gamma-Ray Emission from the Distant BL Lac PG 1553+113. *Astronomy and Astrophysics*, 448:19–23, 2006.

A. A. Al-Rubaiee and A. Jumaah. Investigating of Longitudinal Development Parameters Through Air Shower Simulation by Different Hadronic Models. *IJANS. arXiv:1309.2934 [physics.gen-ph]*, 2:135–140, 2013.

M. Alania, I.J. Araya, A.V.C. Gmez, and et. al. Air Shower Simulation. *AIP Conference Proceedings*, 1123(1):309–319, 2009.

D. Allard, E. Parizot, X. Bertou, J. Beatty, and et al. for the Pierre Auger collaboration. Detecting Gamma-Ray Bursts with the Pierre Auger Observatory using the Single Particle Technique. *29th International Cosmic Ray Conference Pune, India. arXiv:astro-ph/0508441*, 2005.

I.V. Arkhangelskaja. The Properties of the Gamma-Ray Bursts with High-Energy Spectral Component. *Science Direct*, 74:19–23, 2015.

H. Asorey. Los Detectores Cherenkov del Observatorio Pierre Auger y su Aplicacion al Estudio de Fondos de Radiacion. *Thesis from Instituto Balseiro. Universidad Nacional de Cuyo. Comisin Nacional de Energa Atmica Argentina*, 2012.

Pierre Auger-Collaboration. The Pierre Auger Cosmic Ray Observatory. *arXiv:1502.01323 [astro-ph.IM]*, 2015.

M. Bass. Interfacing CORSIKA Air Shower Simulations with LArSoft. *LArSoft Coordination Meeting*, 2015.

D. R. Bergman and J. W. Belz. Cosmic Rays: The Second Knee and Beyond. *arXiv:0704.3721 [astro-ph]*, 2007.

J. Beringer, H Bichsel, and et. al. Passage of Particles through Matter. *Particle Data Group*, PR D86:010001, 2012.

Physics Dictionary. A Dictionary of Physics. *Oxford University Press*, 2009.

L. I. Dorman. Cosmic Rays and Space Weather: Effects on Global Climate Change. *Annales Geophysicae*, 30:9–19, 2012.

T. K. Gaisser. The Cosmic-Ray Spectrum: from the Knee to the Ankle. *Journal of Physics: Conference Series 47*, 2006.

T.K. Gaisser and A.M. Hillas. Reliability of the Method of Constant Intensity Cuts for Recontructing the Average Development of Vertical Shower. *15th International Cosmic Ray Conference*, 1977.

R. C. Gilmore, P. Madau, J. R. Primack, and et al. GeV Gamma-Ray Attenuation and the High-Redshift UV Background. *arXiv:0905.1144 [astro-ph.CO]*, 2009a.

R.C. Gilmore, F. Prada, and J. Primack. Modeling Gamma-Ray Burst Observations by Fermi and MAGIC Including Attenuation due to Diffuse Background Light. *Astronomy Society.arXiv:0908.2830v1 [astro-ph.CO]*, 000:1–13, 2009b.

C. Grupen. *Astroparticle Physics*. Springer Berlin Heidelberg, New York, 2005. ISBN 978-3-540-25312-9.

The ROOT Users Guide. A ROOT Guide for Beginners. Diving Into ROOT. *https://root.cern.ch*, 2015.

HAWC Collaboration H. Salazar. The HAWC Observatory and its Synergies at Sierra Negra Volcano. *31th International Cosmic Ray Conference*, 2009.

S. Haffner. Studies of Cosmic-Ray Electrons with the H.E.S.S. Experiment Using an Advanced Reconstruction Technique. *Erlangen Centre for Astroparticle Physics. Physikalisches Institut*, 2010.

D. Heck, J. Knapp, J. N. Capdevielle, G. Schatz, and T. Thouw. CORSIKA: A Monte Carlo Code to Simulate Extensive Air Showers. *FZKA*, 6019, 1998.

W. Heitler. *The Quantum Theory of Radiation*. Oxford University Press, Oxford, 1954. ISBN 0-486-64558-4.

HESS-Collaboration, F. Aharonian, and et al. Detection of TeV Gamma-Ray Emission from the Shell-Type Supernova Remnant RX j0852.04622 with H.E.S.S. *Astronomy and Astrophysics. arXiv:astro-ph/0505380*, 2005.

HESS-Collaboration, A. Abramowski, F. Aharonian, F. Ait Benkhali, and et al. Search for TeV Gamma-Ray Emission from GRB 100621A, an Extremely Bright GRB in X-Rays, with H.E.S.S. *Astronomy and Astrophysics. arXiv:1405.0488v1 [astro-ph.HE]*, 2014.

A. Hoecker, P. Speckmayer, J. Stelzer, and et. al. TMVA - Toolkit for Multivariate Data Analysis. *arXiv:physics/0703039 [physics.data-an]*, 2009.

H. Hun. Status of the EAS Studies of Cosmic Rays with Energy Below 10^{16} eV. *arXiv:0911.3034 [astro-ph.HE]*, 2009.

H. Krawczynski and E. Treister. Active Galactic Nuclei the Physics of Individual Sources and the Cosmic History of Formation and Evolution. *Cosmology and Nongalactic Astrophysics. arXiv:1301.4179v1 [astro-ph.CO]*, 2013.

R Landi, A. De Rosa, A.J. Dean, and et. al. HESS J1616508: Likely Powered by PSR J16175055. *arXiv:0707.0832v1 [astro-ph]*, 2007.

A. Letessier-Selvon and T. Stanev. Ultrahigh Energy Cosmic Rays. *Rev.Mod.Phys.*, 83:907, 2011.

M. S. Longair. *High Energy Astrophysics*. Third Edition. Cambridge University Press, Cambridge, 2011. ISBN 978-0-521-75618-1.

J.A.J. Matthews, R. Mesler, B.R. Becker, and et al. A Parametrization of Cosmic Ray Shower Profiles Based on Shower Width. *arXiv:0909.4014v2*, 2009.

G. Matthiae. New Results from the Auger Observatory. *arXiv:0807.1024 [astro-ph]*, 2008.

D.G. Melo. *El Detector de Fluorescencia del Observatorio Pierre Auger: Reconstruccin de lluvias de partculas, anlisis de los primeros datos y extensin hbrida del detector de fluorescencia a energas at* 10^{18} *eV*. Thesis from Universidad Nacional General San Martin. Comision Nacional de Energia Atomica. Insituto de Tecnologia, Argentina, 2007.

M. Merck, D.L. Bertsch, B.L. Dingus, and et. al. Observations of High-Energy Gamma-Ray Bursts with EGRET. *Cambridge Core*, 51:358–362, 2016.

L. Nava, G. Ghirlanda, G. Ghisellini, and A. Celotti. Spectral Poperties of 438 GRBs Detected by Fermi/GBM. *Astronomy and Astrophysics. arXiv:1012.2863 [astro-ph.HE]*, 2011.

P. Nieminen. Multilayer Perceptron Tutorial. *Staffordshire University. Beaconside Staffordshire ST18 0DG*, 2005.

L. Noriega. Classification and Multilayer Perceptron Neural Networks. *Lecture in Data Mining Course (TIES445)*, 2010.

K.A. Olive, K. Agashe, C. Amsler, and et. al. Particle Data Group. *Review of Particle Physics. Chinese Physics C.*, 38, 2014.

L. Otiniano, S. Vargas, F. Quispe, and W. Guevara for LAGO Collaboration. Development of a High Altitude LAGO Site in Peru. *34th International Cosmic Ray Conference*, 688, 2015.

M. S. Potgieter. Solar Modulation of Cosmic Rays. *Living Rev. Sol. Phys.*, 2013.

R.R. Prado and V. De Souza. A Study on the Fitting Functions of Energy Deposit Profile of Cosmic Rays Showers. *33RD International Cosmic Rays conference*, 2013.

C.L. Pryke. A Comparative Study of the Depth of Maximum of Simulated Air Shower Longitudinal Profiles. *Astroparticle Physical. arXiv:astro-ph/0003442*, 14:319–328, 2013.

M.A. Rahman, P.N. Bhat, B.S. Acharya, and et al. Gamma Ray and Hadron Generated Cherenkov Photon Spectra at Various Observation Altitudes. *arXiv:astro-ph/0104143*, 2001.

M. Riedmiller. Machine Learning: Multi Layer Perceptrons. *Presentation from Albert-Ludwigs-University Freiburg. AG Maschinelles Lernen*, 2005.

J. Ryan, K. Bennett, W. Collmar, and et. al. COMPTON Measurements of the Gamma-Ray Burst GRB 930131. *Astrophysical Journal*, 442:67–70, 1993.

P. Speckmayer, A. Hocker, J. Stelzer, H. Voss, and et. al. The Toolkit for Multivariate Data Analysis, TMVA 4. *17th International Conference on Computing in High Energy and Nuclear Physics. Journal of Physics*, 2010.

B. C. Thomas, A. L. Melott, C. H. Jackman, and et. al. Gamma-Ray Burst and the Earth: Exploration of Atmospheric, Biological, Climatic and Biogeochemical Effects. *The Astrophysical Journal*, 634:509–533, 2005.

S. Vernetto. Detection of Gamma-Ray Bursts in the 1 GeV-1 TeV Energy Range by Ground Based Experiments. *arXiv:astro-ph/9904324v2*, 2005.

K. F. Weidenhaupt. *Antenna Calibration and Energy Measurement of Ultra-High Energy Cosmic Rays with the Auger Engineering Radio Array.* Publikationsserver der RWTH Aachen University, 2014.

P. W. Younk. Cosmic Rays at the Ankle: Composition Studies Using the Pierre Auger Observatory. *Dissertation from Michigan Technological University*, 2007.

List of Figures

List of Tables

Appendix A

Corsika Installation

This little CORSIKA installation manual is based on our experience.

A.1 Obtaining the CORSIKA software

The first step to access the repository is to send an e-mail to tanguy.pierog@kit.edu or dieter.heck@partner.kit.edu requesting access to the program, mentioning the purpose of the experiment and where the simulations will be used (University or laboratory). After sending the message, your e-mail is added to the principal list of CORSIKA and you will receive the user and password for downloading and unpacking the code: $corsika - 74005.tar.gz$ from: $https://www.ikp.kit.edu$.

Moreover, we can the download it using $lftp$ from a Linux terminal using these commands:

$\$lftp$
$\$lftp :\sim> set\ ssl : verify - certificate\ yes$
$\$lftp :\sim> set\ ssl - allow\ yes$
$\$lftp :\sim> open\ corsika@ikp - ftp.ikp.kit.edu$
$\$password$: (obtained from e-mail)
$\$lftp : corsika@ikp - ftp.ikp.kit.edu :\sim> dir$

Later, change to the subdirectory $corsika - v740$ with the command:
$\$lftp corsika@ikp - ftp.ikp.kit.edu : / >$ cd $corsika - v740$

and transfer the package with the command:
$\$lftp corsika@ikp - ftp.ikp.kit.edu : /corsika - v740 >$ get $\$corsika - 74005.tar.gz$

To exit the $lftp$ use:
$\$lftp corsika@ikp - ftp.ikp.kit.edu : /corsika - v740 >$ quit

This information was collected from
$https://web.ikp.kit.edu/corsika/technicalities.html$ and from the reply e-mail.

A.2 Corsika installation

After obtaining the CORSIKA tarball we have to decompress it using the following command:
$ $tar\ -zxvf corsika-74005.tar.gz$

Later, execute as superuser the coconut script that is inside the folder corsika-74005:

$sudo ./coconut

The program asks some questions about our preferences for the CORSIKA options. This information is related to our simulation.
The first question is about the system architecture, in which we chose 2:

```
Compile in 32 or 64bit mode ?
   1 - Force 32bit mode
   2 - Use compiler default ('-m64' on a 64bit machine) [CACHED]

   r - restart (reset all options to cached values)
   x - exit make

   (only one choice possible):
```

The second question is about the hadronic interaction model at high energies, in which we chose 5, because this model was re-calibrate with latest LHC (Large Hadron Collider) results.

```
Which high energy hadronic interaction model do you want to use ?
   1 - DPMJET 2.55
   2 - EPOS LHC
   3 - NEXUS 3.97
   4 - QGSJET 01C (enlarged commons)
   5 - QGSJETII-04 [CACHED]
   6 - SIBYLL 2.1
   7 - VENUS 4.12

   r - restart (reset all options to cached values)
   x - exit make

   (only one choice possible):
```

The third question is about the hadronic interaction model at low energies in which we chose 1, because this model also was re-calibrate with latest LHC results.

```
Which low energy hadronic interaction model do you want to use ?
   1 - GHEISHA 2002d (double precision) [CACHED]
   2 - FLUKA
   3 - URQMD 1.3cr

   r - restart (reset all options to cached values)
   x - exit make

   (only one choice possible):
```

The fourth question is about detector geometry, in which we choose 2. Although in our case the most adequate had been option 1, because is for detectors

arrays. However we chose 2 because we are interested in the shower development, not its surface. Moreover, with option 1 the files that are obtained would have been much larger and the process to run CORSIKA would have taken longer.

```
Which detector geometry do you have ?
   1 - horizontal flat detector array
   2 - non-flat (volume) detector geometry [CACHED]
   3 - vertical string detector geometry

   r - restart (reset all options to cached values)
   x - exit make

   (only one choice possible):
```

In addition to choosing the options, the script shows some additional options, in which we chose the options 2a, g, 7a, 8a, 9 and 9a. These options were related to the LPM effect, curved atmosphere, for plottin the shower, for the Earth's electric field, external atmosphere and COMPACT particle output. More details for these and other options in Heck et al. (1998)

```
Which additional CORSIKA program options do you need ?
   1a - Cherenkov version for rectangular detector grid
   1b - Cherenkov version for telescope system (using bernlohr IACT C-routines)
   1c - apply atm. absorption, mirror reflectivity & quantum eff.
   1d - Auger Cherenkov longitudinal distribution
   1e - TRAJECTory version to follow motion of source on the sky
   2 - THINning version
   2a - LPM-effect without thinning
   3 - PRESHOWER version for EeV gammas
   4 - NEUTRINO version
   4a - NUPRIM primary neutrino version with HERWIG
   5 - STACK INput of secondaries, no primary particle
   6 - CHARMed particle/tau lepton version with PYTHIA
   6a - TAU LEPton version with PYTHIA
   7 - SLANT depth instead of vertical depth for longi-distribution
   7a - CURVED atmosphere version
   7b - UPWARD particles version
   7c - view-cone version
   8a - shower PLOT version (PLOTSH) (only for single events)
   8b - shower PLOT(C) version (PLOTSH2) (only for single events)
   8c - ANAlysis HISTos & THIN (instead of particle file)
   8d - Auger-histo file & THIN
   8e - MUON-histo file
   9 - external atmosphere functions (table interpolation)
        (using bernlohr C-routines)
   9a - EFIELD version for electrical field in atmosphere
   a - CONEX for high energy MC and cascade equations
   b - PARALLEL treatment of subshowers
   c - CoREAS Radio Simulations
   d1 - Inclined observation plane
   d2 - ROOT particle OUTput file
   e - interaction test version (only for 1st interaction)
   f - Auger-info file instead of dbase file
   g - COMPACT particle output file
   h - MUPROD to write decaying muons
   h2 - prEHISTORY of muons: mother and grandmother
   k - annitest cross-section version (obsolete)
     - ----------------------------------
   y - *** Reset selection ***
   z - *** Finish selection *** [DEFAULT]

   r - restart (reset all options to cached values)
   x - exit make
```

Finally the installation script is executed and the program compiled with the given options.

A.3 Input file

CORSIKA is executed with an input file, which has different options about the shower simulation as detailed in Sec. 5. In Fig. A.1 we show an input file example that we have used in this work.

```
RUNNR   116                        run number
EVTNR   1                        number of first shower event
NSHOW   100                        number of showers to generate
PRMPAR  14                       particle type of prim. particle
ESLOPE  -2.7                     slope of primary energy spectrum
ERANGE  1.E3  1.E3               energy range of primary particle
THETAP  0.  0.                 range of zenith angle (degree)
PHIP    -180.  180.              range of azimuth angle (degree)
SEED    1   0   0                seed for 1. random number sequence
SEED    2   0   0                seed for 2. random number sequence
OBSLEV  1400.E2                   observation level (in cm)
FIXCHI  0.                       starting altitude (g/cm**2)
MAGNET  20.0  42.8               magnetic field centr. Europe
HADFLG  0  0  0  0  0  2         flags hadr.interact.&fragmentation
ECUTS   0.3  0.3  0.003  0.00   energy cuts for particles
MUADDI  T                        additional info for muons
MUMULT  T                        muon multiple scattering angle
ELMFLG  T   T                    em. interaction flags (NKG,EGS)
STEPFC  1.0                      mult. scattering step length fact.
RADNKG  200.E2                   outer radius for NKG lat.dens.distr.
LONGI   T  20.  F  F             longit.distr. & step size & fit & out
ECTMAP  1.E3                     cut on gamma factor for printout
MAXPRT  100                        max. number of printed events
DIRECT  /home/jrengifo/LAGO/     output directory
PLOTSH  T
USER    jrengifo                       user
DEBUG   F  6  F  1000000         debug flag and log.unit for out
EXIT                             terminates input
```

Figure A.1: Input file of CORSIKA.

Appendix B

Condor computer cluster

Condor is a computer cluster, which distributes the jobs in the farm, parallelizing the execution of the scripts and thus reducing the total time of these intensive computing calculations.

We ran CORSIKA with the help of condor. We describe the way to execute it in some steps.

1. We create a script called *script_corsika.sh*, which will execute CORSIKA. This script will use the input file. This script is shown here:

```
#!/bin/sh
export ROOTSYS=/home/jrengifo/root/root
export PATH=$ROOTSYS/bin:$PATH
export LD_LIBRARY_PATH= $ROOTSYS/lib: $LD_LIBRARY_PATH
cd /home/jrengifo/corsika-74005/run/ ./corsika74005Linux_QGSII_gheisha_root
</home/jrengifo/LAGO/input> /home/jrengifo/LAGO/output.txt
```

2. We create a script called *lago_condor.submit*, which will be sent to condor

```
Arguments = $(Process)
input = /dev/null
output = Condor_Logs/lago_job_$(Process).out
error = Condor_Logs/lago_job_$(Process).error
log = Condor_Logs/lago_job_$(Process).log
# Email notification (can be NEVER, ALWAYS, COMPLETE, or ERROR)
Notification = NEVER
```

3. We send this script with the command:
condor_submit lago_condor.submit

Other options were:
condor_rm (number of job): Delete some work.
condor_q: Look at the jobs which are been executed or in queue in condor.