A STUDY OF THE SIGNALS MEASURED WITH THE WATER-CHERENKOV DETECTORS OF THE PIERRE AUGER OBSERVATORY TO INFER THE MASS COMPOSITION OF ULTRA-HIGH ENERGY COSMIC RAYS

Juan Miguel Carceller

Universidad de Granada

July 2020

Departamento de Física Teórica y del Cosmos y CAFPE

Programa de Doctorado en Física y Ciencias del Espacio

Editor: Universidad de Granada. Tesis Doctorales
Autor: Carceller López, Juan Miguel
ISBN: 978-84-1306-655-4
URI: http://hdl.handle.net/10481/63939

Contents

Resumen

Desde hace más de cien años se sabe que hay partículas que llegan a la Tierra desde el espacio. Estas partículas se conocen como rayos cósmicos y su estudio ha sido muy fructífero y contribuido a la física fundamental desde la primera mitad del siglo 19. De estos rayos cósmicos, algunos tienen energías por encima de 1 J y se llaman Rayos Cósmicos de Ultra-Alta Energía. Hay muchas preguntas sobre ellos que todavía no tienen una respuesta, como qué son, de dónde vienen o cómo son acelerados hasta tales energías. En esta tesis pretendemos contestar a la pregunta de qué son, lo que se conoce como estudiar su composición en masa.

Estudiamos Rayos Cósmicos de Ultra-Alta Energía con los datos obtenidos por el Observatorio Pierre Auger. Los detectores del Observatorio miden señales que dejan partículas secundarias producidas en la lluvia o cascada de partículas generada cuando un rayo cósmico de ultra-alta energía colisiona con una molécula de aire en la atmósfera. El Observatorio Pierre Auger es el detector más grande de rayos cósmicos que se ha construido y con sus 3000 km^2 de superficie y alrededor de 15 años de toma de datos ha conseguido reunir el más grande y preciso conjunto de datos para estudiar las propiedades de este intrigante tipo de radiación.

La tesis está dividida en varios bloques. Aunque el objetivo principal es obtener información sobre la composición en masa de los rayos cósmicos, las herramientas para su estudio cambian a medida que la tesis se desarrolla. El primer bloque es una introducción a los rayos cósmicos (Capítulo 1) y al Observatorio Pierre Auger (Capítulo 2). El siguiente bloque está formado por dos estudios del risetime $t_{1/2}$ de las señales en el suelo medidas por el Detector de Superficie del Observatorio Pierre Auger (Capítulos 3 y 4). El tercer y último bloque tiene una introducción a las técnicas de machine learning (Capítulo 5) y la extracción de la señal que dejan los muones en el detector usando redes neuronales (Capítulos 6 y 7). La tesis termina con un capítulo dedicado a comparar datos y simulaciones y describir una de las discrepancias que hay entre ambos (Capítulo 8).

En los Capítulos 3 y 4 obtenemos información sobre la composición en masas con dos análisis usando el risetime $t_{1/2}$ de las señales medidas de un modo original: definiendo un nuevo observable, el promedio del Risetime dividido por la distancia $\overline{\mathrm{ToD}}$. Este observable es estudiado en ambos capítulos; en el Capítulo 3 se estudia su dependencia con la secante del ángulo cenital del suceso y en el Capítulo 4 se estudia el segundo momento de su distribución.

Durante el trabajo para la tesis, se estableció una colaboración con expertos de ciencia de computadores. Esta colaboración tenía como objetivo predecir la señal que dejan los muones en los detectores usando métodos del campo de machine learning, con el objetivo final de mejorar las estimaciones de composición con esta nueva información. En el

Capítulo 5 se hace una introducción a las técnicas de machine learning. En este capítulo se dan varios ejemplos de estas técnicas con resultados concretos y ejemplos hechos para esta tesis. En particular, se explica cómo funciona una red neuronal simple que es implementada desde cero. Esta introducción sirve para entender las técnicas usadas en los Capítulos 6 y 7. El Capítulo 6 describe el trabajo realizado en colaboración con expertos de ciencias de computadores. En este capítulo la señal muónica se predico para cada estación individual del Observatorio. El Capítulo 7 describe el siguiente paso del trabajo: predecir la serie temporal completa de las trazas medidas por el Observatorio.

El Capítulo 7 es el resultado más importante de la tesis. Este capítulo proporciona un método para predecir la señal temporal que dejan los muones en el Detector de Superficie del Observatorio Pierre Auger. Esta información puede mejorar enormemente las capacidades del Observatorio y permitir que se puedan hacer inferencias sobre la composición en masa de los rayos cósmicos suceso a suceso.

El último capítulo, el Capítulo 8, está relacionado con los dos anteriores en los que se utilizan técnicas de machine learning. Como entrenamos las redes neuronales con simulaciones, la calidad del modelo que se obtenie depende de cómo de bien las simulaciones describen los datos. Como veremos, hay algunas discrepancias entre datos y simulaciones y es necesario entenderlas para poder aplicar los métodos de machine learning a los datos correctamente.

Summary

Since more than a hundred years ago, it has been known that there are particles that arrive to the Earth from outer space. These particles are called cosmic rays and their study was very fruitful and has contributed to fundamental physics since the first half of the 19th century. From those particles, there are some that have energies above 1 J, and they are called Ultra High Energy Cosmic Rays. There are many questions about them that have not been answered yet, such as what they are, where they come from or what is the mechanism that gives them those huge energies. In this thesis we focus on answering what they are, that is, inferring their mass composition.

We study UHECRs with the data obtained by the Pierre Auger Observatory. The detectors of the Observatory measure signals left by secondary particles produced in the shower or cascade of particles generated when a UHECR collides with a molecule of air in the atmosphere. The Pierre Auger Observatory is the largest cosmic ray detector built so far, and with its 3000 km^2 of surface area and about fifteen years of data taking it has collected the biggest and most precise data sample to study the properties of this intriguing radiation.

The work in this thesis is divided into several blocks. While the main goal is to obtain information about mass composition, the tools used to do so change as the thesis develops. The first block is an introduction to cosmic rays (Chapter 1) and the Pierre Auger Observatory (Chapter 2). The next block is comprised by a study of the risetime $t_{1/2}$ of the signals at the ground measured by the Surface Detector of the Pierre Auger Observatory (Chapters 3 and 4). The third and last block has an introduction to machine learning techniques (Chapter 5) and the extraction of the signal left by muons in the detector using neural networks (Chapters 6 and 7). The thesis ends with a chapter dedicated to compare data and simulations and highlight one of the discrepancies between them (Chapter 8).

In Chapters 3 and 4 we infer information about mass composition with two analyses using the risetime $t_{1/2}$ of the signals measured in a novel way: by defining a new observable, called the average Risetime over Distance $\overline{\mathrm{ToD}}$. This observable is studied in both chapters; in Chapter 3 we study its dependence with the secant of the zenith angle of the event and in Chapter 4 the second moment of its distribution.

As the thesis evolved, a collaboration with experts from computer science was established. This collaboration had the goal of predicting the signal left by muons in the detectors using novel methods from the field of machine learning, with the objective of improving composition estimations using this new information. An extensive introduction to the techniques of machine learning is done in Chapter 5. Several examples of these techniques are explained with concrete results and examples done for this thesis. In particular, it is explained how a simple neural network works and it is programmed from scratch. This

introduction lays the foundations for Chapters 6 and 7. Chapter 6 describes the work done within the collaboration with experts from computer science. In this chapter, the muon signal is predicted for each individual station of the Observatory. Chapter 7 describes the next step of the work: to predict the whole temporal series of the traces measured by the Observatory.

Chapter 7 is the most important result of the thesis. It provides a method to predict the signal left by muons in the Surface Detector of the Pierre Auger Observatory. This information can enhance greatly the capabilities of the Observatory and allow it to make mass composition inferences on an event by event basis.

The last chapter, Chapter 8, is related to the previous ones that use machine learning techniques. Since we train our neural networks with simulations, the quality of the model obtained will depend on the quality with which simulations describe the data. As we will see, there are some discrepancies that have to be taken into account when applying methods from machine learning to data.

Summary of Variables

This is a short summary for some of the variables that appear in the thesis.

One value for each event or shower

- E_{SD} is the reconstructed energy of the cosmic ray. It is obtained by a fit of the total signal S measured at each station as a function of its distance to the reconstructed core of the shower r: $S(r)$. The value at $r = 1000\,\mathrm{m}$, S_{1000}, is picked and converted to E. See SD event reconstruction starting on page 27.
- E_{MC} is the Monte Carlo energy, that is, the energy of the primary cosmic ray that begins the shower of particles.
- θ is the reconstructed zenith angle of the arrival direction of the cosmic ray. It is obtained by fitting a spherical wave form to the times when stations measure signal from the shower. See SD event reconstruction starting on page 27.
- X_{max} is the depth (measured in $\mathrm{g\,cm^{-2}}$) at which the maximum energy deposited is reached in a shower of particles. It is measured by the fluorescence telescopes.
- $\langle \Delta_s \rangle$ is the Delta, an observable built from the risetime that has information about its deviation from a defined benchmark. It is obtained by an average. See Summary of the Delta Method on page 43.

One value for each station

- r is the distance of the station to the reconstructed position of the core of the shower in the plane of the shower.
- S is the total signal measured at each station, obtained by integrating or summing the signal measured over time.
- $t_{1/2}$ is the risetime, the time it takes for the total signal to rise between 10% and 50% of the total signal. $t^{\mu}_{1/2}$ has the same definition using the muon signal instead of the total signal. See Introduction of Chapter 3 starting on page 38.
- ζ is the polar angle of the station in the plane perpendicular to the shower. See Polar angle correction on page page 40.
- S^{μ} and S^{EM} are the total muon signal and total electromagnetic signal, respectively, obtained by integrating the muon signal or electromagnetic signal over time.
- $\widehat{S^{\mu}}$ and $\widehat{t^{\mu}_{1/2}}$ have the same definition than before but are obtained from the predicted muon trace. $\widehat{S^{\mu}}$ is also obtained directly from a neural network in Chapter 6.

One value for each event or shower

- $\overline{\mathrm{ToD}}$ is the average risetime divided by the distance, defined in Equation 3.12 on page 45.
- $\sigma^2_{\mathrm{total}}$ is the variance of an observable, σ^2_{det} is the contributions to the variance due to the detector and σ^2_{f} the contribution due to
- S^{μ}_{1000} and S^{EM}_{1000} are the values at $r = 1000\,\mathrm{m}$ of a fit of the muon signal and electromagnetic signal, respectively, as a function of the distance r.

One value for each station

- $\widehat{S^{EM}}$ is obtained from the predicted electromagnetic signal, which is obtained by subtracting the predicted muon signal to the total signal.

1

Cosmic Rays

Cosmic rays are particles that travel through space and come from outside the Earth. Since their discovery, they have sparkled a lot of interest. There have been and are many experiments dedicated to the study of cosmic rays both in the Earth and in space. In addition, the study of cosmic rays has been very successful since its beginning. It played a crucial role in the discovery of many particles during the 30s and 40s, making important contributions to fundamental physics [1].

Even though cosmic rays have been studied for more than a century, there are still many questions unanswered about them. In particular, it is known that some cosmic rays have energies which can be considered macroscopic as they are above 10^{18} eV (0.16 J), orders of magnitude above what can be produced by human-made accelerators. These cosmic rays are called Ultra-High Energy Cosmic Rays (UHECRs). It is still not clear what is their origin, what are the mechanisms that accelerate these UHECRs up to those extraordinary energies or what is their chemical composition.

This chapter begins with a historical overview of the discovery of cosmic rays in Section 1. Then, the physical phenomenon that is studied in this thesis, cosmic ray showers, is described in Section 2. The chapter ends with features and physics results about cosmic rays in Section 3 that help to set a context of this work in the field of cosmic rays in Section 4.

1 History of cosmic rays

The history of cosmic rays begins with the discovery of radiation. In 1895 X-rays were discovered by Röntgen by studying the light emitted by cathode ray tubes under high voltage [2]. In 1896 Becquerel discovered spontaneous radioactivity [3]. He was interested in phosphorescence and was studying the radiation of uranium crystals after being exposed to the sunlight. The discovery of spontaneous radioactivity was a coincidence: Becquerel did not have time to obtain results for a conference to be held the next week because it had been cloudy. But he developed the photographic plates anyway where it could be seen that a crystal had generated radiation without an external source of energy. A few years later, Marie and Pierre Curie discovered that polonium and radium also generated radioactivity [4]. It was, then, found that an electroscope discharges when in presence of a radioactive material. Note that by 1785 Coulomb had already discovered that electroscopes can discharge spontaneously in presence of air and with a good insulation [5].

The electroscope is a device that can detect whether a body has charge or not. The device evolved with the years but its working principle remained the same: detecting charge by means of the attraction or repulsion generated by the Coulomb force, see Figure 1.1.

After the discovery that an electroscope can be discharged spontaneously, it would be concluded that there is charged radiation in the air. A new research line opened: to find where this radiation came from. There were two possibilities: either the radiation came from the Earth or it came from outside. Wilson made the suggestion that the radiation could come from outside of the Earth. He studied the rate of ionization underground, in tunnels below rocks and could not find a decrease that would support his hypothesis [6].

The first one to test the variation of the ionization with height was Father Wulf, a Jesuit priest. In 1909 he went to the top of the Eiffel Tower in Paris [7]. If the radiation came from the ground it would be more absorbed at larger heights, so it was expected to measure a much lower rate of ionization that on the surface. He measured a decrease in the rate of ionization but this decrease was not enough to agree with what the calculations at the time had predicted assuming radiation came from the ground. Even so, it was still thought that the radiation came from the Earth.

Domenico Pacini made several important experiments to measure the rate of ionization in different environments [8–10]. He compared the rate of ionization when an electroscope was located at the ground and at sea on a ship. If cosmic rays were coming from the soil, water should absorb them and the rate of ionization should be lower on the ship. However, he measured similar rates at ground and on a ship, contradicting the hypothesis of the terrestrial origin. He also measured the rate of ionization under the sea, and found a decrease compatible with the expected absorption in water if the radiation was coming from above.

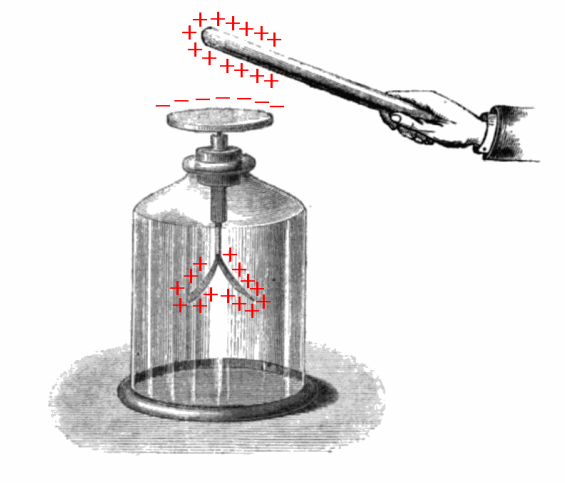

Figure 1.1: Drawing of a gold-leaf electroscope. When a charged object is close to the disk at the top of the electroscope, the disk becomes charged with opposite sign charge. There are two parallel strips that are connected to the disk by a metal bar and become charged. They repel each other and that is how charge in the object can be measured.

With contradictory results about the origin of the radiation it would be clear soon that balloon flights were needed. Alfred Gockel was the first one: he ascended up to 4500 m during three successive flights and found that ionization did not decrease with height as expected from a terrestrial origin [11,12].

In spite of Pacini's conclusions and of Wulf's and Gockel's puzzling results on the dependence of radioactivity on altitude, physicists were reluctant to give up the hypothesis of a terrestrial origin. The situation was cleared up thanks to a long series of balloon flights by the Austrian physicist Victor Hess, who established the extra-terrestrial origin of at least part of the radiation causing the observed ionization.

In 1912 Victor Hess made a series of 7 balloon flights [13,14]. He carried devices in the globe that allowed him to measure the rate of ionization as a function of the altitude. He found that the rate of ionization decreased as the altitude increases for low altitudes, as it could be expected if the radiation came from Earth. Then, above a certain altitude, the rate of ionization increased exponentially with height. Hess concluded from this that the radiation should come from outside of the Earth. The Sun was ruled out as a possible source because he also measured the rate of ionization during the day and night and there was not a significant difference. These results were later confirmed by Kolhörster [15,16] with more balloon flights reaching altitudes up to 9200 m, see Figure 1.2. See ref. [14] for a more detailed history of the discovery of cosmic rays. Hess was awarded the Nobel Prize in 1936 for the discovery of cosmic rays. The Nobel Prize was shared with Carl Anderson for the discovery of the positron while studying cosmic rays in cloud chambers.

After a pause due to the First World War, research in cosmic rays continued, helped by

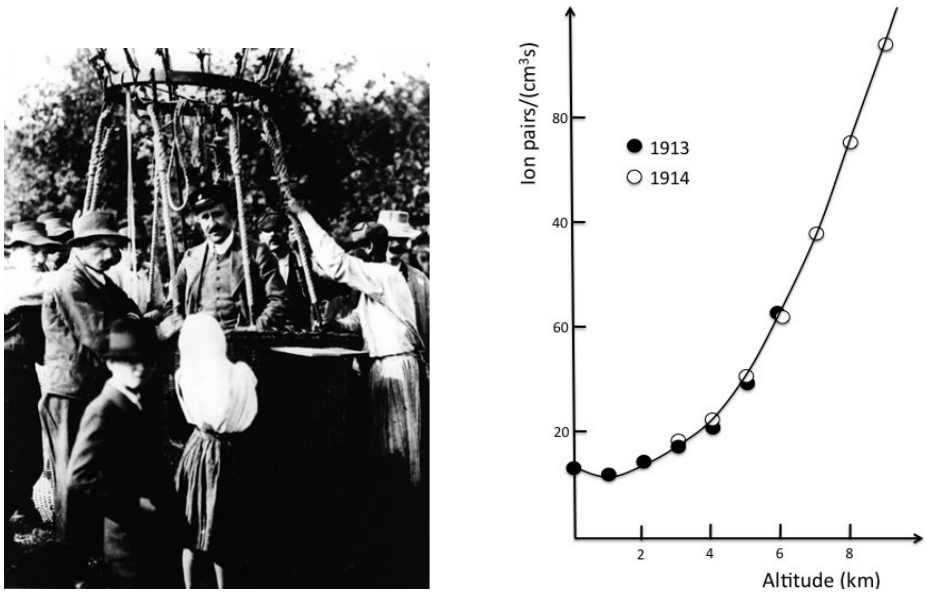

Figure 1.2: Left: Victor Hess before one of his balloon flights. Right: Ionization rate measured by Hess (1913) and Kolhörster (1914) as a function of altitude in balloon flights.

the development of the detectors and, in particular, the introduction of the Geiger-Müller tube. Soon after, the coincidence technique would be introduced. This technique allows to measure only when signals from several detectors arrive within a defined window of time. Photographs of clouds chambers could be taken when a cosmic ray passed through, making their study easier.

Experiments from the field of cosmic rays made very important contributions to fundamental physics. For 30 years, and until particle accelerators were being made, cosmic rays experiments were being used to discover new particles. Some of the particles discovered in this time were the positron, the muon, the pion and the kaon, along with their properties such as mass, charge and lifetime. Another finding was the important discovery that cosmic rays can produce showers or cascades of particles.

2 Cosmic ray showers

The discovery of cosmic ray showers begins with the work done by Rossi. In 1933 Rossi found triple coincidences in an arrangement of Geiger counters when lead was placed above them [17]. He concluded that secondary particles where being produced when a cosmic ray collided with lead. However, the work done by Rossi was not noticed by other scientists, possibly because it was written in italian. The discovery of extensive air showers

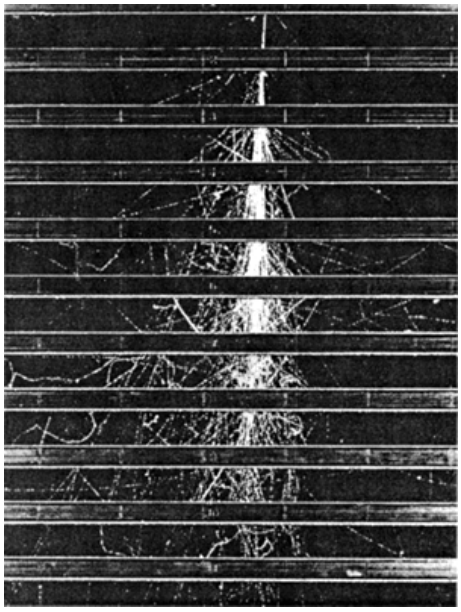

Figure 1.3: Image of a shower or cascade or particles inside a cloud chamber. There are several horizontal lead plates inside the chamber. A cosmic ray enters the cloud chamber from the top and the first interaction appears to have taken place in one the plates.

is usually attributed to Auger. Auger, Maze and Robley measured coincidences in the Swiss Alps separating the detectors by 300 m [18]. Based on the number of particles and assuming that each particle carried the critical energy (defined in the next section), they estimated for the first time that the energy of the cosmic rays was 10^{15} eV.

A shower occurs when a cosmic ray interacts with an atom or molecule (of air in the atmosphere, for example) and new particles are produced. These particles can collide again or decay producing more particles for the third generation of this iterative process. The number of particles present in the shower at some time can reach 10^{10} or even 10^{11} for ultra-high energy cosmic rays. The footprint at the ground can extend over large areas of several km^2, giving these showers the name of Extensive Air Showers (EAS). Figure 1.3 is an example of a cascade or shower of particles produced by a cosmic ray in a cloud chamber. Up to scale, the shower presents the same features as those produced in the atmosphere.

To study a shower it is usual to divide it in several components depending on the process in which particles were produced, see Figure 1.4. In the following sections it is explained, in a simplified way, these components and how they develop. For a historical review on extensive air showers see ref. [19].

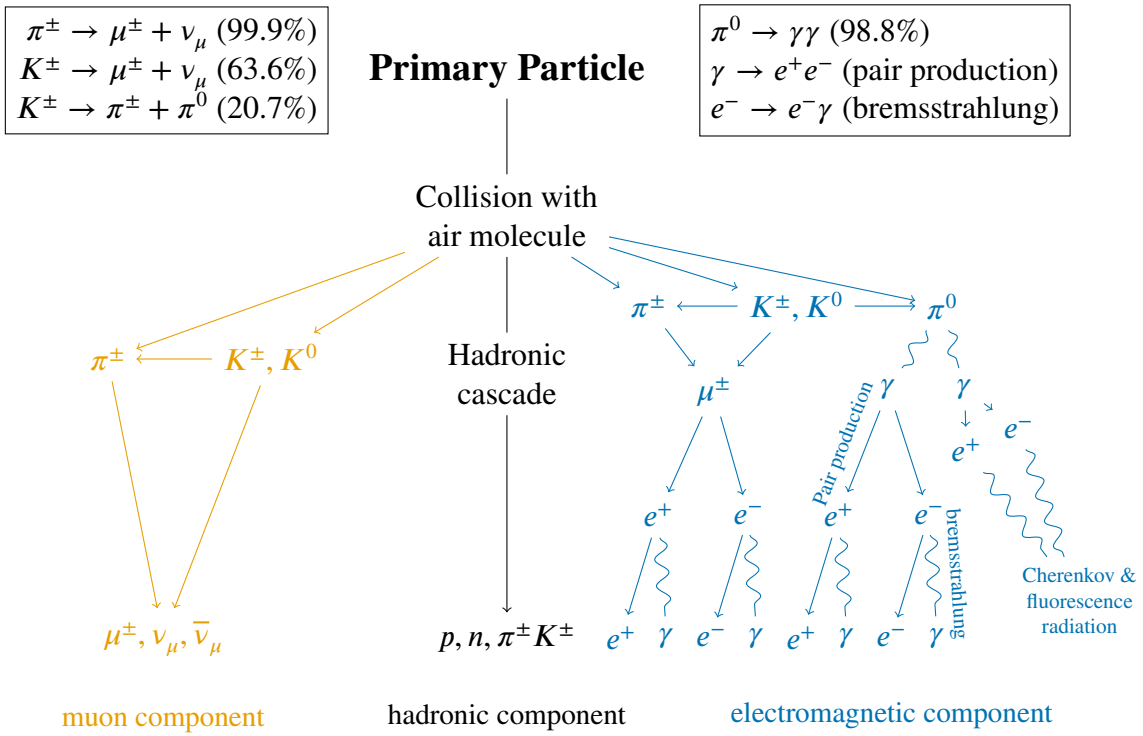

Figure 1.4: Development of a shower produced by a cosmic ray into its three main components with the most frequent particles for each component. Branching ratios for pions and kaons are shown in the boxes [20].

2.1 Electromagnetic showers

An approximate description of electromagnetic showers can be given using the toy model developed by Heitler [21]. In this model the shower develops by bremsstrahlung and pair production. Bremsstrahlung is the process by which a charged particle, such an electron or positron, loses energy by emitting photons. Pair production is the conversion of a photon to a pair electron-positron. At each step of the model the number of particles is doubled such that after n steps the number of particles is $N = 2^n$, see the left panel of Figure 1.5. These processes continue until a critical energy E_c (around 87 MeV in air [22]) is reached. This happens when the energy loses by ionization are equal to the loses due to bremsstrahlung and pair production. At this point the shower reaches its maximum number of particles. Afterwards, particles do not have enough energy for bremsstrahlung or pair production and keep travelling until they interact or reach the ground.

For a given slant depth X[1] of the atmosphere (usually measured in g/cm^2), the number

[1]The slant depth X for a path dr along a medium with density ρ is defined as $X = \int \rho \mathrm{dr}$. X takes into

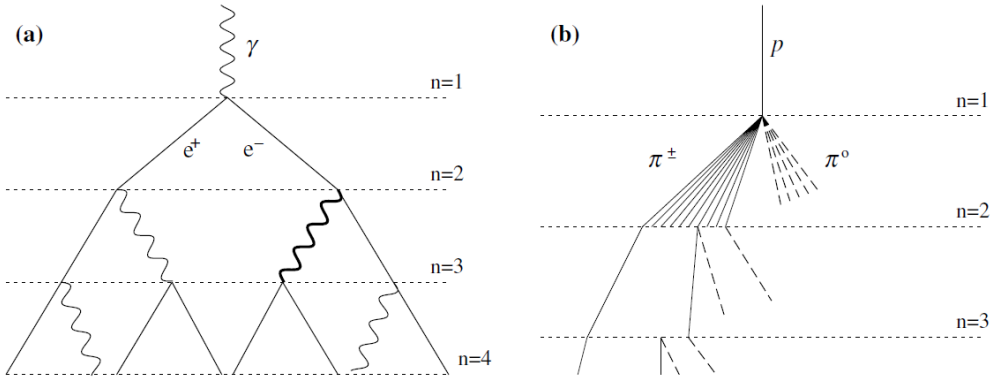

Figure 1.5: Left: Development of a electromagnetic cascade: A photon produces a pair of an electron and positron, these emit photons by bremsstrahlung and the process continues. Right: Development of a hadronic cascade.

of branchings that have taken place is:

$$n = \frac{X}{\lambda \ln 2} \tag{1.1}$$

for a radiation length λ for both the processes of bremsstrahlung and pair production. Since there are 2^n particles at each step and assuming that energy is shared equally amongst particles, at each step each particle carries an energy:

$$E(n) = \frac{E_0}{2^n} \tag{1.2}$$

where E_0 is the energy of the particle that initiated the shower. When the energy of each particle is equal to the critical energy, $E(n) = E_c$, then $X = X_{\max}$ and they are related by Equation 1.1 and Equation 1.2:

$$\frac{E_0}{2^{\frac{X_{\max}}{\lambda \ln 2}}} = E_c \;\Rightarrow\; X_{\max} = \lambda \ln\left(\frac{E_0}{E_c}\right) \tag{1.3}$$

$X_{\max}$ is the position of the shower maximum, where the number of particles is maximum and the deposited energy (proportional to the number of particles) is also maximum.

From Equation 1.3 we have obtained that $X_{\max} \propto \ln(E_0)$. The rate of change of $X_{\max}$ with $\log_{10}$ is called the elongation rate and can be computed from Equation 1.3:

$$\frac{\mathrm{d}X_{\max}}{\mathrm{d}\log_{10}(E)} = \lambda \ln 10 \tag{1.4}$$

account the density of the medium in which the cosmic ray is travelling. The probabibity of interacting depends not only on the distance travelled but also on the density along that distance.

which is about 85 g/cm^2 and it is confirmed by the current models in simulations without great deviations. On the other side, the maximum number of particles in the shower is given by the ratio E_0/E_c, which scales linearly with the energy of the primary particle.

The electromagnetic component of the showers is the one that carries most of the energy of the shower. Furthermore, the hadronic shower feeds the electromagnetic shower with the production of π^0 that decay to photons, as it is explained in the next section.

2.2 Hadronic showers

Hadronic showers can be described by extending the toy model of Heitler. This description was done by J. Matthews [23]. The model is similar to the one for electromagnetic showers: after an interaction length, a hadron produces pions, twice the number of charged compared to the number of neutral ones. Neutral pions decay immediately to photons and contribute to the electromagnetic shower while charged pions continue the hadronic shower. Pions collide and keep producing more pions until their energy is equal to the critical energy. Then, they decay to muons that can reach the ground.

After n interactions, there are a total of $N_\pi = (N_{\text{ch}})^n$ pions, where N_{ch} is the number of charged pions produced in each interaction. Then, assuming that the energy of the initial particle has been shared equally amongst the pions, the energy of each pion is equal to:

$$E_\pi = \left(\frac{2}{3}\right)^n \frac{E_0}{\left(N_{\text{ch}}\right)^n} \tag{1.5}$$

and the factor $2/3$ comes from the fact that in each interaction charged pions only carry $2/3$ of the current energy. Assuming that all the pions produced have the same energy, neutral pions decay into photons carrying $1/3$ of the energy of the parent pion. After the charged pions reach the critical energy E_c^π, they will decay to muons, giving one muon each. The number of muons can then be obtained as the total number of charged pions: $N_\mu = \left(N_{\text{ch}}\right)^{n_{\text{max}}}$. n_{max} is the number of iterations needed to reach the critical energy. From Equation 1.5 n_{max} can be obtained by making $E_\pi = E_c^\pi$:

$$n_{\text{max}} = \frac{\ln(E_0/E_c^\pi)}{\ln\left(\frac{3}{2}N_{\text{ch}}\right)} \tag{1.6}$$

And then the muons as:

$$N_\mu = \left(N_{\text{ch}}\right)^{n_{\text{max}}} \Rightarrow \ln N_\mu = n_{\text{max}} \ln N_{\text{ch}} = \frac{\ln(E_0/E_\pi)}{\ln\left(\frac{3}{2}N_{\text{ch}}\right)} \ln N_{\text{ch}} \tag{1.7}$$

which means that the number of muons does not increase linearly with the energy but as E_0^β for $\beta = \ln N_{ch} / \ln\left(\frac{3}{2} N_{ch}\right)$.

Note that the descriptions given for electromagnetic and hadronic showers are only approximate descriptions of extensive air showers. A complete description of the shower requires simulations where particles are followed. However, the amount of particles generated is very large to be treated even with modern computers. In practice, a technique called thinning and first developed by Hillas is used [24]. The idea is only to follow only a subsample of all the particles. One method to do so is to follow every particle above a certain energy and only a fraction of the particles below this energy. These particles with lower energies are given weights so that energy is conserved. Using the technique of thinning, computing time and memory needed for simulations can be greatly reduced.

2.3 Superposition principle

For showers initiated by nuclei with A nucleons the superposition principle is an approximation that tells us that a nucleus with energy E_0 and mass number A can be modelled as A nucleons having energy E_0/A [23]. The shower can then be treated as a sum of proton showers initiated at the same point. From Equation 1.3 the position of the shower maximum X^A_{max} for a nucleus of mass number A would be the same as in a shower initiated by a proton, X^p_{max}, with energy E_0/A:

$$X^A_{max}(E_0) = X^p_{max}(E_0/A) = X_{max} - \lambda \ln A \tag{1.8}$$

The number of muons is larger for heavier primaries than it is for lighter primaries. It is A times the number of muons when the energy is E_0/A and from Equation 1.7:

$$N_\mu \propto E_0^\beta A^{1-\beta} \tag{1.9}$$

The elongation rate is the same independently of the mass of the primary. From Equation 1.8 and using Equation 1.4:

$$\frac{dX^A_{max}}{d\log_{10}(E)} = \lambda \ln 10 \tag{1.10}$$

3 Properties of cosmic rays and recent results

In this section some of the most important properties and current knowledge of cosmic rays are explained.

3.1 Spectrum

The flux of cosmic rays has been measured by several experiments over many years [25–33]. It extends from 10^9 eV, a GeV, to over 10^{20} eV, more than 100 EeV. The spectrum has a characteristic dependence with energy: it decreases with a power law $E^{-\alpha}$, where E is the energy of the cosmic ray and α is called the spectral index and is approximately equal to 3. In Figure 1.6 the spectrum measured by some experiments has been plotted as a function of the energy. Because of this fast decrease in the flux, ultra-high energy cosmic rays arrive with a very low frequency: cosmic rays having energies around 10^{20} eV arrive with a flux of one per km^2 and century. Thus it is impossible to study cosmic rays of very high energy with experiments in space as huge detectors are needed.

The spectral index α is not constant with the energy. Because of this, the shape of the spectrum curve changes and where those changes occur have been given the names of knee and ankle. At the knee, at around 10^{15} eV, the spectral index changes from 2.7 to 3 while at the ankle, at around 10^{18} eV, the spectrum becomes less steep and α becomes 2.7 again, see Figure 1.6. It is generally thought that most of the cosmic rays with energies below and around the ankle come from our galaxy, the Milky Way. From energies above the ankle, cosmic rays are thought to have extragalactic origin. In this scenario, the knee could be an effect due to propagation of the cosmic rays [34].

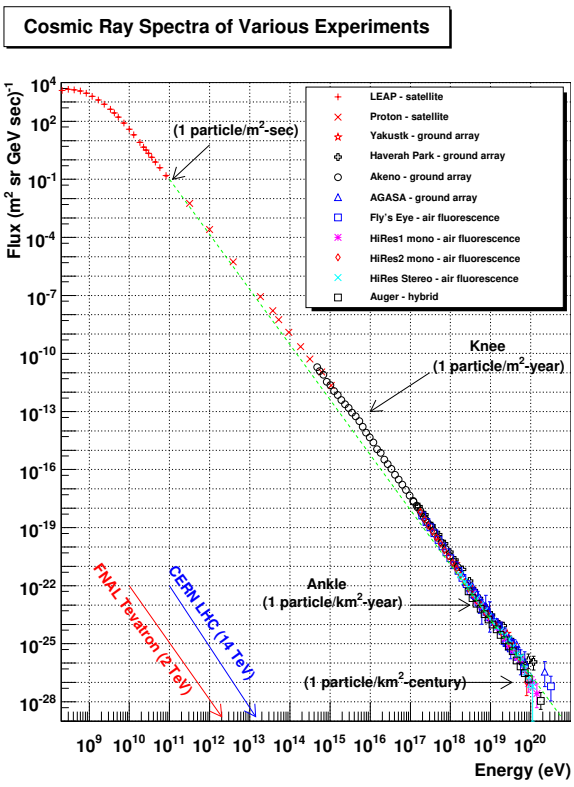

Figure 1.6: Cosmic ray spectra measured by several experiments.

For the highest energies, there is a strong suppression of the flux. This suppression has been measured by the Pierre Auger Observatory to start at
$E = (4.21 \pm 0.17 \pm 0.76) \cdot 10^{19}$ eV
(statistical and systematic uncertainties, respectively) [35]. One reason for this could be the interaction of particles with the cosmic microwave background (CMB). Above a certain threshold, cosmic rays have enough energy to produce a Δ^+ when interacting with the CMB that decays to pions, losing energy in the process. This threshold is known as the Greisen-Zatsepin-Kuzmin (GZK)

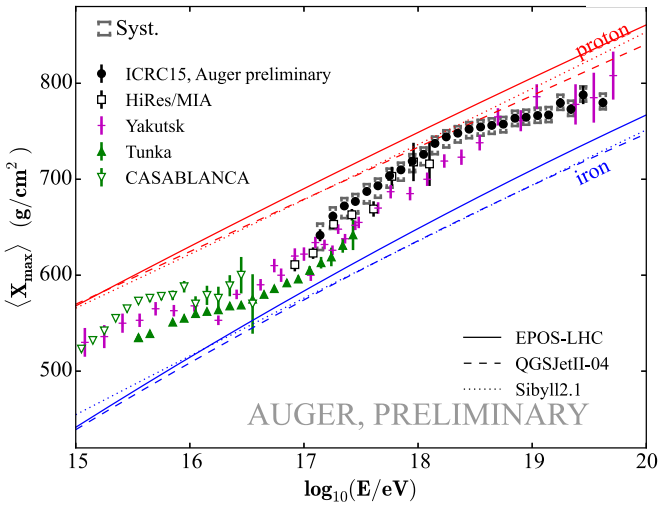

Figure 1.7: Average value of the position of the shower maximum measured by several experiments. Plot taken from ref. [39].

.

limit [36,37], predicted in 1966. For protons, the following process takes place when these protons have energies above this limit:

$$p + \gamma \rightarrow \Delta^{+} \rightarrow p + \pi^{0} \quad \text{or} \quad n + \pi^{+} \tag{1.11}$$

where γ is a photon from the CMB. Protons lose energy in this process and it would not be possible to observe protons with energies above the GZK limit coming from distant sources. However, that this process is happening can not be inferred from the flux suppression, since sources may not be able to accelerate cosmic rays above certain energies.

3.2 Composition

For energies below the knee, most cosmic rays are protons. About 10% are helium nuclei and about 1% are nuclei of heavier elements. That comprises about 99% of the total cosmic rays while electrons, positrons and photons make the rest 1% [34]. Because at these energies the flux is large, there are several experiments that can study these cosmic rays in space, such as AMS or PAMELA. For a review on the composition of cosmic rays see ref. [38].

At higher energies the composition is mostly nuclei. See Figure 1.7 for the evolution of the position of the shower maximum X_{max} with the energy measured by several experiments from 10^{15} eV to 10^{20} eV.

For ultra-high energy cosmic rays, mass composition can be studied with the Fluorescence Detector of The Pierre Auger Observatory. Fluorescence light is emitted by atmo-

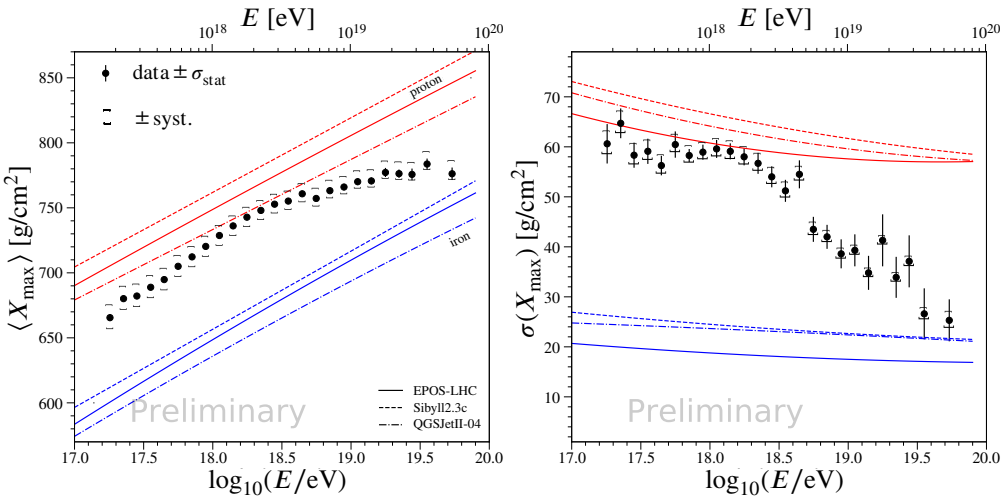

Figure 1.8: Left: Mean position of the shower maximum as a function of the energy for data (black points) and simulations. Right: Second moment of the distribution of X_{max} for data and simulations.

spheric nitrogen when it is ionized by particles coming from a cosmic ray shower. With this information X_{max} can be measured.

The latest results on X_{max} by the Pierre Auger Observatory are shown in Figure 1.8 [40]. For data, the elongation rate (defined in Equation 1.4) of $\langle X_{\mathrm{max}} \rangle$ is $77 \pm 2\,(\mathrm{stat})\,\mathrm{g\,cm^{-2}}$ per decade of energy below $E_0 = 10^{18.32 \pm 0.03}$ and is $26 \pm 2\,(\mathrm{stat})\,\mathrm{g\,cm^{-2}}$ per decade for energies above E_0. The elongation rate for simulations is $60\,\mathrm{g\,cm^{-2}}$ independently of the hadronic model used. This means that the composition changes from heavier to lighter below E_0 and then to heavier again above E_0. The results on the second moment of X_{max} agree with this measurement. Previous published measurements of X_{max} can be found in refs. [41,42].

There are other results on mass composition using the Surface Detector of The Pierre Auger Observatory instead of the Fluorescence Detector. The Surface Detector has the advantage of being operative almost 100% of the time and having increased statistics.

The Delta Method [43,44] uses the information from the risetime of the signals measured by the Surface Detector to infer the mass composition of UHECRs. It is explained later, on page 43, since it is related to the work done in of the chapters of this thesis.

In another work the correlation of the position of the shower maximum X_{max} and the signal measured at the ground is used to obtain information about the spread of the masses in a sample of events [40,45]. Showers initiated by heavier nuclei have a smaller X_{max} and a larger number of muons at the ground than showers initiated by lighter nuclei. This correlation has been studied by computing a correlation coefficient called r_{G} [46]. In Figure 1.9

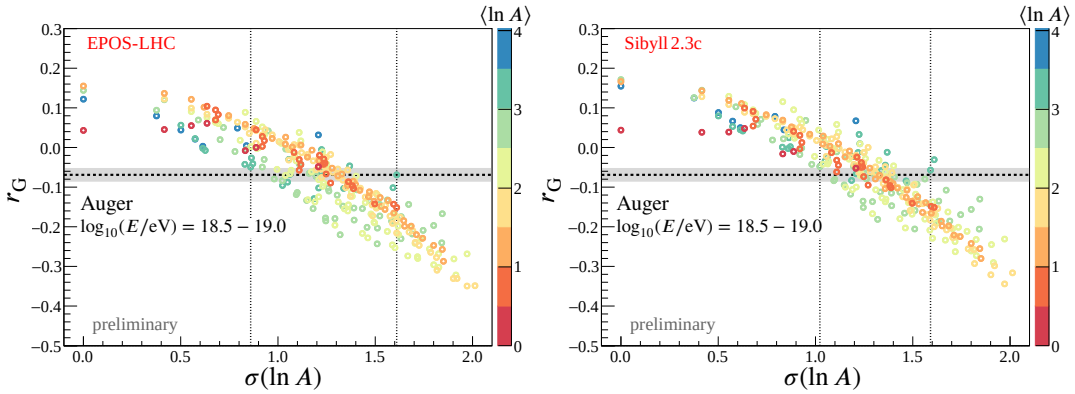

Figure 1.9: Dependence of the correlation coefficient r_G on $\sigma(\ln A)$ for the hadronic models EPOS-LHC [47] (left) and Sibyll 2.3c [48] (right). Each simulated point corresponds to a mixture with different fractions of (p, He, O, Fe). Colours of the points indicate the value of $\langle \ln A \rangle$ of the corresponding simulated mixture. The shaded area shows the observed value for data. Vertical dotted lines indicate the range of $\sigma(\ln A)$ in simulations compatible with the observed correlation in data.

r_G has been plotted for data and for different samples of simulations. The negative correlation found in data can not be reproduced with a pure composition (when $\sigma(\ln A)$ is zero). Correlation for samples with only proton and helium is non-negative, so the values found for data can only be explained when nuclei with $A > 4$ are included in the sample.

3.3 Origin

The origin of cosmic rays is better known for the lowest energies. Cosmic rays with energies around one GeV come from the Sun. For higher energies they can not come from there since at the Sun there are not any processes involving such energies. Up to the knee, cosmic rays are thought to come from Supernova Remnants (SNRs) from our own galaxy [34]. When studying the arrival direction of cosmic rays with energies up to the knee, the flux is found to be very isotropic, consistent with the smearing that the galactic magnetic field would produce.

Cosmic rays with very high energies, around the ankle and above, are thought to be extragalactic. One reason for that is that magnetic fields in our galaxy are not strong enough to confine them. There are also few astrophysical objects that can accelerate cosmic rays up to those energies. There are two main characteristics that influence the maximum energy attainable by cosmic rays at these sources: their size and their magnetic field. In the left panel of Figure 1.10, a comparison of the magnetic field and size of different astrophysical objects has been made. The maximum energy is proportional to the product of the mag-

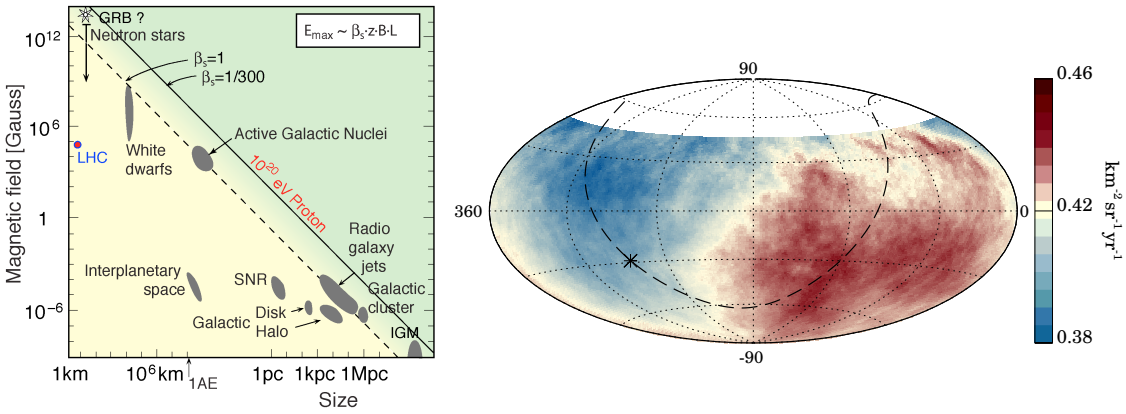

Figure 1.10: Possible cosmic ray accelerators plotted as a function of their magnetic field and size. The dashed and continuous line give the relationship between the magnetic field and size needed to accelerate a proton of 10^{20} eV, at $\beta = 1/300$ and $\beta = 1$ respectively. β is the velocity of the shock that would accelerate the cosmic rays in units of the speed of light c. This plot is usually known as "Hillas Plot" [49]. Right: Sky map in equatorial coordinates, using a Hammer projection, showing the cosmic-ray flux above 8 EeV. The Galactic center is marked with an asterisk and the Galactic plane is shown by a dashed line.

netic field and the size of the object, so only a few objects have the necessary properties to accelerate ultra-high energy cosmic rays. A recent measurement by The Pierre Auger Collaboration finds a large-scale anisotropy in the direction of arrival of UHECRs [50]. This anisotropy indicates an extragalactic origin of these cosmic rays, see the right panel of Figure 1.10.

There are two possible scenarios for cosmic rays to reach those energies: a top-down and a bottom-up scenario. In the top-down scenario, a particle of high mass (for example, a dark matter particle) decays and the products of this process are cosmic rays. In the bottom-up scenario, particles are accelerated by strong turbulent magnetic fields that arise in shock fronts, such as those in a supernova. Current experimental data does not support the existence of top-down processes.

3.4 Hadronic interactions

Hadronic interactions are very present in the study of cosmic rays and their knowledge is linked to the study of how cosmic ray showers develop. We focus on hadronic interactions at high energies, since hadronic interactions at low energies are better understood. We describe a few recent results that show that the current hadronic models can not explain correctly the experimental data. Even though the hadronic models used for high-energy interactions of particles in simulations can be tuned for data measured at the LHC at en-

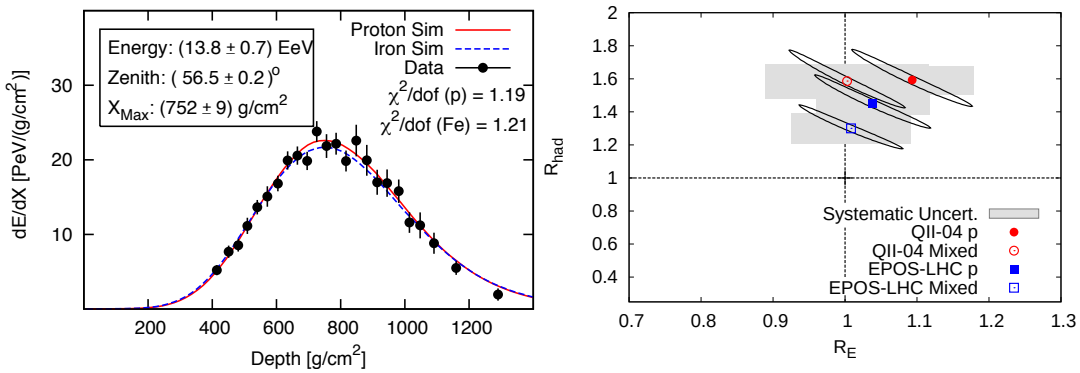

Figure 1.11: Left: Energy deposited as a function of depth for one event in data, a proton and an iron simulation. Right: Best-fit values of R_E and R_{had} for the hadronic models QGSJetII-04 [51] and EPOS-LHC, for pure proton (solid symbols) and mixed composition (open symbols). The ellipses and gray boxes show the 1-σ statistical and systematic uncertainties.

ergies of 14 TeV in the centre of mass, these energies are still below the energies that ultra-high energy cosmic rays carry. Hadronic models have to rely on extrapolations and parameters that are not well known at the energies of ultra-high energy cosmic rays.

In ref. [52] simulations of events measured by The Pierre Auger Observatory are done until the longitudinal profiles measured by the Fluorescence Detector are found to be similar to those in experimental data, see the left panel of Figure 1.11. Then, the signal measured at the ground is compared between data and simulations. With a maximum likelihood method, they find the best parameters R_{had} and R_E that rescale respectively the hadronic and electromagnetic components of the signal at the ground. With this method, simulations with the same longitudinal profile as data are forced to have the same signal at the ground that the data have. For a given shower i and a primary mass j the rescaled signal at the ground in simulations is written as:

$$S_{\text{resc}}(R_E, R_{\text{had}})_{i,j} = R_E S_{EM,i,j} + R_{\text{had}} R_E^{\alpha} S_{\text{had},i,j} \tag{1.12}$$

where R_E and R_{had} are the free parameters of the rescaling, $S_{EM,i,j}$ and $S_{\text{had},i,j}$ are the electromagnetic and hadronic signals at the ground and R_E^{α} is a parameter whose value can be found in simulations. It is found that, independently of the composition, the rescaling on the energy scale is compatible with one (no rescaling needed) while the rescaling of the hadronic component is always above one, see the right panel of Figure 1.11.

This is not the only result on discrepancies between data and simulations. In ref. [53], the number of muons is shown to be larger in data than in simulations for inclined events. In inclined events it is easier to measure the muon number because the effective length of the atmosphere increases, the electromagnetic component of the shower is absorbed and the muon component can be determined directly. In Figure 1.12 the muon content has been compared between data and simulations. There is a large discrepancy even when

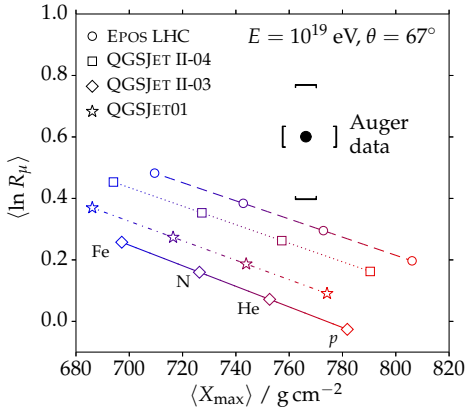

Figure 1.12: Average logarithmic muon content as a function of the average depth of the shower maximum at 10^{19} eV. Model predictions are obtained from showers simulated at $\theta = 67°$. The predictions for proton and iron showers are directly taken from simulations. Values for intermediate masses are computed with the Heitler model.

uncertainties are taken into account. This discrepancy is even larger when assuming the mass composition given by the measured X_{max} in Figure 1.8.

The Pierre Auger Observatory is not the only experiment that has measured a deficit of muons, see ref. [54] for a recent article from The Telescope Array Collaboration.

4 This thesis in the context of studying UHECRs

UHECR were discovered more than 50 years ago [55]. However, we still lack a plausible theory that explains the chemical composition of this radiation, where it is produced and what mechanisms are capable of conferring to these nuclei such extraordinary energies. While the Fluorescence Detector of the Pierre Auger Observatory measures accurately the depth of shower maximum X_{max}, that can be used as a proxy for mass composition, it is limited to operate on nights with good weather conditions. The statistics obtained with this detector are thus not enough at the highest energies, with only a hundred of events with energies above $10^{19.5}$ eV. The Surface Detector (SD) operates all the time but it is limited to sampling the footprint of the shower at the ground level.

In this thesis we focus on the measurements done with the SD to do mass composition studies with the largest amount of statistics available. We study the time series of signals measured by the SD with an observable that we define and with neural networks to predict the muon component of these signals. Knowing the signal left by muons would enhance the information that can be obtained about mass composition of cosmic rays.

2

The Pierre Auger Observatory

1 Introduction

The Pierre Auger Observatory, located in western Argentina, is the largest cosmic ray observatory in the world. It studies cosmic rays above 10^{17} eV, the most energetic particles observed in nature. The design of the Observatory features an array of 1660 water Cherenkov particle detector stations spread over 3000 km^2 overlooked by 24 air fluorescence telescopes. In addition, three high elevation fluorescence telescopes overlook a 23.5 km^2, 61-detector infilled array with 750 m spacing. The Observatory has been in successful operation since completion in 2008. A key feature of the Pierre Auger Observatory is its hybrid design, in which ultra-high energy cosmic rays are detected simultaneously by a surface array and by fluorescence telescopes. The two techniques are used to observe air showers in complementary ways, providing important cross-checks and measurement redundancy [56,57]. For a thorough review of the Observatory see ref. [58].

The surface detector array (SD when referring to the whole array) views a slice of an air shower at ground level, with water Cherenkov stations which respond to both the electromagnetic and muonic components of the shower. The SD operates 24 hours per day and has the important property that the quality of the measurements improves with the shower energy.

The fluorescence detector (FD when referring to all the telescopes) is used to image the longitudinal development of the shower cascade in the atmosphere. The fluorescence light is produced predominantly by the electromagnetic component of the shower. Observation periods are limited to dark nights of good weather, representing a duty cycle up to 15 %. The technique provides a near-calorimetric method for determining the primary cosmic ray energy and the depth at which a shower reaches maximum size, X_{max}. X_{max} is the most direct of all accessible mass composition indicators.

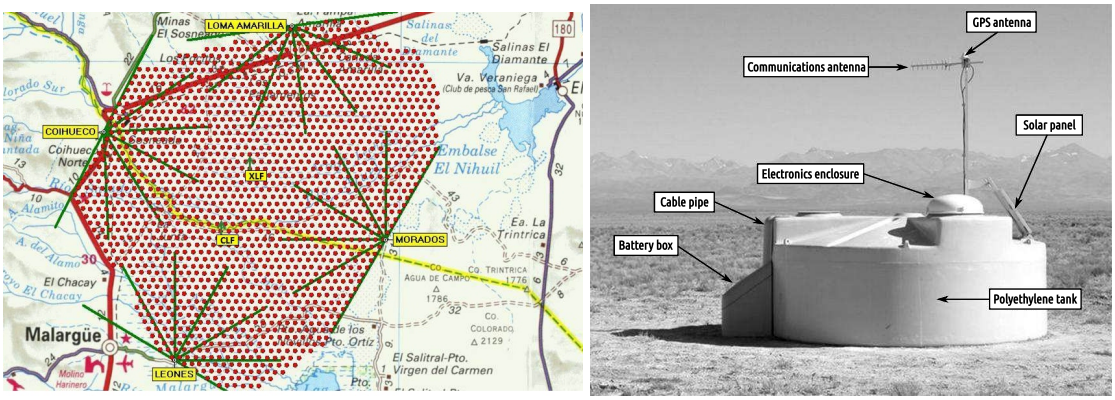

Figure 2.1: Left: The Pierre Auger Observatory. Each dot corresponds to one of the 1660 surface detector stations. The four fluorescence detector enclosures are shown, each with the 30 ° field of view of its six telescopes. Right: A schematic view of a surface detector station in the field, showing its main components.

This chapter is structured as follows. In Section 2 a description of the SD is given and in Section 3 the FD is described. The next important information about the Observatory is how the reconstruction is done for events measured only by the SD in Section 4. The chapter ends with an overview of an improvement of the detector to measure the muon component in Section 5.

2 Surface Detector

The Surface Detector consists on stations or tanks that record Cherenkov light produced by the passage of relativistic charged particles through the water. These stations have 12000 litres of ultra-pure water and three photomultipliers (PMTs) that look into the water. In the left panel of Figure 2.1, the layout of the surface array and the FD buildings at its periphery are shown. The components of a surface detector station are shown in the right panel of Figure 2.1 and described in detail in the next sections.

2.1 SD station

The tanks are made of polyethylene and are low cost, tough and have robustness against the environmental elements. The selected compounded polyethylene resins contained additives to enhance ultraviolet protection. The tanks have an average wall thickness of 1.3 cm and a nominal weight of 530 kg. The tanks do not exceed 1.6 m in height so that they can

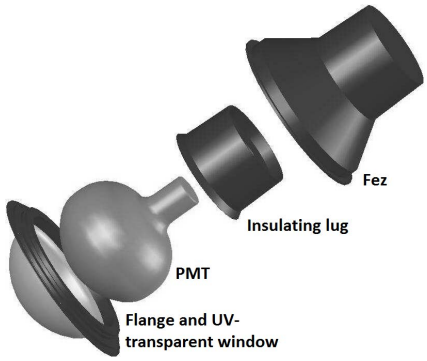

Figure 2.2: Mechanical housing for the SD PMT. Top to bottom: outer plastic housing (fez), insulating lug, PMT, flange, UV-transparent window.

be shipped over the roads within transportation regulations.

Three hatches, located above the PMTs, provide access to the interior of the tank for water filling. They also provide access for installation and servicing of the interior parts. The hatches are covered with light- and water-tight polyethylene hatchcovers. One hatchcover is larger than the other two and accommodates the electronics on its top surface. The tanks also possess molded-in lugs, for lifting and to support the solar panel and antenna mast assembly.

Electrical power is provided by two 55 Wp (watt-peak) solar panels which feed two lead-acid batteries wired in series to produce a 24 V system. The electronics assembly at each SD station possesses a Tank Power Control Board (TPCB) which monitors the power system operation [59]. The TPCB allows the remote operator to set into hibernation any of the SD stations if the charge of the batteries falls below a critical level. The solar panels are mounted on aluminium brackets, which also support a mast. Antennas for radio communication and GPS reception are mounted at the top of this mast.

The tank liners are circular cylinders made of plastic conforming to the inside surface of the tanks up to a height of 1.2 m. They enclose the water volume, provide a light-tight environment and diffusively reflect the Cherenkov light produced in the water volume. The liner has three windows through which the PMTs look into the water volume from above. These windows are made of UV-transparent polyethylene. Each PMT is optically coupled to a window with optical GE silicone and shielded above by a light-tight plastic cover, designated as the “fez”. In Figure 2.2 the PMT enclosure is shown. The fez has four ports, including a light-tight air vent for pressure relief. The other ports are for cable feedthroughs.

Once deployed in their correct positions in the field, the tanks are filled with ultrapure water. Water quality (resistivity) exceeds 15 MΩ cm at the output of the water plant, and the water is transported in clean specialized transport tanks. The water is expected to maintain its clarity without significant degradation for the lifetime of the Observatory.

2.2 SD electronics

To collect the Cherenkov light produced in the water volume of the detectors by the air showers, three PMTs view the water volume from above. The PMTs have a 9 inch diameter photocathode and eight dynodes, are operated at a nominal gain of $2 \cdot 10^5$, are specified for operation at gains up to 10^6 and are required to be linear within 5 % up to 50 mA anode current. Each PMT has two outputs. An AC coupled anode signal is provided. In addition, the signal at the last dynode is amplified and inverted by the PMT base electronics to provide a signal with 32 times the charge gain of the anode. The filtered analog signals are fed to 10 bit 40 MHz semi-flash ADCs, which means that the signal is sampled in bins of 25 ns.

Each SD station contains a GPS receiver for event timing and communications synchronization. This receiver outputs a timed one-pulse-per-second (1 PPS). Event timing is determined using a custom ASIC which references the timing of shower triggers to the GPS 1 PPS clock. The ASIC implements a 27 bit clock operating at 100 MHz. This clock is latched on the GPS 1 PPS signal at the time of each shower trigger. A counter operating at the 40 MHz ADC clock is also latched on the GPS 1 PPS clock. These data are used to calibrate the frequencies of the 40 MHz and 100 MHz clocks and to synchronize the ADC data to GPS time within 10 ns RMS.

The digital data from the ADCs are clocked into a programmable logic device (PLD). The PLD implements firmware that monitors the ADC outputs for interesting trigger patterns, stores the data in a buffer memory, and informs the station microcontroller when a trigger occurs. There are two local trigger levels (T1 and T2) and a global third level trigger, T3. Details of the local triggers are described on page 23.

The front end is interfaced to a unified board which implements the station controller, event timing, and slow control functions, together with a serial interface to the communications system. The slow control system consists of DACs and ADCs used to measure temperatures, voltages, and currents relevant to assessment of the operation of the station.

The data acquisition system implemented on the station controller transmits the time stamps of the $\sim$ 20 T2 events collected each second to CDAS (Central Data Acquisition System). The station controller then selects the T1 and T2 data corresponding to the T3 requests and builds it into an event for transmission to CDAS. Calibration data are included in each transmitted event.

2.3 SD signal saturation

As it has been explained before, signals measured by PMTs are obtained from two channels. The signal from the anode is usually denoted as the signal coming from the high-gain

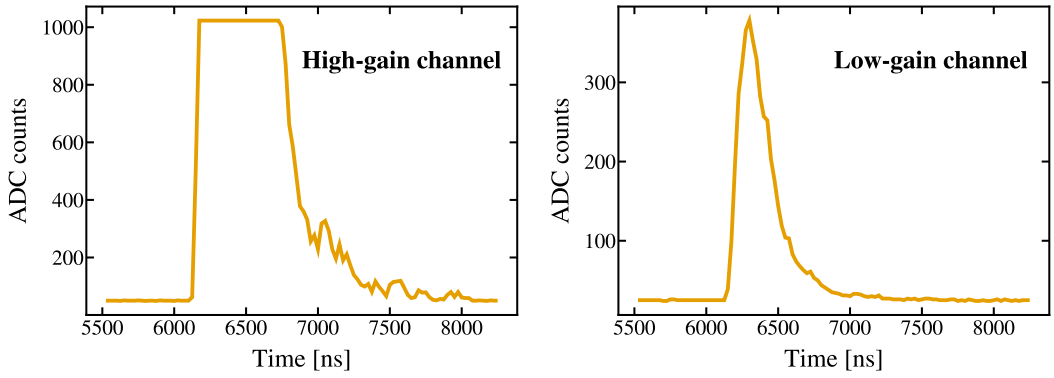

Figure 2.3: Left: Saturated signal from the high-gain channel for one PMT. The signal is capped at around 1000 ADC counts. Right: Corresponding signal from the low-gain channel.

channel while the last-dynode signal that is amplified is denoted as the signal coming from the low-gain channel. Sometimes, signals are large and the PMT can not measure their shape. This is known as saturation. The saturation is caused by the overflow of the FADC read-out electronics with finite dynamic range and a modification of the signal due to the transition of the PMTs from a linear to a non-linear behavior. In the majority of cases the missing part of the signals are recovered using the procedure described in ref. [60].

In the left panel of Figure 2.3, there is an example of a signal from the high-gain channel that is saturated. In this case the signal from the low-gain channel, shown in the right panel of Figure 2.3, can be used instead. However, un-saturated signals from the high-gain and low-gain channels have different shapes. This fact complicates the use of these signals from the low-gain channel, that have to be studied separately from the signals from the high-gain channel.

Signals measured at stations that are very close to the core of the shower can be saturated both in the high-gain and low-gain channels. In this case and with the current design of the SD, nothing can be done to recover the true shape of the signal.

2.4 SD calibration

The Cherenkov light recorded by a surface detector is measured in units of the signal produced by a muon traversing the tank on a vertical trajectory, see Figure 2.4. This unit is termed the Vertical Equivalent Muon (VEM). The goal of the surface detector calibration is to measure the value of 1 VEM in hardware units (integrated FADC channels). The conversion to units of VEM is done both to provide a common reference level between tanks and to calibrate against the detector simulations.

We define $Q_{\mathrm{VEM}}^{\mathrm{peak}}$ (denoted simply by Q_{VEM} hereafter) as the bin containing the peak in

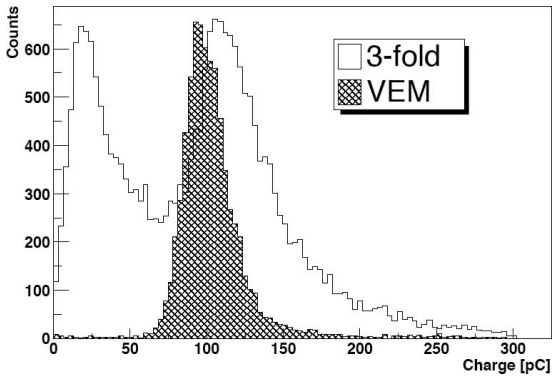

Figure 2.4: Charge spectrum obtained when a surface detector is triggered by a threefold coincidence among its photomultipliers (open histogram). The hatched histogram shows the spectrum when triggered on central vertically aligned plastic scintillators. The bin containing the peak of the scintillator triggered spectrum is defined as a vertical equivalent muon. The leftmost peak in the open histogram is due to low energy and corner-clipping muons convolved with the threefold low threshold coincidence.

the charge histogram of an individual PMT response, and $I_{\mathrm{VEM}}^{\mathrm{peak}}$ (denoted by I_{VEM} hereafter) as the bin containing the peak in the pulse height histogram. These quantities are used in the three main steps in the calibration procedure:

1. Set up the end-to-end gains of each of the three PMTs to have I_{VEM} at 50 channels. The choice of 50 ch/I_{VEM} results in a mean gain close to $3.4 \cdot 10^5$ for a mean n_{pe}/VEM of around 94 photo-electrons.

2. To compensate for drifts, adjust the electronics level trigger by continually performing a local calibration to determine I_{VEM} in channels.

3. Determine the value of Q_{VEM} to high accuracy using charge histograms, and use the known conversion from Q_{VEM} to 1.0 VEM to obtain a conversion from the integrated signal of the PMT to VEM units.

The high voltages, and thus the gains of each of the three PMTs, are tuned to match a reference event rate. This tuning implies that the PMTs in the SD stations will not have equivalent gains, even for PMTs in the same tank.

In addition to the primary conversion from integrated channels to VEM units, the calibration must also be able to convert the raw FADC traces into integrated channels. The primary parameters needed for this are the baselines of all six FADC inputs, and the gain ratio between the dynode and anode. The dynode/anode ratio, or D/A, is determined by averaging large pulses and performing a linear time-shifted fit to obtain both D/A and the phase delay between the dynode and anode. This method determines D/A to 2 %.

The calibration parameters are determined every 60 s. The most recently determined parameters are returned to CDAS with each event and stored along with the event data. Each event therefore contains information about the state of each SD station in the minute preceding the trigger, allowing for an accurate calibration of the data [61].

2.5 SD local triggers

Several independent local trigger functions are implemented in the front-end electronics: the scaler trigger, the calibration trigger, and the main shower trigger.

The scaler trigger records pulses with a very low threshold for auxiliary physics purposes such as space weather. The calibration trigger collects low threshold pulses using a small number of bins (20), which is one bin above 0.1 I_{VEM}, thus providing high rate cosmic ray data. Data from the three high-gain channels are stored from three samples before the trigger to 20 samples after the trigger. These data are used to build calibration histograms such as the one shown in Figure 2.4, and are also used to convert offline the six FADC traces into VEM units.

The main trigger is the shower trigger that results in the recording of 768 samples (19.2 µs) of the six FADCs. It has two levels of selection. The first level, called T1, has 2 independent modes. The first one is a simple threshold trigger (TH) requiring the coincidence of all three PMTs being above 1.75 I_{VEM}. This trigger is used to select large signals that are not necessarily spread in time. It is particularly effective for the detection of very inclined showers that have penetrated through a large atmospheric depth and are consequently dominantly muonic. The threshold has been adjusted to reduce the rate of atmospheric muon triggers from about 3 kHz to 100 Hz. The second T1 mode is a time-over-threshold trigger (ToT) requiring that at least 13 bins within a 3 µs window (120 samples) exceed a threshold of 0.2 I_{VEM} in coincidence for two out of the three PMTs. The ToT trigger selects sequences of small signals spread in time, and is thus efficient for the detection of vertical events, and more specifically for stations near the core of low-energy showers, or stations far from the core of high-energy showers. The rate of the ToT trigger depends on the shape of the muon pulse in the tank and averages 1.2 Hz with a rather large spread (about 1 Hz rms). The second trigger level, called T2, is applied to decrease the global rate of the T1 trigger down to about 23 Hz. While all T1-ToT triggers are promoted T2-ToT, only T1-TH triggers passing a single threshold of 3.2 I_{VEM} in coincidence for the three PMTs will pass this second level and become T2-TH. All T2s send their timestamp to CDAS for the global trigger (T3) determination [62].

Since June 2013, there are two more T1 triggers. The time-over-threshold-deconvolved (ToTd) trigger deconvolves the exponential tail of the diffusely reflected Cherenkov light pulses before applying the ToT condition. This has the effect of reducing the influence of muons in the trigger, since the typical signal from a muon, with fast rise time and

Figure 2.5: FD building at Los Leones during the day. Shutters are open because of maintenance. Behind the building there is a communication tower.

around 60 ns decay constant, is compressed into one or two time bins. The multiplicity-of-positive-steps trigger (MoPS), on the other hand, counts the number of positive-going signal steps in two of three PMTs within a 3 µs sliding window. The steps are required to be above a small FADC value ($\approx$ 5$\times$ RMS noise) and below a moderate value ($\approx \frac{1}{2}$ vertical muon step). This reduces the influence of muons in the trigger. Both the ToTd and MoPS triggers also require the integrated signal to be above $\approx$ 0.5 VEM. Because these triggers minimize the influence of single muons, they reduce the energy threshold of the array, while keeping random triggers at an acceptable level. Thus they improve the energy reach of the SD, as well as improve the trigger efficiency for photon and neutrino showers.

3 Fluorescence Detector

The 24 telescopes of the Fluorescence Detector (FD) overlook the SD array from four sites: Los Leones, Los Morados, Loma Amarilla and Coihueco [63]. Six independent telescopes are located at each FD site in a clean climate controlled building [64], an example of which is seen in Figure 2.5. A single telescope has a field of view of 30 ° $\times$ 30 ° in azimuth and elevation, with a minimum elevation of 1.5 ° above the horizon. The telescopes face towards the interior of the array so that the combination of the six telescopes provides 180 ° coverage in azimuth.

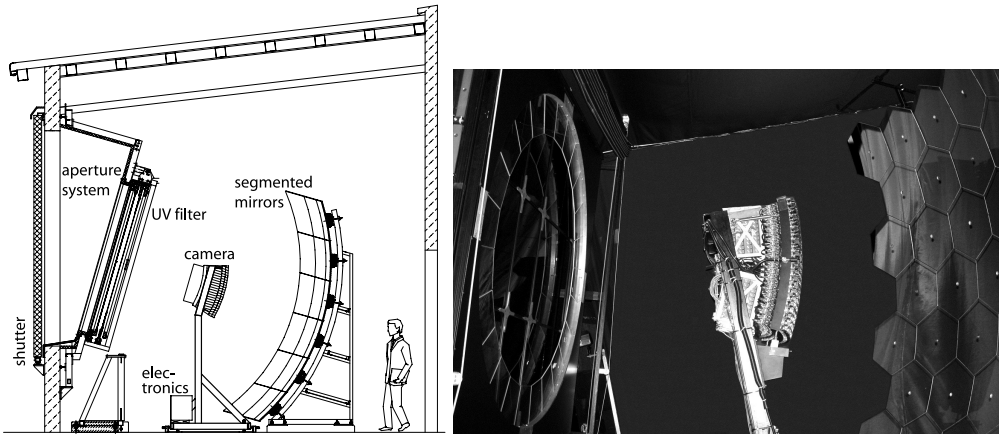

Figure 2.6: Left: Schematic view of a fluorescence telescope with a description of its main components. Right: Photo of a fluorescence telescope at Coihueco.

3.1 FD telescopes

The details of the fluorescence detector telescope and an actual view of an installed telescope are shown in Figure 2.6. The telescope design is based on Schmidt optics because it reduces the coma aberration of large optical systems. Nitrogen fluorescence light, emitted isotropically by an air shower, enters through a circular diaphragm of 1.1 m radius covered with a filter glass window. The filter transmission is above 50 % (80 %) between 310 and 390 nm (330 and 380 nm) in the UV range. The filter reduces the background light flux and thus improves the signal-to-noise ratio of the measured air shower signal. It also serves as a window over the aperture which keeps the space containing the telescopes and electronics clean and climate controlled. The shutters seen in Figure 2.6 are closed during daylight and also close automatically at night when the wind becomes too high or rain is detected. In addition, a fail-safe curtain is mounted behind the diaphragm to prevent daylight from illuminating a camera in case of a malfunction of the shutter or a failure of the Slow Control System, in charge of allowing remote operations of the FD system.

The light is focused by a spherical mirror of around 3400 mm radius of curvature onto a spherical focal surface with radius of curvature close to 1700 mm. Due to its large area (13 m^2), the primary mirror is segmented to reduce the cost and weight of the optical system. Two alternative segmentation configurations are used: one is a tessellation of 36 rectangular anodized aluminium mirrors of three different sizes; the other is a structure of 60 hexagonal glass mirrors (of four shapes and sizes) with vacuum deposited reflective coatings [64]. The average reflectivity of cleaned mirror segments at a wavelength $\lambda = 370$ nm is more than 90 %.

The camera body is machined from a single aluminium block of 60 mm thickness,

with an outer radius of curvature of 1701 mm and an inner curvature radius of 1641 mm. The hexagonal photomultiplier tubes are positioned inside 40 mm diameter holes drilled through the camera block at the locations of the pixel centres. The pixels are arranged in a matrix of 22 rows by 20 columns.

The PMT boundaries are approximate hexagons with a side to side distance of 45.6 mm. The PMTs are separated by simplified Winston cones secured to the camera body which collect the light to the active cathode of the photomultiplier tube. The light collectors serve to prevent photons from landing in the dead spaces between the PMT cathodes. The upper edge of the light collectors lie on the focal surface of 1743 mm radius. The pixel field of view defined by the upper edges corresponds to an angular size of 1.5 °.

The contribution of reflection and scattering inside the optical system of the telescope has been measured in situ and with an airborne remotely controlled platform carrying an isotropic and stabilized UV light source [65]. The measured point spread function of the light distribution in pixels has been implemented in the software used in the air shower reconstruction.

Cleaning and maintenance work has been required during years of detector operation. The cleaning of the UV filter from outside has been performed several times because of deposited dust layers. The equipment inside the building is cleaned less frequently, because it is not exposed to the outside environment. The reflectivity of a few selected mirror segments is measured once or twice each year and it changes less than 1 % per year.

From the end-to-end calibration, the appropriate constants are found to be approximately 4.5 photons/ADC count for each pixel. To derive a flux of photons for observed physics events, the integrated ADC number is multiplied by this constant and divided by the area of the aperture. The flux in photons per m^2 perpendicular to the arrival direction is thus obtained.

The relative spectral efficiencies, or multi-wavelength calibrations, of FD telescopes were measured using a monochromator-based drum light source with a xenon flasher. The measurement was done in steps of 5 nm from 270 nm to 430 nm. As described on page 25, there are two types of mirrors and two different glass materials used for the corrector rings in the FD telescopes. In total eight telescopes were measured to have a complete coverage of the different components and a redundant measure of each combination. The uncertainty of these measurements is close to 3 %. An example of measured relative efficiency of an FD telescope is shown in Figure 2.7.

3.2 FD operation

All FD telescopes are operated remotely from the central campus and other places around the world by shift personnel. Their responsibilities include preparation of the FD for a run, making relative calibrations, starting and stopping runs and online checking of the quality

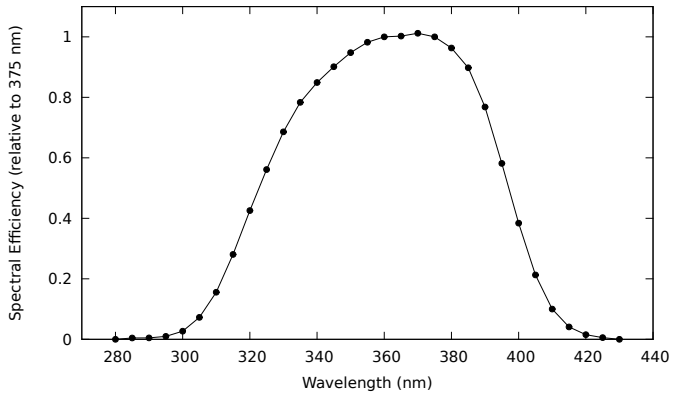

Figure 2.7: Relative efficiency between 280 nm and 430 nm measured for the telescope 3 at Coihueco. The curve is taken relative to the efficiency of the telescope at 375 nm.

of measured data [66]. The observation of air showers via fluorescence light is possible only at night. Moreover, night sky brightness should be low and thus nights without a significant amount of direct or scattered moonlight are required. The mean length of the dark observation period is then 17 nights each month with an on-time of the FD telescopes of ∼15%. The telescopes are not operated when the weather conditions become dangerous (high wind speed, rain, snow, etc.).

4 SD event reconstruction

The reconstruction of the energy and the arrival direction of the cosmic rays producing air showers that have triggered the surface detector array is based on the sizes and times of signals registered from individual SD stations.

4.1 Event selection

To ensure good data quality for physics analysis there are two additional off-line triggers. The physics trigger, T4, is needed to select real showers from the set of stored T3 data that also contain background signals from low energy air showers. This trigger is mainly based on a coincidence between adjacent detector stations within the propagation time of the shower front. In selected events, random stations are identified by their time incompatibility with the estimated shower front. The time cuts were determined such that 99 % of the stations containing a physical signal from the shower are kept. An algorithm for

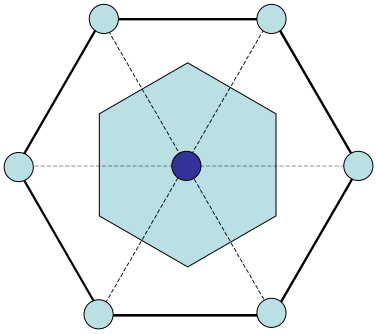

Figure 2.8: Schematic view of the area (shaded region) where the core of a vertical shower must be located inside an elementary hexagonal cell of the SD array to pass the quality trigger for a complete hexagon with 6 active neighbors.

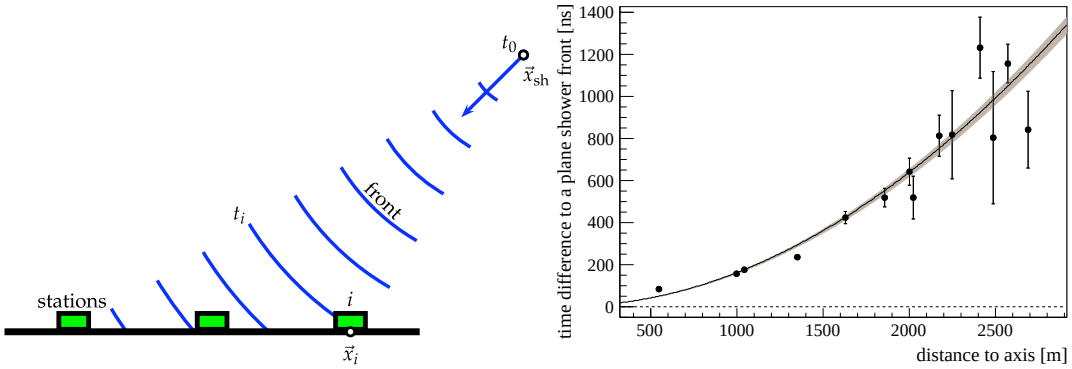

Figure 2.9: Left: schematic representation of the evolution of the shower front. Right: dependence of signal start times (relative to the timing of a plane shower front) on perpendicular distance to the shower axis. The shaded line is the resulting fit of the evolution model and its uncertainty.

the signal search in the time traces is used to reject signals produced by random muons by searching for time-compatible peaks.

To guarantee the selection of well-contained events, a fiducial cut (called the 6T5 trigger) is applied so that only events in which the station with the highest signal is surrounded by all 6 operating neighbors (i.e., a working hexagon) are accepted, see Figure 2.8. This condition assures an accurate reconstruction of the impact point on the ground, and at the same time allowing for a simple geometrical calculation of the aperture and exposure [62].

4.2 Shower geometry

A rough approximation for the arrival direction of the shower is obtained by fitting the start times of the signals, t_i, in individual SD stations to a plane front. For events with enough

triggered stations, these times are described by a more detailed concentric-spherical model, see the left panel of Figure 2.9. This model approximates the evolution of the shower front with a speed-of-light inflating sphere,

$$c(t_i - t_0) = |\vec{x}_{\rm sh} - \vec{x}_i| \tag{2.1}$$

where $\vec{x}_i$ are the positions of the stations on the ground and $\vec{x}_{\rm sh}$ and t_0 are a *virtual* origin and a start-time of the shower development, see the right panel of Figure 2.9. From this 4-parameter fit the radius of curvature of the inflating sphere is determined from the time at which the core of the shower is inferred to hit the ground.

4.3 Lateral distribution function

The impact points of the air showers on the ground, $\vec{x}_{\rm gr}$, are obtained from fits of the signals in SD stations. This fit of the lateral distribution function (LDF) is based on a maximum likelihood method which also takes into account the probabilities for the stations that did not trigger and the stations close to the shower axis with saturated signal traces.

An example of the footprint on the array of an event produced by a cosmic ray and the lateral distribution of the signals are depicted in Figure 2.10. The function employed to describe the lateral distribution of the signals on the ground is a modified Nishimura-Kamata-Greisen function [67,68],

$$S(r) = S(r_{\rm opt}) \left(\frac{r}{r_{\rm opt}}\right)^{\beta} \left(\frac{r + r_1}{r_{\rm opt} + r_1}\right)^{\beta+\gamma} \tag{2.2}$$

where $r_{\rm opt}$ is the optimum distance, $r_1 = 700\,\text{m}$ and $S(r_{\rm opt})$ is a free parameter and also an estimator of the shower size used in an energy assignment. For the SD array with station spacing of 1.5 km the optimum distance [69] is $r_{\rm opt} = 1000\,\text{m}$ and the shower size is thus $S(1000)$, also written as S_{1000}. The free parameter β depends on the zenith angle and shower size. Events up to zenith angle 60 ° are observed at an earlier shower age than more inclined ones, thus having a steeper LDF due to the different contributions from the muonic and the electromagnetic components at the ground. For events with only 3 stations, the reconstruction of the air showers can be obtained only by fixing the two parameters β and γ to a parametrization obtained using events with a number of stations larger than 4.

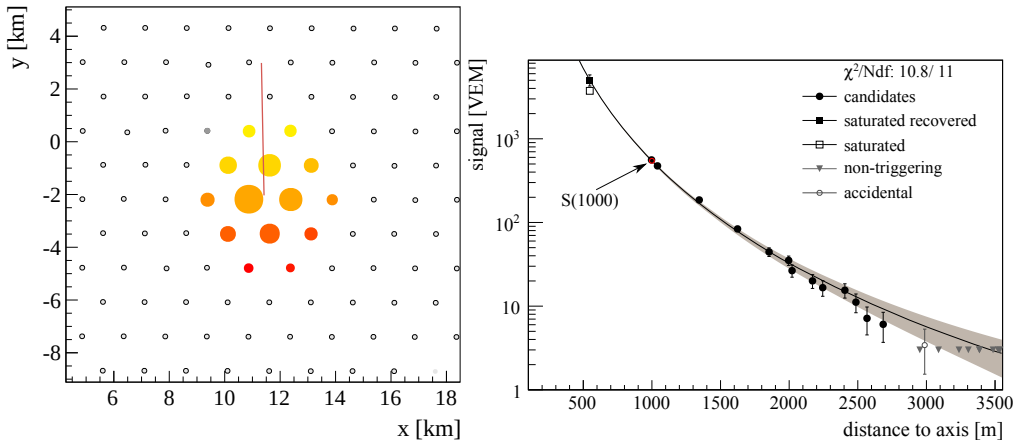

Figure 2.10: Left: Example of signal sizes an extensive air shower induces in the stations of the surface detector array. Colours represent the arrival time of the shower front from early (yellow) to late (red) and the size of the markers is proportional to the logarithm of the signal. The line represents the shower arrival direction. Right: Dependence of the signal size on distance from the shower core.

4.4 Shower arrival direction

The shower axis $\hat{a}$ is obtained from the virtual shower origin (of the geometrical reconstruction) and the shower impact point on the ground (from the LDF reconstruction),

$$\hat{a} = \frac{\vec{x}_{\mathrm{sh}} - \vec{x}_{\mathrm{gr}}}{|\vec{x}_{\mathrm{sh}} - \vec{x}_{\mathrm{gr}}|}. \tag{2.3}$$

To estimate an angular resolution of the whole reconstruction procedure a single station time variance is modeled [70] to take into account the size of the total signal and the time evolution of the signal trace. As shown in the left panel of Figure 2.11, the angular resolution achieved for events with more than three stations is better than 1.6 °, and better than 0.9 ° for events with more than six stations [71].

4.5 Energy calibration

For a given energy, the value of $S(1000)$ decreases with the zenith angle θ due to the attenuation of the shower particles and geometrical effects. We extract the shape of the attenuation curve (see the right panel of Figure 2.11) from the data using the Constant Intensity Cut (CIC) method [72]. The attenuation curve $f_{\mathrm{CIC}}(\theta)$ has been fitted with a third degree polynomial in $x = \cos^2\theta - \cos^2\bar{\theta}$, i.e., $f_{\mathrm{CIC}}(\theta) = 1 + a\,x + b\,x^2 + c\,x^3$, where

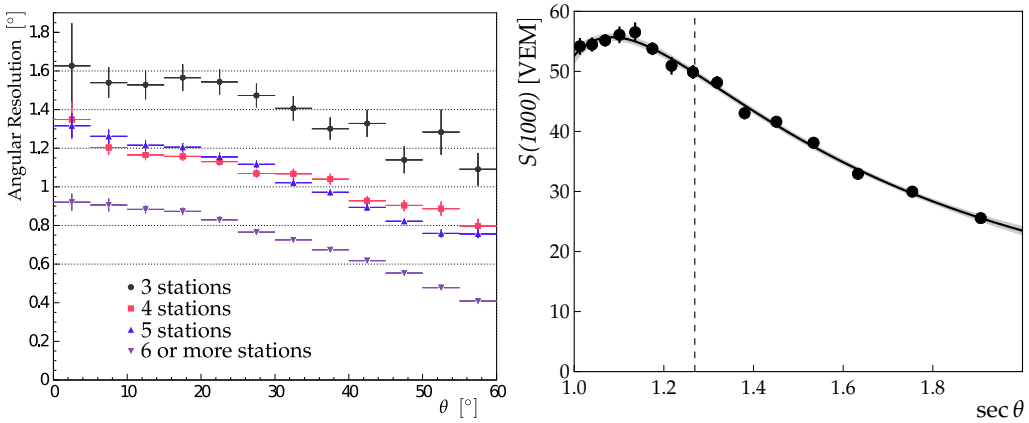

Figure 2.11: Left: Angular resolution as a function of the zenith angle θ for events with an energy above 3 EeV, and for various station multiplicities [71]. Right: Attenuation curve described by a third degree polynomial in $x = \cos^2\theta - \cos^2\bar{\theta}$ where $\bar{\theta} = 38°$ (denoted by the dashed vertical line). In this example the polynomial coefficients are deduced from $S(1000)$ dependence at $S_{38} \approx 50$ VEM which corresponds to an energy of about 10.5 EeV.

$a = 0.980 \pm 0.004$, $b = -1.68 \pm 0.01$, and $c = -1.30 \pm 0.45$ [73]. The median angle, $\bar{\theta} = 38°$, is taken as a reference point to convert $S(1000)$ to $S_{38} \equiv S(1000)/f_{\mathrm{CIC}}(\theta)$. S_{38} may be regarded as the signal a particular shower with size $S(1000)$ would have produced had it arrived at $\theta = 38°$.

To estimate the energy of the primary particle producing the air-showers recorded with the SD, the advantage comes from the hybrid detection: the air-showers that have triggered independently the FD and SD are used for the cross-calibration. High-quality hybrid events, as defined below, with reconstructed zenith angles less than 60° are used to relate the shower size from SD to the almost-calorimetric measurement of the shower energy from FD, E_{FD}. These hybrid events must be such that the reconstruction of an energy estimator can be derived independently from both the SD and FD parts of the event [74,75].

Only a subsample of events that passes strict quality and field of view cuts is used. For the FD part of the event, we require an accurate fit of the longitudinal profile to the Gaisser-Hillas function. Furthermore, the depth of the shower maximum, X_{max}, must be contained within the telescope field-of-view and measured with an accuracy better than 40 g/cm^2. The uncertainty on the reconstructed E_{FD} is required to be less than 18 % and the atmosphere conditions have to be good. To avoid any potential bias of the event selection on the mass of the primary particle, a fiducial cut on the slant depth range observed by the telescopes is also added [74].

The final step in the calibration analysis leads to a relation between S_{38} and E_{FD}. The 1475 high quality hybrid events recorded between Jan 2004 and Dec 2012 which have an energy above the SD full efficiency trigger threshold [62] are used in the calibra-

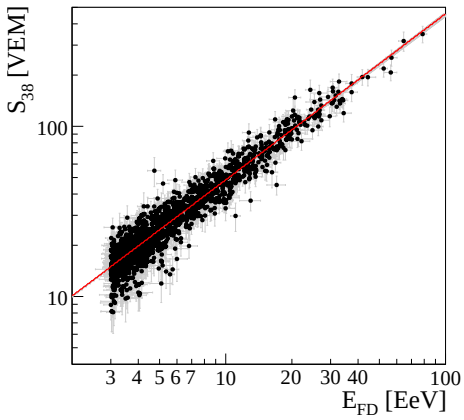

Figure 2.12: Correlation between S_{38} and E_{FD} [73,74].

tion. The correlation between the two variables is obtained from a maximum likelihood method [74,76]. The relation between S_{38} and E_{FD} is well described by a single power-law function,

$$E_{\mathrm{FD}} = A\,(S_{38}/\mathrm{VEM})^{B} \tag{2.4}$$

where the resulting parameters from the data fit are $A = (1.90 \pm 0.05) \times 10^{17}\,\mathrm{eV}$ and $B = 1.025 \pm 0.007$ [73,77], see Figure 2.12.

The final SD energy estimator is:

$$E_{\mathrm{SD}} = A(S(1000)/f_{\mathrm{CIC}}(\theta)/\mathrm{VEM})^{B} \tag{2.5}$$

and its resolution can be inferred from the distribution of the ratio $E_{\mathrm{SD}}/E_{\mathrm{FD}}$. Using the FD energy resolution of 7.6 %, the resulting SD energy resolution with its statistical uncertainty is $\sigma_{E_{\mathrm{SD}}}/E_{\mathrm{SD}} = (16 \pm 1)\%$ at the lower energy edge in Figure 2.12 and $(12 \pm 1)\%$ at the highest energies. Due to the large number of events accumulated until December 2012, the systematic uncertainty on the SD energy due to the calibration is better than 2 % over the whole energy range. The systematic uncertainties in the energy scale, shown in Table 2.1, are dominated by the absolute FD calibration [77].

The dataset recorded extends up to larger angles of 90 °. For the inclined events, with zenith angles larger than 60 °, we employ a different reconstruction method [76,78,79].

5 Auger Muon Infilled Ground Array (AMIGA)

The AMIGA enhancement, a dedicated detector to directly measure the muon content of air showers [80–82], is a joint system of water Cherenkov and buried scintillator detectors that spans an area of 23.5 km^2 in a denser array with 750 m spacing nested within the

Absolute fluorescence yield	3.4%
Fluorescence spectrum and quenching parameters	1.1 %
Subtotal, Fluorescence yield	**3.6 %**
Aerosol optical depth	3–6 %
Aerosol phase function	1 %
Wavelength dependence of aerosol scattering	0.5 %
Atmospheric density profile	1 %
Subtotal, Atmosphere	**3.4–6.2 %**
Absolute FD calibration	9 %
Nightly relative calibration	2 %
Optical efficiency	3.5 %
Subtotal, FD calibration	**9.9 %**
Folding with point spread function	5 %
Multiple scattering model	1 %
Simulation bias	2 %
Constraints in the Gaisser-Hillas fit	3.5–1 %
Subtotal, FD profile reconstruction	**6.5–5.6 %**
Invisible energy	**3–1.5 %**
Statistical error of SD calibration fit	**0.7–1.8 %**
Stability of the energy scale	**5 %**
Total	**14 %**

Table 2.1: Systematic uncertainties in the energy scale.

1500 m array, see Figure 2.13. The 750 m array is fully efficient from $3 \cdot 10^{17}$ eV onwards for air showers with zenith angle $\leq 55\,^{\circ}$ [83], allowing the study of the region between the second knee [84] and the ankle of the cosmic ray spectrum.

The first prototype hexagon of buried scintillators, the *Unitary Cell*, consists of seven water Cherenkov detectors paired with 30 m^2 scintillators segmented in two modules of 10 m^2 plus two of 5 m^2 in each position. In addition, two positions of the hexagon were equipped with *twin* detectors (extra 30 m^2 scintillators) to allow the accuracy of the muon counting technique to be experimentally assessed [85] and one position has 20 m^2 of extra scintillators buried at a shallower depth to analyze the shielding features. The proven tools and methods used for the analysis of the 1500 m SD array data have been extended to reconstruct the lower energy events [86].

The buried scintillators are the core of the detection system for the muonic component

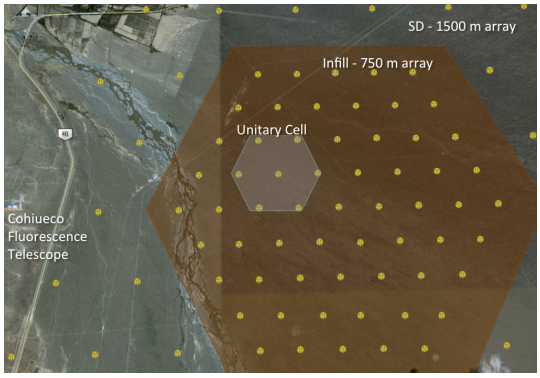

Figure 2.13: AMIGA layout: an infill of surface stations with an inter-detector spacing of 750 m plus plastic scintillators of 30 m^2 buried under $\approx$ 540 g/cm^2 of vertical mass to measure the muon component of the showers.

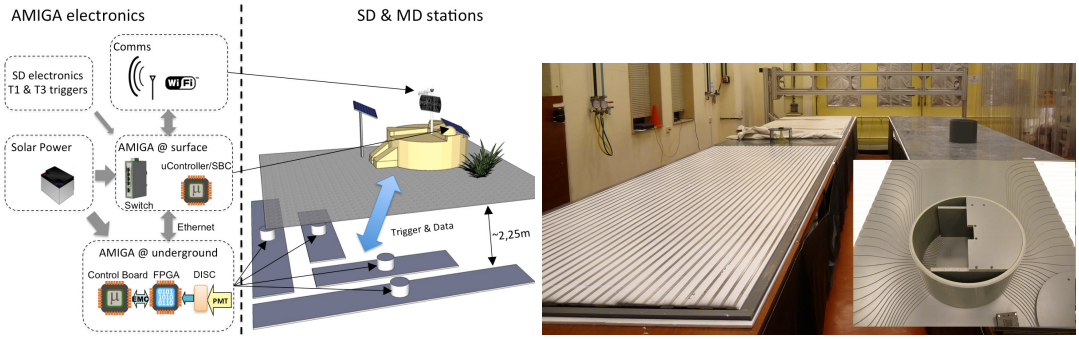

Figure 2.14: Left: AMIGA station: SD+MD paired detectors. The buried front end electronics is serviceable by means of an access pipe which is filled with local soil bags. Right: AMIGA scintillator detector, illustrating the assembly of a 10 m^2 module. Strips are grouped in two sets of 32 strips on each side of the electronics dome located at the centre of the detector. The multi-anode PMT and front end electronics board are hosted in the central dome.

of air showers (the muon detector, MD). To effectively shield the electromagnetic component, the MD is placed under a depth of 2.3 m in the local soil while the shallower extra scintillators are at 1.3 m. The layout of SD+MD paired stations is shown in Figure 2.14. The scintillator surface of each MD station is highly segmented. It consists of modules made of 64 strips each. The manifold of fibers of each module ends in an optical connector matched to a 64 multi-anode PMT. Scintillator strips are grouped in two sets of 32 strips on each side of the PMT and front end electronics board.

The bandwidth of the front end electronics is set to 180 MHz to determine the pulse width. Signal sampling is performed by a Field Programmable Gate Array (FPGA) at 320 MHz. MD scintillator modules receive the trigger signal from their associated SD station. Incoming analog signals from each pixel of the PMT are digitized with a discrim-

inator that provides the input to the FPGA. Samples can be either a logical "1" or "0" depending on whether the incoming signal was above or below a given (programmable) discrimination threshold. This method of *one-bit* resolution is very robust for counting muons in a highly segmented detector [87]. The MD station power is supplied by an additional solar panel and battery box and a dedicated WiFi communication system, see the left panel of Figure 2.14.

Recently the muon component of extensive air showers has been measured between $10^{17.5}$ and 10^{18} eV [88]. This work reports discrepancies between data and simulations on the muon density at the ground. The density of muons for data is larger, around 40% or 50% depending on the hadronic model. These measurements are in agreement with previous measurements of the muon component in inclined air showers [53].

3

The Risetime over Distance

In this chapter we study a novel way to use the data of the Surface Detector (SD) to infer the mass composition of ultra-high energy cosmic rays. We introduce a new observable related to the *risetime* $t_{1/2}$, which we abbreviate by $\overline{\text{ToD}}$ (Average Time over Distance). This observable characterizes each event with a single value: the average value of the risetime divided by the distance to the core r. With the aid of Monte Carlo simulations, a relationship between the mass of the primary cosmic ray and this new observable is established.

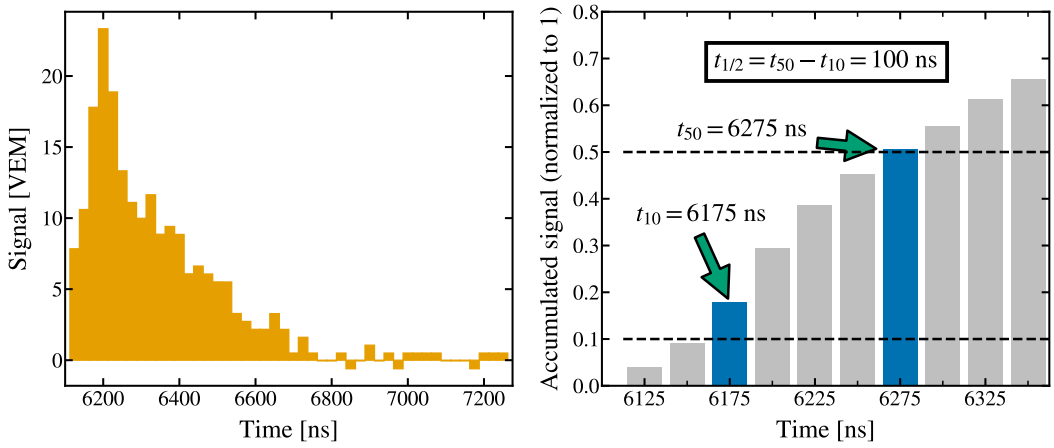

Figure 3.1: Example of the calculation of a risetime. Left: An example of the signals registered by the PMTs of the water-Cherenkov detectors. Right: For each of the first 10 bins of 25 ns, the sum of the signal up to and including that bin, normalized by the sum for all the bins, is represented. The first bin reaching above 10 % of the total signal happens at time $t_{10} = 6175$ ns, while the first bin reaching above 50 % of the total signal happens at $t_{50} = 6275$ ns, giving a risetime $t_{1/2} = 100$ ns.

1 Introduction

The *risetime* $t_{1/2}$ is defined as the time it takes for the measured signal of a triggered station to rise from 10 % to 50 % of the total signal. In Figure 3.1 an example of the computation of a risetime is shown. Each station has 3 PMTs which can measure one trace each. We compute the risetime for each of the traces measured by the operative PMTs. Then, we take the average to obtain a single value of the risetime for each triggered station.

The risetime is a measurement of the spread in the arrival time of the particles produced in a shower. Muons arrive earlier to the stations than particles from the electromagnetic component, so the risetime can give us information about these components. It is well known that it carries information about the composition of the primary cosmic ray [44,89,90]. Furthermore, this physical observable is also interesting because since the duty cycle of the SD is close to 100 %, we can obtain a sizeable amount of statistics even at the highest energies.

This chapter is structured as follows. In Section 2 we study the risetime and its dependence with the distance to the core r to motivate the definition of the $\overline{\text{ToD}}$. We also give a short summary of the $\langle\Delta\rangle$ Method, which is a previous work on the same topic: extract information about mass composition using the information from the risetime. We define the $\overline{\text{ToD}}$ in Section 3. After defining it, in Section 4 we study its dependence with the zenith

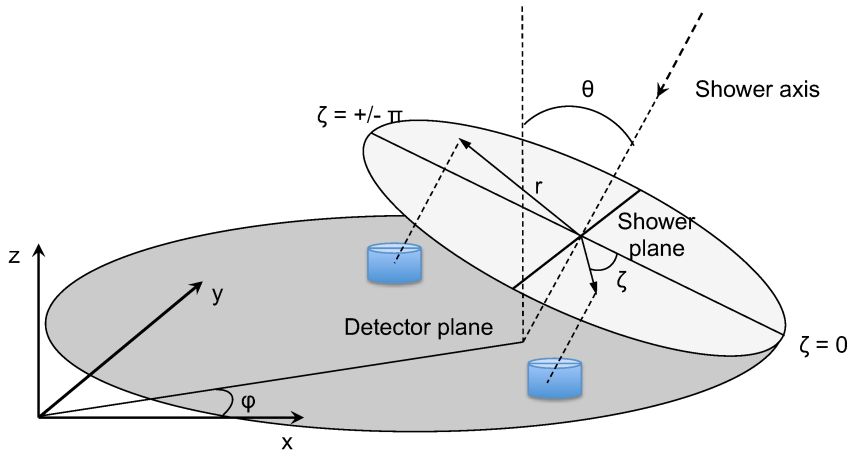

Figure 3.2: Geometry of an event. The angle θ is the angle between the shower axis and the zenith, the distance r is the distance between the shower axis and the projection of the station in the shower plane and the polar angle ζ is the angle between the projection of the station in the shower plane and the projection of the direction of the cosmic ray.

angle θ and compute the evolution of this observable with the reconstructed energy E_{SD} of the primary cosmic ray. We translate the results to $\langle \ln A \rangle$, the average logarithm of the mass number A, to show how the composition inferred from our observable behaves for the two different hadronic models employed for the simulations. We compare our results to those obtained with the $\langle \Delta \rangle$ Method. We give a short summary and the conclusions of this study in Section 5.

2 The risetime

The risetime depends on the distance to the core r from the shower axis to the projection of the station in the shower plane, see Figure 3.2. Stations located further away will measure a wider spread in the times of particle arrival and larger $t_{1/2}$, while stations located close to the core will measure smaller values for $t_{1/2}$.

From Figure 3.3 we can see that the risetime increases with r. This increase is approximately linear up to a distance of about 2000 m for non-saturated stations and about 1200 m for saturated stations[1]. At high distances there are fluctuations for saturated stations, due to low statistics. The amount of signal measured at each station decreases with distance

[1]We do not use stations with the low-gain channel saturated since the risetime can not be reliably computed with those stations. When we refer to saturated stations we refer to stations with the high-gain channel saturated. See page 20 for a description of saturation.

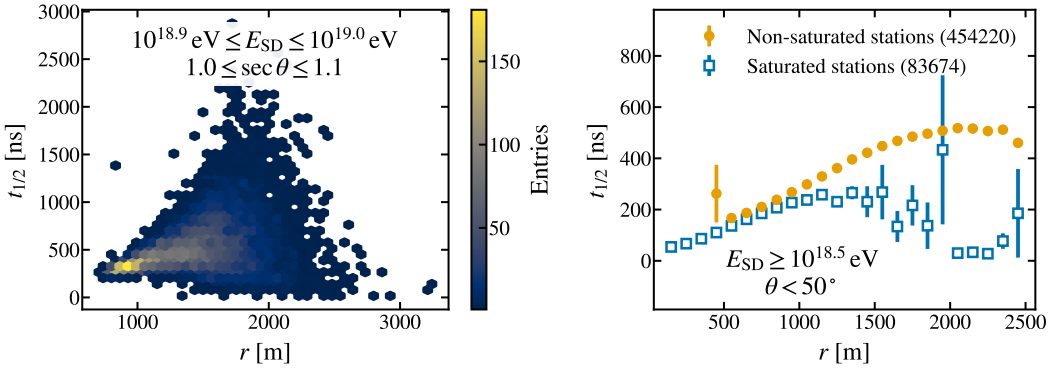

Figure 3.3: Left: Risetime as a function of the distance r for non-saturated stations, reconstructed energy in the range $10^{18.9}$ eV ≤ E_{SD} ≤ $10^{19.0}$ eV and a value of the secant of the zenith angle in the range $1 \leq \sec\theta \leq 1.1$. Right: Mean value of the risetime as a function of the distance r for all the data (all energies from $10^{18.5}$ eV and all zenith angles up to 50 °), separated in saturated and non-saturated stations. The error bars represent the error of the mean and the numbers inside the parenthesis are the total number of stations of each kind.

and saturation is unlikely for smaller signals so that there are few saturated stations in bins of high values of r.

The risetime also depends on the zenith angle θ. This dependence arises from the fact that the muon and electromagnetic components are attenuated differently as the shower develops through the atmosphere: the electromagnetic component is attenuated faster, while muons are very penetrating particles and are less attenuated. The thickness of atmosphere that the shower travels through is proportional to the factor $\sec\theta = 1/\cos\theta$. Then, as θ increases, so does $\sec\theta$ and the contribution from the electromagnetic component, responsible for the spread in time of the signals, decreases for signals measured at the ground. This explains the decrease of the risetime as a function of $\sec\theta$ that can be seen in Figure 3.4.

2.1 Polar angle correction

To be able to compare risetimes from different stations measuring the same shower or event one correction has to be made. The polar angle is the angle ζ around the shower axis as it can be seen in Figure 3.2. A station having $\zeta = 0$ is a station located on top of the projection of the shower axis on the shower plane while a station having $\zeta = \pm\pi$ is a station located in the opposite direction.

In Figure 3.5, the mean value of the risetime as a function of the polar angle has been plotted. The value of the risetime depends strongly on the polar angle. This happens for

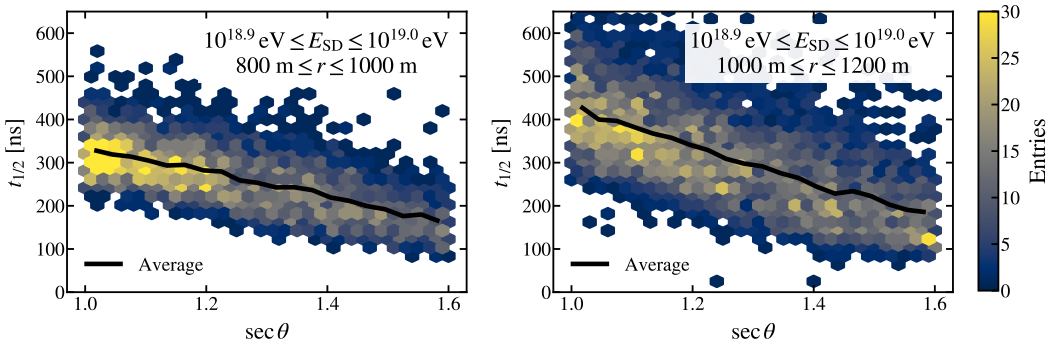

Figure 3.4: Risetimes as a function of the secant of the zenith angle using stations from events with energy in the range $10^{18.9}$ eV ≤ E_{SD} ≤ $10^{19.0}$ eV, non-saturated stations and two different bins for the distance r in meters. The black line is the average of the risetimes. Left: 800 m ≤ r ≤ 1000 m. Right: 1000 m ≤ r ≤ 1200 m.

the same geometrical reason that causes the risetime to decrease with sec θ: the attenuation of the electromagnetic component of the shower is stronger and particles reaching stations with large values of ζ have had to travel a longer path than those stations with smaller values of ζ. There is also another effect related to the geometry, because stations with large ζ measure more muons emitted closer to the shower axis [90].

The risetime as a function of the polar angle can be well described using a cosine function. In what follows, we use the parameterization described in ref. [91]. By picking a certain arbitrary reference angle, which is chosen to be $\zeta = 90°$, we can correct the risetimes using the following expression:

$$t_{1/2}^{\text{corrected}} = t_{1/2}^{\text{measured}} - g(r,\theta)\cos\zeta \tag{3.1}$$

where

$$g(r,\theta) = m(\theta)r^2 \tag{3.2}$$

$$m = (a\sec\theta + b\sec^3\theta + c)\sqrt{\sec\theta - 1} \tag{3.3}$$

and the parameters a, b and c have the following values:

$$a = (-3.9 \pm 2.3)\cdot 10^{-5}\ \text{ns m}^{-2} \tag{3.4}$$

$$b = (-1.9 \pm 0.4)\cdot 10^{-5}\ \text{ns m}^{-2} \tag{3.5}$$

$$c = (2.0 \pm 0.2)\cdot 10^{-4}\ \text{ns m}^{-2} \tag{3.6}$$

After the correction has been applied, the risetime does not depend anymore on ζ, see Figure 3.5.

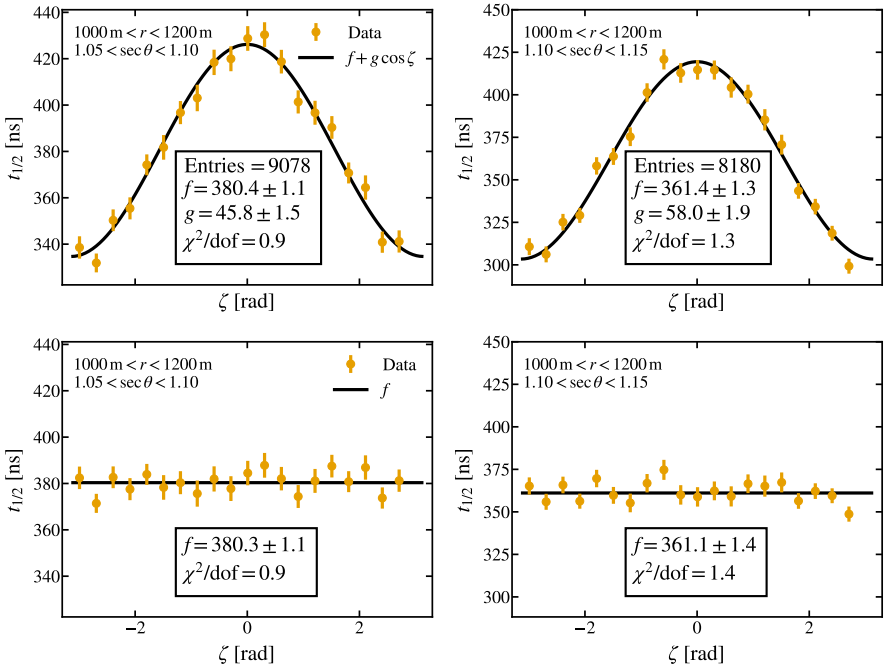

Figure 3.5: Risetime as a function of the polar angle ζ before (top) and after the correction (bottom) using Equation 3.1 for two different bins with 1000 m $< r <$ 1200 m. Left: Bin of secant of the zenith angle $1.05 \leq \sec\theta \leq 1.10$. Right: Bin of secant of the zenith angle $1.10 \leq \sec\theta \leq 1.15$.

We note that the energy does not appear anywhere in this parameterization. However, after a careful study of the polar angle dependence and its relationship with the energy of the cosmic ray we conclude that it is not necessary to include it. There is not a significant gain on how much flatter the distribution of risetimes becomes when the energy dependence is included. Another objection is that the parameters a, b and c do not show a behaviour with the energy but rather they seem to change randomly with it. Based on these two reasons, energy is not taken into account when correcting the polar dependence. **All the following results in this chapter and the next one will have the risetime corrected using Equation 3.1.**

2.2 Risetime uncertainty

There is an uncertainty associated to the measurement of the risetime. This uncertainty has already been studied and parameterized and we will not discuss it in detail. We use the parameterization described in ref. [44]. For each value of the risetime, its uncertainty

will be:

$$\sigma_{1/2} = \sqrt{\left(\frac{J(r,\theta)}{\sqrt{S}}\right)^2 + \left(\sqrt{2}\frac{25}{\sqrt{12}}\right)^2} \tag{3.7}$$

where S is the integral of the trace measured at a station and $J(r,\theta)$ is a linear fit on the distance that depends on two parameters:

$$J(r,\theta) = p_0(\theta) + p_1(\theta)r \tag{3.8}$$

and these parameters $p_0(\theta)$ and $p_1(\theta)$ are given by:

$$\begin{aligned} p_0(\theta) &= \begin{cases} (-340 \pm 30) + (186 \pm 20)\sec\theta & \text{if } r \leq 650\,\text{m} \\ (-447 \pm 30) + (224 \pm 20)\sec\theta & \text{if } r > 650\,\text{m} \end{cases} \\ p_1(\theta) &= \begin{cases} (0.94 \pm 0.03) + (-0.44 \pm 0.01)\sec\theta & \text{if } r \leq 650\,\text{m} \\ (1.12 \pm 0.03) + (-0.51 \pm 0.02)\sec\theta & \text{if } r > 650\,\text{m} \end{cases} \end{aligned} \tag{3.9}$$

2.3 Summary of the $\langle\Delta\rangle$ Method

The $\langle\Delta\rangle$ Method [43,44] is an analysis developed to infer information about the mass composition of cosmic rays from the risetime. The first step of the analysis consists in defining a benchmark. The benchmark is a function that describes the average behaviour of the risetime as a function of the distance to the core and zenith angle for a chosen energy bin. Then, Δ_i is defined for each station:

$$\Delta_i = \frac{t_{1/2} - t_{1/2}^{bench}(r,\theta)}{\sigma_{1/2}} \tag{3.10}$$

where $t_{1/2}$ is the risetime measured at that station, $t_{1/2}^{bench}$ is the value of the benchmark function and $\sigma_{1/2}$ is the associated uncertainty to the measured risetime, see Figure 3.6. $\langle\Delta_s\rangle$ [2] is obtained from the average of all the stations that belong to the same event:

$$\langle\Delta_s\rangle = \frac{1}{N}\sum_{i=1}^{N}\Delta_i = \frac{1}{N}\sum_{i=1}^{N}\frac{t_{1/2} - t_{1/2}^{bench}(r,\theta)}{\sigma_{1/2}} \tag{3.11}$$

With $\langle\Delta_s\rangle$ every event can be characterized with a single number. The evolution of $\langle\Delta_s\rangle$ with the energy is plotted in the left panel of Figure 3.7. This evolution has been

[2] It is defined as $\langle\Delta_s\rangle$ in ref. [43] and as $\langle\Delta\rangle$ in ref. [44].

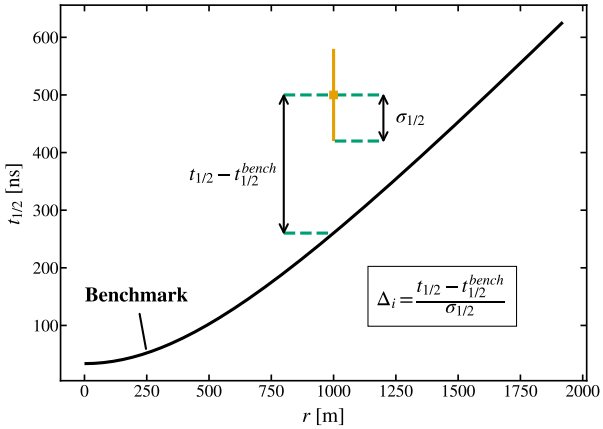

Figure 3.6: Diagram with the computation of Δ_i for a single risetime.

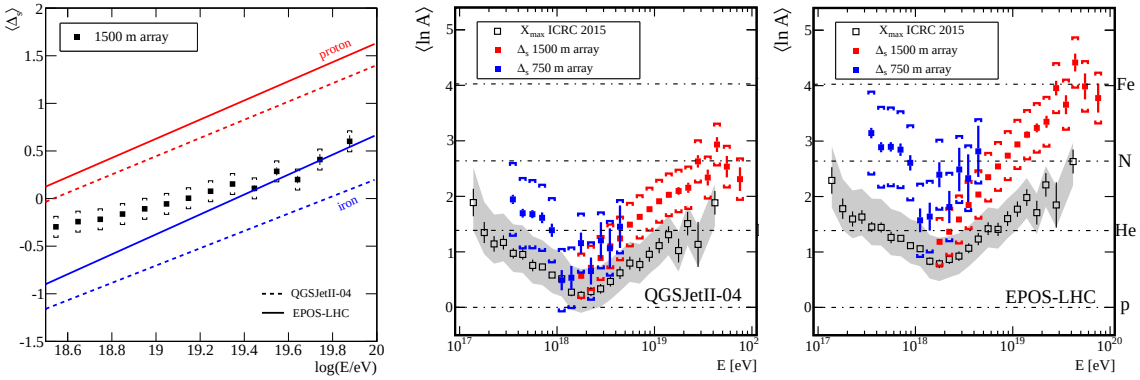

Figure 3.7: Left: Values of the $\langle\Delta_s\rangle$ as a function of the energy for data (black squares) and simulations (lines). Middle and right: Evolution of $\langle \ln A\rangle$ obtained from $\langle\Delta_s\rangle$ with the energy.

transformed to $\langle \ln A\rangle$, the average logarithm of the mass number A. The results for $\langle \ln A\rangle$ are compared to those obtained using the measurements of the depth of shower maximum, $X_{\rm max}$, done with the FD. Both sets of points follow the same trend: mass composition becomes lighter up to $10^{18.1}$ eV and then it becomes heavier for higher energies. There is a difference between the values obtained from $X_{\rm max}$ and $\langle\Delta_s\rangle$ that can be attributed to the fact that the FD only sees the electromagnetic component of the shower for the measurement of $X_{\rm max}$ while $\langle\Delta_s\rangle$ has contributions from both the electromagnetic and muonic components of the shower. There are several results by The Pierre Auger Observatory that hint at an incorrect modelling of the muonic component of the shower by the hadronic interaction models [52,53,88]. In these results there is a deficit of muons in simulations. This causes that when comparing data recorded with the SD and FD detectors, data of the surface detectors seems to favour heavier compositions (as can be seen in Figure 3.7).

The $\langle\Delta\rangle$ Method has played an important role to motivate the risetime over distance. For the analysis with the $\langle\Delta_s\rangle$, it is necessary to define and study a benchmark. This process is very involved: the dependence of the risetime on r and $\sec\theta$ has to be studied thoroughly and the behaviours are different for saturated and non-saturated stations, so two different benchmarks are needed. It is, however, a very powerful way of obtaining physical information since there is one value for each event. With $\langle\Delta_s\rangle$, event-by-event studies can be done instead of average studies, where a physical quantity is obtained by an average in a sample in which many events are included. For example, $\langle\Delta_s\rangle$ can be calibrated using the values of $X_{\max}$ for hybrid events and then a value of $X_{\max}$ can be given to each event measured by the SD. The risetime over distance arises naturally from the next question: can we have a simpler physical observable that uses the information from the risetime and characterizes each event with a single value?

3 The risetime over distance

In this section the new observable that we have built is introduced and studied: *The average risetime over distance* that we will abbreviate by $\overline{\text{ToD}}$ (Time over Distance), where a bar has been put to indicate that it is an average value. The $\overline{\text{ToD}}$ is defined as the average of the values of the risetime $t_{1/2}$ measured at each individual station divided by the distance r from each station to the shower axis of the event:

$$\overline{\text{ToD}} = \left\langle \frac{t_{1/2}}{r} \right\rangle = \frac{1}{n}\sum_{i=1}^{n} \frac{t_{1/2_i}}{r_i} \tag{3.12}$$

where the sum runs over all the n selected stations of an event. It is motivated by the approximate linear dependence of $t_{1/2}$ with r.

The uncertainty of $t_{1/2}/r$ is obtained through standard error propagation:

$$\begin{aligned}\Delta\left(\frac{t_{1/2}}{r}\right) &= \sqrt{\left(\frac{\partial\frac{t_{1/2}}{r}}{\partial t_{1/2}}\right)^2 \left(\Delta t_{1/2}\right)^2 + \left(\frac{\partial\frac{t_{1/2}}{r}}{\partial r}\right)^2 (\Delta r)^2} \\ &= \sqrt{\left(\frac{1}{r}\right)^2 \left(\Delta t_{1/2}\right)^2 + \left(-\frac{t_{1/2}}{r^2}\right)^2 (\Delta r)^2}\end{aligned} \tag{3.13}$$

The main contribution comes from the first term in the square root, since the uncertainty of the distance to the core Δr is usually very small (lower than 3 %) compared to r itself.

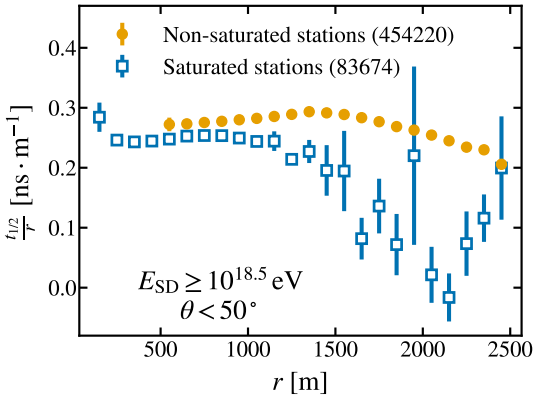

Figure 3.8: Risetime over distance as a function of the distance r computed for each station and grouped in bins of distance r. All the data (all energies from $10^{18.5}$ eV and all zenith angles up to 50 °) has been included and the numbers inside the parenthesis are the total number of stations of each kind.

Since the $\overline{\text{ToD}}$ is a mean, the uncertainty on the $\overline{\text{ToD}}$ is obtained from the standard formula for the uncertainty of the mean, modified to take into account the uncertainty of each term:

$$\Delta\overline{\text{ToD}} = \frac{1}{n}\sqrt{\sum_{i=1}^{n}\left[\Delta\left(\frac{t_{1/2_i}}{r}\right)\right]^2} \tag{3.14}$$

In Figure 3.8, the risetime divided by the distance has been plotted as a function of the distance[3].

3.1 Data selection

In this analysis we have used data from March 2004 to June 2017, and zenith angles up to 50 °. Above this angle, the behaviour of the $\overline{\text{ToD}}$ as a function of $\sec\theta$ is no longer linear.

We use the following cuts:

- We do not use saturated stations in our analysis. For those stations that have the high-gain channel saturated in any of the three photomultipliers the traces are computed using the low gain channel and have worse resolution, see SD signal saturation on page 20. The behaviour of the risetimes obtained from saturated and non-saturated stations is different, as it can be seen in Figure 3.3 and Figure 3.8.

[3]Note that this is not the $\overline{\text{ToD}}$ (one value for each event) but is the risetime measured at each station divided by the distance to the core of each station.

- We use those stations that measure a value of the risetime $t_{1/2} > 40\,\text{ns}$. The polar correction causes a few values of the risetime to be below zero, which does not make sense and it is artificially caused by the correction. This usually happens for large values of r where the signals measured are very low and these stations would not pass the cut on the signal. The cut on 40 ns is due to 40 ns being the minimum risetime that can be measured with the electronics of a station. This risetime is obtained when a vertical muon passes through a station. There are also a few risetimes above 3000 ns that have been checked and they are discarded since those are cases with anomalous traces and, in some cases, the photomultipliers have been flagged as defective [43].

- We use events with a reconstructed energy E_{SD} equal or greater than $10^{18.5}$ eV. Above this energy the 1500 m array is fully efficient, that is, the probability that a shower triggers the 1500 m array of the SD is 100 %.

- We use only stations that measure a total signal $S > 5$ VEM because the behaviour of the risetime divided by the distance is different for very low signals than for higher signals. We elaborate more on this below.

- After applying the previous cuts, events will be required to have, at least, 2 stations that survive the cuts discussed before. In this way, we are using at least 2 values to compute the $\overline{\text{ToD}}$. With this cut we avoid computing the $\overline{\text{ToD}}$, which is an average, with a single value of the risetime divided by the distance.

In the following subsections, the cut on the signal and the selection efficiency are discussed.

The signal cut

When studying the risetime over distance as a function of the signal S, we observed a peculiar behaviour for low values of S. As it is shown in Figure 3.9, the average value of $t_{1/2}/r$ for the lowest signals is smaller than for larger signals. This happens because the risetime stops increasing linearly with distance at high distances, which is equivalent to low signals. For this reason, it was decided not to include in the analysis those stations for which the value of the total signal measured is below a threshold S_0. Now, we explain how we obtained the threshold $S_0 = 5$ VEM for the cut $S > S_0$.

To find S_0 the average of $t_{1/2}/r$ was characterized by fitting a constant function $f(r) = a$ to the values of the average $t_{1/2}/r$ with $S > 10$ VEM, in the region where its behaviour is almost constant with the signal, see Figure 3.9. Then, the value of the signal S where the average $t_{1/2}/r$ began being very close (with a difference of less than 0.02 VEM, although any small value gives the same result) to the value obtained with the fit was considered as the analysis threshold. It was also studied whether the cut should depend on the energy.

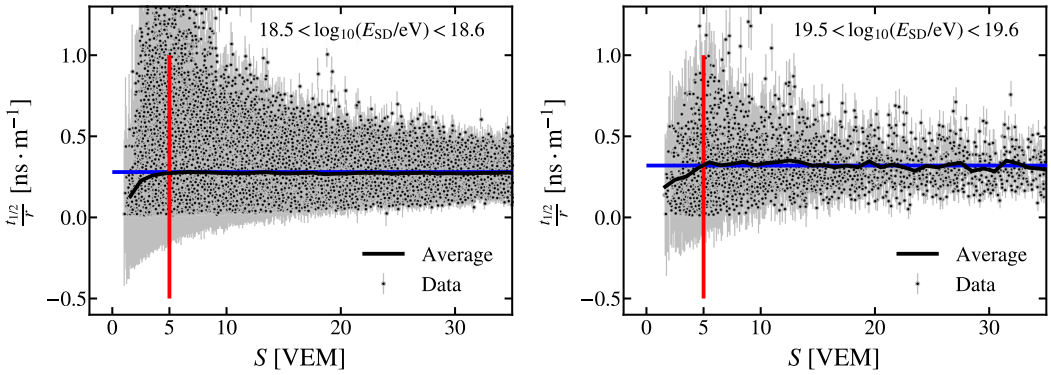

Figure 3.9: Risetime over distance as a function of the signal S for two energy bins. The black line is the average of the risetime over distance and the blue horizontal line is a fit of the average value of $t_{1/2}/r$ for $S > 10$ VEM. The red vertical line is the value for which the average is close to the fit for $S > 10$ VEM.

We made bins of energy of 0.1 width in $\log_{10} E$. Then, for each energy bin we repeated the process described to obtain the threshold. We found $S_0 = 5$ VEM as the valid threshold for all the energy bins.

Selection efficiency

As it can be seen in Table 3.1, the selection efficiencies are very close for data and simulations and also for different primaries at the lowest energies, while at the highest energies they are 100 %. This is important to ensure that the cuts do not have any preference towards a certain composition. Full efficiency at large energies happens because for those events the footprint is larger and therefore events trigger many stations. For these events our cuts do not remove enough stations and most of the events do not have a number of stations n lower than 2.

In Table 3.2 the selection efficiency is broken down for the cuts done. Without the saturated stations and with energy E above $10^{18.5}$ eV, there are a total of 454220 stations distributed in 91991 events. Data measured during bad periods (inoperative detectors, storms, lightning, etc.) have been excluded. Only 6T5 events have been selected.

4 Results

We have studied the dependence of the $\overline{\text{ToD}}$ with the zenith angle θ. As it can be seen in Figure 3.10, it follows a linear relationship. The values of the $\overline{\text{ToD}}$ decrease with $\sec\theta$

$\log_{10}(E/\text{eV})$	% of events selected				
	Exp. data	Proton (QGSJetII-04)	Iron (QGSJetII-04)	Proton (EPOS-LHC)	Iron (EPOS-LHC)
18.5 - 18.6	98.4	97.6	97.8	98.2	98.2
18.8 - 18.9	99.7	99.7	99.6	99.7	99.7
19.0 - 19.1	99.8	99.9	99.9	99.8	99.9
19.4 - 19.5	100	100	100	100	100

Table 3.1: Percentage of the initial events remaining after the cuts used for different energies.

	stations	events with $n \geq 2$
Initial sample	454220 (100%)	91991 (100%)
40 ns $< t_{1/2} <$ 3000 ns	447131 (98.4%)	91976(99.9%)
$S > 5$ VEM	353671 (78%)	91914 (99.9%)
$n \geq 2$	352953 (78%)	91196 (99%)

Table 3.2: Selection efficiency for data on the number of stations n and on the number of events with the cuts used for the data measured by the SD.

because the risetimes also decrease as it was shown in Figure 3.4. In view of the relationship between the two previous parameters, we have done a linear fit in each energy bin to characterize the evolution of the mean value of the $\overline{\text{ToD}}$ as a function of $\sec\theta$, see Figure 3.10. The evolution with the energy of the free parameters of the fits, such as the slope or the intercept, has been studied and it has been found that the most stable one is to pick a value of the fit for a certain reference angle.

The linear fits can be expressed as $f(\sec\theta) = a + b\sec\theta$ with free parameters a and b; then, the reference value will be $\xi = f(\sec\theta_0) = a + b\sec\theta_0$. We will use $\theta_0 = 30°$ since it is a very close value to the median of all the values of θ of the data selected. Our results do not depend heavily on the value of this reference angle. The evolution with the energy of ξ is shown in Figure 3.11. Simulations have been fitted with a straight line and only this line is shown. The computation of systematic uncertainties is discussed with detail in the next sections. In Figure 3.11 there is a trend of the data to move towards heavier composition as the energy increases. Then, around $10^{19.6}$ eV, the data seems to start moving towards a lighter composition.

The evolution with the energy shown in Figure 3.11 has been studied. If a straight line is fitted to the data points the goodness of the fit obtained is $\chi^2/\text{dof} = 1.95$. Using the maximum likelihood ratio test, we found that it is not significant to fit the elongation rate with a second degree polynomial when compared to a fit with a straight line. With a second degre polynomial one extra parameter is introduced and $\chi^2/\text{dof} = 2.1$. Another possibility is using two straight lines that intersect at a certain energy E_0. In that case, the goodness of the fit obtained is $\chi^2/\text{dof} = 1.94$ and again the maximum likelihood ratio test

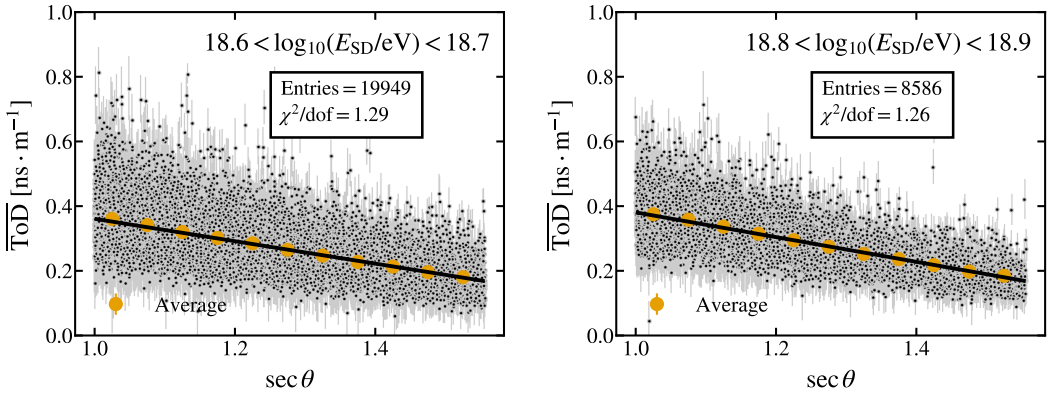

Figure 3.10: Values of $\overline{\text{ToD}}$ as a function of $\sec\theta$ for two energy bins. The average values have been plotted for several bins of $\sec\theta$. In Figure 1 on page 143 more energy bins are shown.

shows that this is not significant with a p-value of $p = 0.135$ and $E_0 = 10^{19.78}$ eV. When one standard deviation is subtracted to the value of the elongation rate of the last bin of energy, the p-value obtained is $p = 0.6$, and $E_0 = 10^{19.79}$ eV; most of the significance for a fit with two straight lines is coming from the last point.

4.1 Systematic uncertainties

The next step is to obtain the systematic uncertainties, that have been classified as follows:

- **Ageing of the detectors**. In the left panel of Figure 3.12, the $\overline{\text{ToD}}$ has been plotted as a function of the year. The Observatory was completely installed in 2008 which could explain the increase from 2004 to 2008. From 2008 onwards, there is a steady decrease. A linear fit has been done, which shows that the variation in 10 years (2008 to 2018) is approximately $-0.01\,\text{ns}\,\text{m}^{-1}$. The contribution to the systematic uncertainty has been obtained by taking half of the total change for both the positive and negative shift: $0.005\,\text{ns}\,\text{m}^{-1}$.

  A study of the dependence of the risetime with the year was carried out to establish where the differences come from. It was found that many variables depend on the year and that there is an effect of ageing for these variables too. For example, the length of the trace used to compute the risetime, the risetime and the falltime (time for the signal to rise from a 50 % to a 90 % of the total signal) decrease with the years. The behaviour found on these variables was a linear decrease with time.

  It was also found that the behaviour of the risetime seems to be strongly correlated to that of the Area over Peak (AoP) defined as the total charge divided by the largest

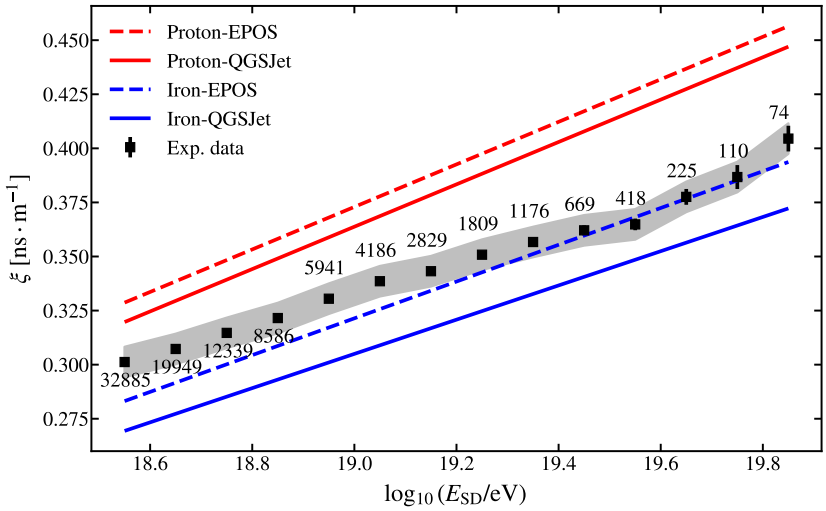

Figure 3.11: Evolution with the energy of ξ. The shadowed area corresponds to the systematic uncertainty while the error bars are the statistical uncertainties, that is, the uncertainty of the values obtained from the linear fits of ξ as a function of $\sec\theta$. All events with $E_{SD} > 10^{19.8}$ eV have been used to compute the point with the highest energy.

value in the signal of the PMTs. When studying the AoP as a function of time, the AoP can be seen to have a trend which is similar to the one found for the risetime [92].

- **The seasonal effect**. A dependence with the seasons of the year was studied. It was found that there is a periodicity on the value of the $\overline{\text{ToD}}$. A fit of a sine was done and the amplitude found to be 0.002 ns m^{-1} so the systematic contribution is taken to be 0.001 ns m^{-1} for both the positive and negative shifts, see Figure 3.13.

- **Dependence on the atmospheric pressure, temperature and humidity**. The $\overline{\text{ToD}}$ depends on the atmospheric conditions. In particular we found that there is a dependence with pressure and temperature. These variables, however, are very correlated between themselves and also with the season, see Figure 3.13. Because a systematic contribution from the effect of the seasons has been included already, it is not necessary to include an additional contribution from these variables.

- **Energy uncertainty**. The reconstruction of the absolute energy of the cosmic ray has an uncertainty which is not worse than a 14%, and improves as the energy increases. In the right panel of Figure 3.12, the evolution with the energy has been obtained when the energy of the experimental data is shifted by $\pm 14\,\%$. The difference between the shifted values and the unshifted is, at most, 0.005 ns m^{-1}. For the bin from $10^{19.6}$ eV to $10^{19.7}$ eV and shifting the energy upwards this difference

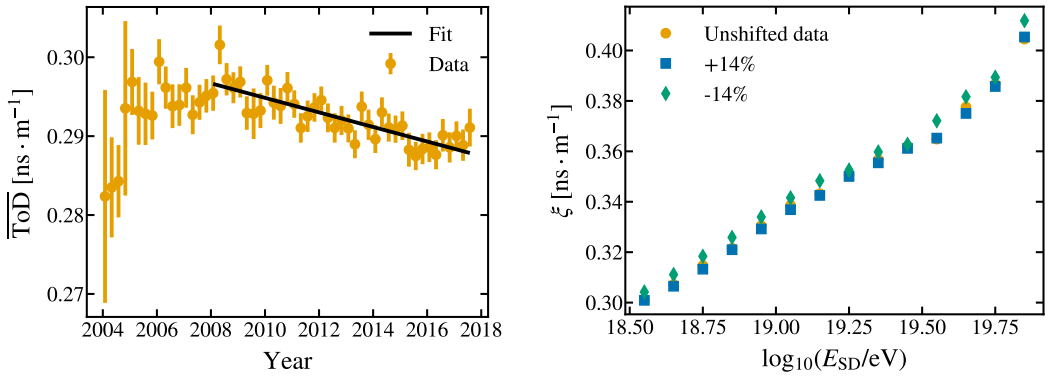

Figure 3.12: Left: Plot of the average values of $\overline{\text{ToD}}$ as a function of the year. The error bars are the error of the mean of each point. The black line is a fit for data measured from 2008 onwards. Right: Evolution with the energy of ξ computed by shifting the energy E of all the events by $\pm 14\%$.

Ageing of the detectors	± 0.005 ns·m^{-1}
Seasonal effect	± 0.001 ns·m^{-1}
Energy uncertainty	± 0.005 ns·m^{-1}
Total systematic uncertainty	± 0.007 ns·m^{-1}

Table 3.3: Summary of the systematic uncertainties and final computation for the value of $\Delta\xi_{\text{syst}}$, obtained as the sum in quadrature of the systematic uncertainties.

is larger than 0.005 ns m but the energy resolution improves with the energy so the real shift with the correct energy resolution is lower.

The most relevant sources of systematic uncertainties, listed in Table 3.3, are added in quadrature giving a final value of ± 0.007 ns·m^{-1}. This value amounts to approximately 23% and 26% of the separation between p and Fe for QGSJetII-04 and EPOS-LHC, respectively.

4.2 Average logarithm of the mass

Figure 3.11 can be translated to a plot where the change of the composition with energy obtained with this analysis can be seen more clearly. The superposition principle allows us to assume a logarithmic dependence as it happens with other observables such as X_{max} [93]. The average logarithm of the mass will be:

$$\langle \ln A \rangle = \ln 56 \cdot \alpha = \ln 56 \cdot \frac{P - D}{P - I} \tag{3.15}$$

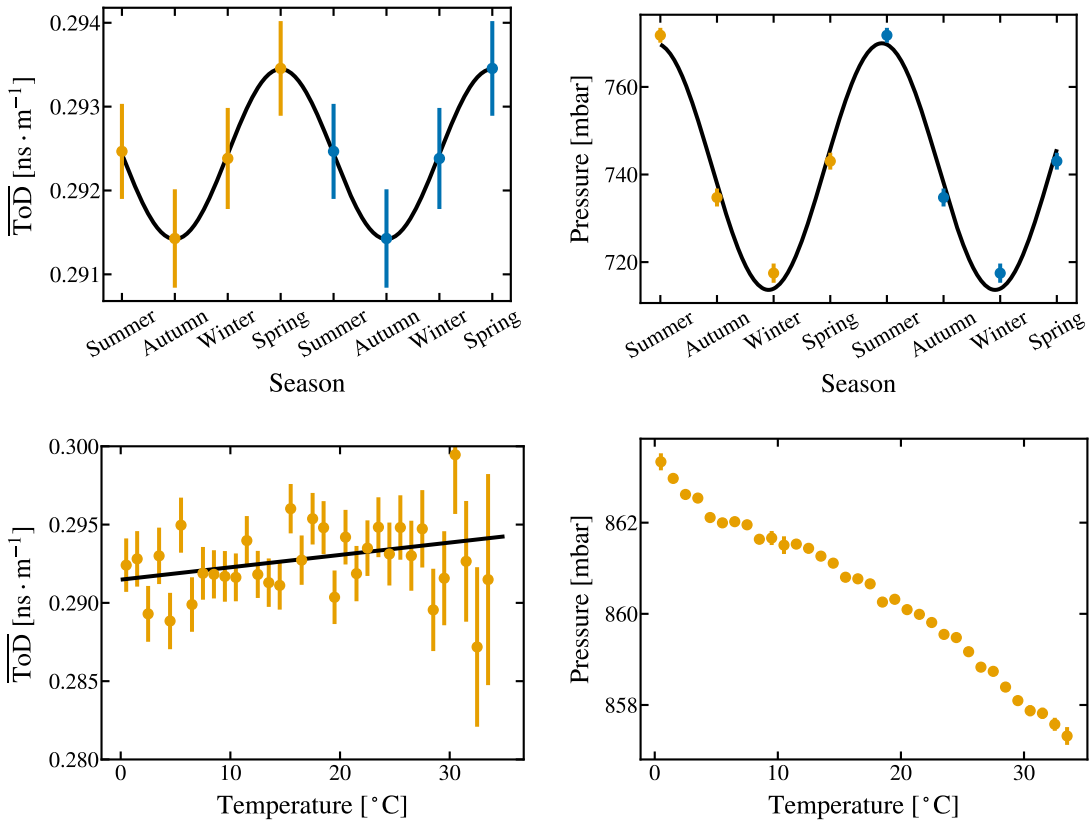

Figure 3.13: Top left: $\overline{\text{ToD}}$ as a function of the season for all the data. The points have been repeated so that periodicity can be noticed better. Top right: pressure as a function of the season. Bottom left: $\overline{\text{ToD}}$ as a function of the temperature. Bottom right: pressure as a function of the temperature.

where 56 corresponds to the number of nucleons in an iron nucleus, P to the values obtained from proton simulations, I to the values obtained from iron simulations and D to those obtained from measured data. This is the equivalent to assign a continuous change in composition from the line for proton to the line for iron and choosing a certain composition based on where the point for data lies between these lines. Because we have different values for simulations depending on the hadronic model employed, we have different interpretations for the composition. The uncertainty of Equation 3.15 is computed through standard quadratic propagation of the uncertainties of the values for the proton, iron and data points.

In Figure 3.14 the evolution of the composition with energy is shown using the two hadronic models that were employed in the simulations. We can see that the average masses are different depending on the model used to infer them and the trend of increasing mass

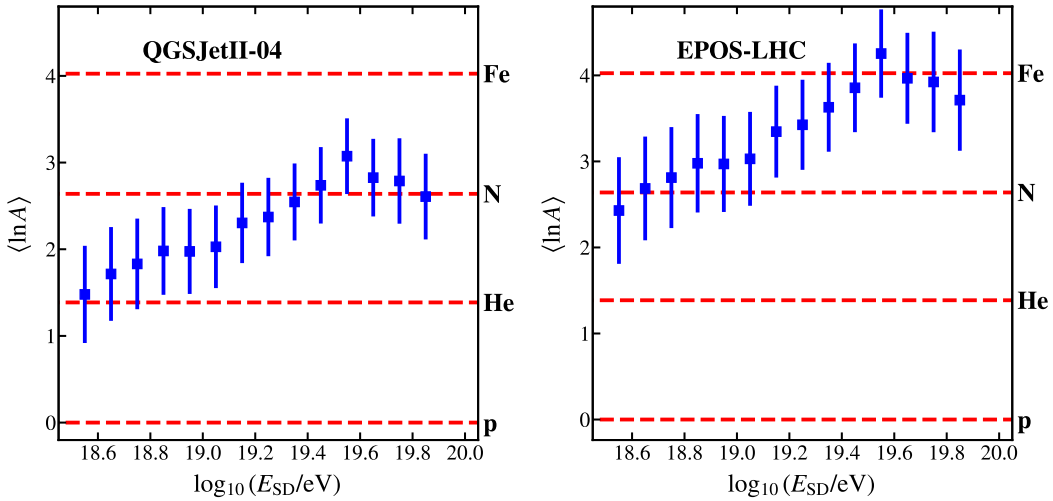

Figure 3.14: Values of $\langle \ln A \rangle$ computed using Equation 3.15 and the results obtained in Figure 3.11. The statistical and systematic uncertainties have been added in quadrature and propagated. Left: Proton and iron values are taken from the results for the hadronic model QGSJetII-04. Right: Proton and iron values are taken from the results for the hadronic model EPOS-LHC.

until $10^{19.6}$ eV.

4.3 Comparison to the $\langle \Delta \rangle$ Method

Our results have been compared to those obtained using the $\langle \Delta \rangle$ Method [44]. In Figure 3.15 both results are plotted as $\langle \ln A \rangle$. All the bins have a very similar value and both our results are perfectly compatible within the one σ level. The differences can be explained by a combination of different factors: the $\overline{\text{ToD}}$ is a variable that shares some information with $\langle \Delta_s \rangle$ but it still is a different one, the data that we use includes more recent years (where ageing contributes more, shifting the risetimes to lower values or heavier composition) and the simulations used are generated and reconstructed with newer versions of the software used.

5 Summary and conclusions

In this chapter we have introduced a new observable: the average risetime over distance abbreviated by $\overline{\text{ToD}}$ from Time over Distance. After applying the quality cuts for data

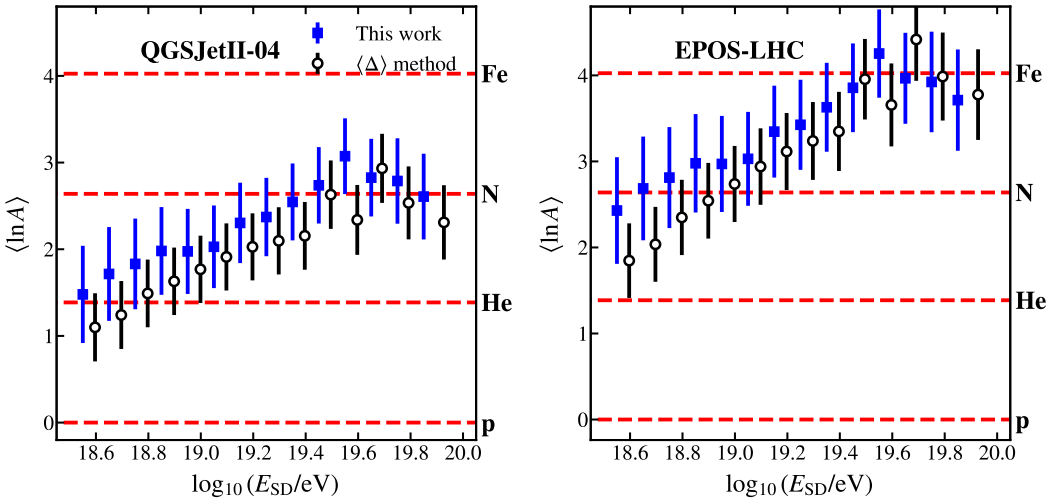

Figure 3.15: Comparison of the results obtained in this work (squares) with the results obtained in [44] (circles). The circles have been shifted half of the bin in energy to the right so that they can be compared to the results obtained in this work. Uncertainties are the obtained by adding in quadrature the systematic and statistical contributions.

and simulations, we have studied the dependence of the $\overline{\text{ToD}}$ with the zenith angle θ. We have found a linear relationship, fitted the $\overline{\text{ToD}}$ as a linear function of $\sec\theta$ and picked a reference angle to obtain one value for each bin of energy. We have called this value ξ.

As our final plot, which is Figure 3.11, we have obtained the evolution with energy of ξ with both its systematic and statistical uncertainties. This plot tells us that regardless of the hadronic model used, the mass of the primary cosmic ray grows with the energy up to a certain energy of $\sim 10^{19.6}$ eV. We have seen that a fit with two lines has a high p-value. This p-value is very dependent on the value of the last bin of energy, as we have seen by subtracting one standard deviation and recomputing the p-value.

With this analysis we can explore the ultra-high energy region with good statistics. The last energy bin includes 74 events above $10^{19.8}$ eV (63 EeV). We have 24 events above $10^{19.9}$ eV ($\sim$ 80 EeV).

Regarding the systematic uncertainties, we have seen how the time evolution of both the detectors and the atmosphere plays a role in determining the systematic uncertainties. The absolute energy uncertainty is also one of the most important contributions to the systematic uncertainty.

This work was published as an internal note of the Pierre Auger Collaboration [94].

4

Extensive Air Shower Fluctuations

The study of shower-to-shower fluctuations is useful in efforts to determine the mass composition. Using an observable that has already been studied, the $\overline{\text{ToD}}$, we provide two methods to measure the fluctuations, obtaining results that are in very good agreement. We observed that the detector resolution is much larger than the effects due to the fluctuations associated with the physics and the composition of the primary flux but nonetheless we are able to demonstrate that the shower-to-shower fluctuations show a dependency with the energy. Large uncertainties caused by a lack of statistics due to strong cuts and due to the small area of the water-Cherenkov detectors prevent us from making strong claims.

1 Introduction

The fluctuations that a physical observable exhibits are a convolution of two effects. One of them is caused by the sampling and conditions of the detector. The other has physical information and we call it shower-to-shower fluctuations[1]. These intrinsic fluctuations are therefore the differences found in an observable due to physics only. Both effects are entangled following the next equation:

$$\sigma^2_{\text{total}} = \sigma^2_{\text{det}} + \sigma^2_{\text{f}} \tag{4.1}$$

where we denote by σ^2_{total} the total variance of an observable, σ^2_{det} is the contribution due to the detector and σ^2_{f} are the fluctuations due to physics and a possible spread in mass of the events in a sample.

The determination of σ^2_{f} is a valuable tool when doing composition analysis, for the fluctuations induced by light or heavy nuclei are different, see the right panel of Figure 4.1 [40]. In particular, the distributions of X_{max} are predicted to be wider in showers initiated by protons than in showers initiated by iron nuclei. There have been several studies of the shower-to-shower fluctuations [95] and some of them also use the information from the Surface Detector of different experiments [96,97].

This study provides two methods for measuring the fluctuations σ^2_{f} using the $\overline{\text{ToD}}$. One method is based on comparing values of the same observable when it is computed in different subdivisions of the same event. The other one is based on the Analysis of Variance (ANOVA).

We follow the analysis done in Chapter 3 where we study the risetime over distance or $\overline{\text{ToD}}$, defined as the average of all the risetimes divided by the distance to the core over each station of an event:

$$\overline{\text{ToD}} = \left\langle \frac{t_{1/2}}{r} \right\rangle = \frac{1}{n} \sum_{i=1}^{n} \frac{t_{1/2_i}}{r_i} \tag{3.12 revisited}$$

where, as reminder, n is the number of stations in a certain event, $t_{1/2_i}$ is the risetime measured in the i-th station and r_i is the distance to the core of that station. With this observable we estimate σ^2_{det} and from Equation 4.1 we compute the fluctuations subtracting from the total variance of our observable:

$$\sigma^2_{\text{f}} = \sigma^2_{\text{total}} - \sigma^2_{\text{det}} \tag{4.2}$$

[1] Based on the results of the previous chapter, since the composition is not pure σ^2_{f} also contains contributions due to a spread in primary masses.

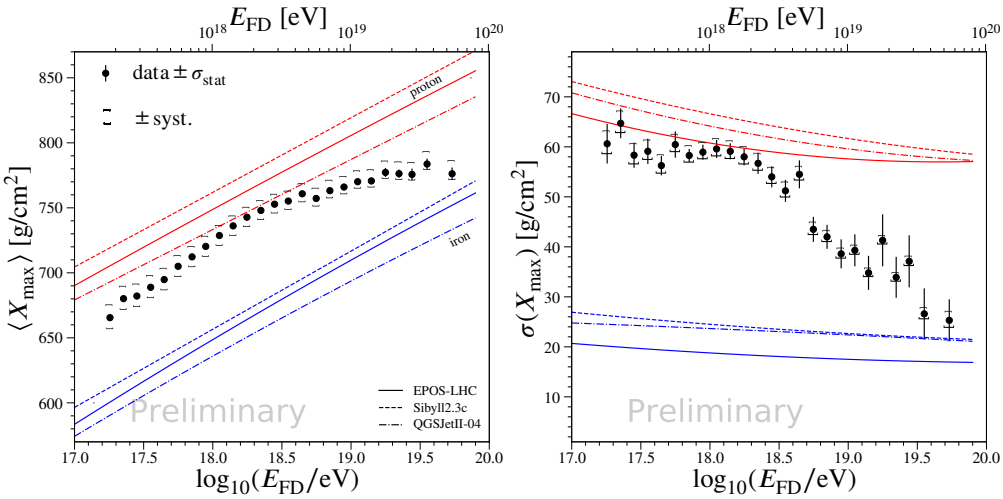

Figure 4.1: Left: Average value of the distribution of measured X_{max} for data and simulations. Right: Standard deviation of the distribution of the measured X_{max}.

This chapter is structured as follows. In Section 2 we present the methods employed and the selection cuts and efficiency for data and simulations. In Section 3 the results obtained with the two methods are shown, and later combined and compared. The chapter ends with a short summary and the conclusions of this study in Section 4.

2 Methodology

In this section, the two methods employed are explained in detail. Afterwards, the quality cuts applied and the selection efficiency in data and simulations are broken down. The results obtained with these methods are explained in the next section.

2.1 The method of splitting

The main idea is to split each event in two subdivisions and compute the $\overline{\text{ToD}}$ in each of the groups separately. The difference between the two values obtained, that we denote by $\overline{\text{ToD}}_1$ and $\overline{\text{ToD}}_2$, can only be due to the resolution of the detector. $\overline{\text{ToD}}_1$ and $\overline{\text{ToD}}_2$ are measurements of the same physical observable within the same shower, so their values should be the same up to the smearing caused by the resolution of the detector.

More precisely, an event with n stations after the cuts have been applied is divided in some way, that we will explain in detail later, in two groups with n_1 and n_2 stations respectively such that $n = n_1 + n_2$. We keep the same definition for the Time over Distance for each subdivision:

$$\overline{\mathrm{ToD}}_1 = \frac{1}{n_1}\sum_{i=1}^{n_1}\frac{t_{1/2_i}}{r_i} \quad \text{and} \quad \overline{\mathrm{ToD}}_2 = \frac{1}{n_2}\sum_{j=1}^{n_2}\frac{t_{1/2_j}}{r_j} \tag{4.3}$$

Then, we compute the resolution of the detector from these two values using the following formula:

$$\sigma^2_{\mathrm{det}} = \sigma^2\left(\frac{\overline{\mathrm{ToD}}_1 - \overline{\mathrm{ToD}}_2}{2}\right) \tag{4.4}$$

where $\sigma^2(x)$ is the variance of the distribution of x and the factor 2 in the denominator is a statistical factor that normalizes σ^2_{det}. This is necessary because the number of stations with which the measurement of the $\overline{\mathrm{ToD}}$ is done is different from the number of stations used to measure $\overline{\mathrm{ToD}}_1$ and $\overline{\mathrm{ToD}}_2$. Without this factor we would see differences that are not only caused by physics but also by the difference in the number of stations.

We explain now the scheme chosen to divide the stations of each event in two groups to compute $\overline{\mathrm{ToD}}_1$ and $\overline{\mathrm{ToD}}_2$. The method that we have chosen is to divide based on the total signal measured at each station. The stations are sorted from largest total signal to smallest total signal. Then, the odd stations in this list are in the first group and the even ones are in the second. That is, the first station, the one with the highest signal in the event, would be in the first group, the second station would be in the second group, the third would be again in the first group and so on.

Another possible choice is to pick as the first group the stations that have a positive polar angle $\zeta > 0$ and for the other group those having a negative polar angle $\zeta < 0$. This choice has been explored before [96] but we have used it only to test that our results do not depend on the scheme chosen to define the stations set.

Regarding the uncertainties, the statistical uncertainties are taken from a fit of a Gaussian function to the distribution of the differences of $\overline{\mathrm{ToD}}_1$ and $\overline{\mathrm{ToD}}_2$ (with the factor 2 in the denominator). It will be seen later that these distributions have a gaussian shape, centred close to zero. The uncertainty of σ^2_{det} is taken from the uncertainty of the parameters of the fit:

$$f(x, a, \mu, \sigma^2) = a\, e^{-\frac{(x-\mu)^2}{2\sigma^2}} \tag{4.5}$$

where a, μ and σ^2 are free parameters. σ^2 is used as a parameter instead of σ so that the uncertainty of σ^2 can be obtained directly from the uncertainty of the parameters of the fit.

In principle, the uncertainty of σ^2_{det} could also be obtained by a simple quadratic propagation of the uncertainties from the risetime since we are working with analytic expressions. However, uncertainties obtained this way are overestimated because the assumptions of independence of the terms that are involved when propagating are not valid: $\overline{\text{ToD}}_1$ and $\overline{\text{ToD}}_2$ are not independent since they are computed from risetimes of the same event. It would be necessary to do an involved study of how the uncertainties are correlated while the procedure that we follow only requires doing a fit. The computation of systematic uncertainties is explained later, on page 69.

2.2 Analysis of variance (ANOVA)

The analysis of variance (ANOVA) is a technique for the study of deviations of a population from a mean value based on a division on groups of this population. This method can be applied to any dataset that can be divided in groups, although we give a version of this method adapted to the problem that we are dealing with. ANOVA has been applied to the study of fluctuations using risetimes before [97].

In this work our population is the values of the risetime divided by the distance $t_{1/2}/r$ and the groups are the events. First, we define some notation to explain the ANOVA method. x is a vector of numbers and in our case it is a vector with all the values of the risetime divided by the distance to the core. $\langle x \rangle$ is the average value of x. We denote each event with the index g. $\langle x^g \rangle$ is the mean for the group or event g and that is the definition of the $\overline{\text{ToD}}$, because we are using $x = t_{1/2}/r$. Each event or group has n_g stations. And the j-th value of $\frac{t_{1/2}}{r}$ in an event is denoted by x^g_j. ANOVA starts from the following general equality, valid for any definition of groups:

$$\underbrace{\sum_i \left(x_i - \langle x \rangle\right)^2}_{\text{total}} = \underbrace{\sum_g n_g \Big(\langle x^g \rangle - \langle x \rangle\Big)^2}_{\text{between groups}} + \underbrace{\sum_g \sum_{j \in g} \Big(x^g_j - \langle x^g \rangle\Big)^2}_{\text{within groups}} \tag{4.6}$$

In ANOVA, the null hypothesis is that all the groups are sampled randomly from the population. Statistical significance for the hypothesis that groups are not sampled randomly is obtained via a F-test of the following quantity:

$$F = \frac{\sigma_1^2}{\sigma_2^2} \quad \text{with} \quad \sigma_1^2 = \frac{\sum_g n_g \Big(\langle x^g \rangle - \langle x \rangle\Big)^2}{N_g - 1} \quad \text{and} \quad \sigma_2^2 = \frac{\sum_g \sum_{j \in g} \Big(x^g_j - \langle x^g \rangle\Big)^2}{N - N_g} \tag{4.7}$$

with degrees of freedom $\nu_1 = N_g - 1$, where N_g is the number of groups, and $\nu_2 = N - N_g$, where N is the total number of samples. After the statistical significance is computed, the null hypothesis can be accepted or rejected.

We do not follow the standard ANOVA but we identify σ_2^2 with σ_{det}^2. This is motivated by the fact that Equation 4.6 is very similar to Equation 4.1, where we have that the total variance is equal to the sum of two components. In Equation 4.6, one term takes into account the variations between different groups or events while the other takes into account the variations within the same group. This is similar to what was done in [97] although here the fluctuations are computed in a different way, using Equation 4.2. We obtain σ_{det}^2 from:

$$\sigma_{\text{det}}^2 = \frac{1}{4} \frac{\sum_g \sum_{j \in g} \left(x_j^g - \langle x^g \rangle \right)^2}{N - N_g} \tag{4.8}$$

where the factor $1/4$ is a normalization factor that takes into account that the number of stations that we use in each event is 4 (explained later in Data selection).

Regarding the uncertainties, we do the same as we did for the method of splitting. Since Equation 4.8 is very similar to the expression of a variance, we do a Gaussian fit of the distributions of $x_j^g - \langle x^g \rangle$ with Equation 4.5. However, Equation 4.8 is not exactly the expression of a variance because of the factor in the denominator. We rescale the distribution of $x_j^g - \langle x^g \rangle$ with the factor $4(N - N_g)/(N-1)$, so that there is an equivalence between σ_{det}^2 and σ^2 in the gaussian fit, and the uncertainty can be taken from the uncertainty of the parameters of the fit.

2.3 The total variance: σ_{total}^2

The other term that is needed to obtain σ_{f}^2 besides σ_{det}^2 is σ_{total}^2. σ_{total}^2 is the total variance of the distribution of the $\overline{\text{ToD}}$. However, it is not as easy as computing directly the variance from the distribution of the $\overline{\text{ToD}}$, since we saw in the previous chapter that the $\overline{\text{ToD}}$ depends linearly on $\sec\theta$. That means that the variance of the distribution of $\overline{\text{ToD}}$ has a contribution due to the real variance of the $\overline{\text{ToD}}$ and another contribution due to the dependence on $\sec\theta$. Because this dependence is linear, it is easy to correct for it to obtain the real variance of the $\overline{\text{ToD}}$ distribution. For each energy bin we make a linear fit of $\overline{\text{ToD}}$ as a function of $\sec\theta$. Then, we pick a reference angle, for example $\sec\theta = 1$ [2], and then use the following equation to refer all the values to $\sec\theta = 1$:

$$\overline{\text{ToD}} := \overline{\text{ToD}} + (f(1) - f(\sec\theta)) \tag{4.9}$$

where $f(x) = a + bx$ is the function we use for the fit with a and b being the free parameters. That means that every value of the $\overline{\text{ToD}}$ will be its previous value plus a correction which

[2] The choice of this value has no effect since choosing another value would shift the whole $\overline{\text{ToD}}$ distribution by a constant value that leaves the variance unchanged.

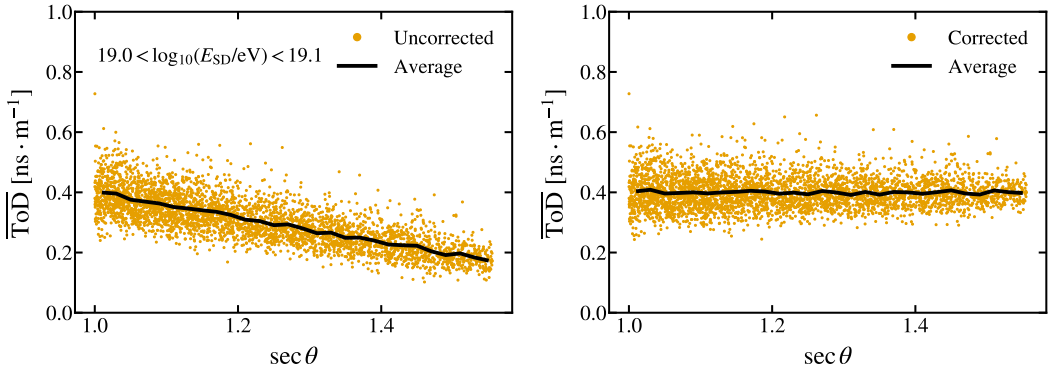

Figure 4.2: $\overline{\text{ToD}}$ as a function of $\sec\theta$ before (left) and after (right) applying the correction in Equation 4.9.

is the difference between the value of the fit at $\sec\theta = 1$ and the value of the fit at the corresponding $\sec\theta$ of each event. Once the correction in Equation 4.9 is applied, then σ^2_{total} is obtained by taking the variance of the corrected values of $\overline{\text{ToD}}$ for each energy bin. See Figure 4.2 for an example of the application of this correction: in the right panel the dependence with $\sec\theta$ has been removed.

2.4 Data selection

This work is based on the $\overline{\text{ToD}}$, thus some of the cuts are very similar to those explained in the previous chapter on page 46. The data used come from the same sample from March 2004 to June 2017. The cuts applied in this analysis are the following:

- No saturation in any of the low-gain or high-gain channels.
- $40\,\text{ns} < t_{1/2} < 2000\,\text{ns}$. The upper bound removes most of the extremely large risetimes that sometimes contribute greatly to the variances. This is a slightly stricter cut than in the previous chapter with 392 values with $2000\,\text{ns} < t_{1/2} < 3000\,\text{ns}$ and half of those values having $r > 2000\,\text{m}$ and $S > 5\,\text{VEM}$.
- We only include events that have a reconstructed energy E_{SD} above $10^{18.5}$ eV, so that the 1500 m array is fully efficient.
- Total signal $S > 5\,\text{VEM}$. The dependence of the risetime divided by the distance with the signal is only important at very low signals. 5 VEM is the value of the signal where the dependence of the risetime divided by the distance with S flattens and becomes constant.

- We use events that have a zenith angle $\theta < 50°$, because the $\overline{\text{ToD}}$ does not behave linearly with $\sec\theta$ above this angle.

- We only use events with a number of stations $n >= 4$. For events that have more than four stations we pick only the four stations with the largest total signal measured and we do not use the other stations in the analysis. The quality of the measurements depends on the number of stations used. In particular, the $\overline{\text{ToD}}$, which is defined as an average, has a resolution that improves with $1/\sqrt{n}$. Four stations is a trade-off between a number of stations large enough and enough statistics because the number of events with more stations decreases very rapidly as n increases. Furthermore, we pick only the four stations with the largest signal because with four stations it is possible to divide the event in two halves with two stations each. The following number that allows to divide an event in two parts with the same number of stations is six and that is a much stronger cut: less than 1% of the events would be selected betwen $10^{18.5}$ eV and $10^{18.6}$ eV.

2.5 Selection efficiency

The selection efficiency is shown for each energy bin in Table 4.1 and Table 4.2. The most demanding cut is to use only events that have four or more stations. This is specially true at the lowest energies where most events have two or three stations and that is the reason why the selection efficiencies are so low at these energies. There is a slight difference in the selection efficiency between simulations with the same model and different primary. For both QGSJetII-04 and EPOS-LHC the difference is between a 7% and a 10%.

3 Results

The results obtained are presented in the following subsections. First, the evolution of σ^2_{total} with the energy, then, the resolution of the detector σ^2_{det} and last, the fluctuations obtained with Equation 4.1 for the two methods studied.

3.1 The total variance: σ^2_{total}

After applying the correction in Equation 4.9, the variance of the distribution of the $\overline{\text{ToD}}$ for each energy bin has been computed in Figure 4.3. The total variance decreases with

	% of events selected				
$\log_{10}(E_{SD}/eV)$	Exp. data	Proton QGSJetII-04	Iron QGSJetII-04	Proton EPOS-LHC	Iron EPOS-LHC
18.5 - 18.6	31	24	29	24	32
18.8 - 18.9	72	60	69	62	71
19.0 - 19.1	90	81	86	83	87
19.4 - 19.5	99.7	99.0	99.5	99.2	99.7

Table 4.1: Percentage of the initial events remaining after the cuts used for different energies.

	stations	events with $n \geq 2$
Initial sample	454220 (100%)	91991 (100%)
40 ns $< t_{1/2} <$ 2000 ns	446735 (98.4%)	91976 (99.9%)
$S > 5$ VEM	353463 (78%)	91914 (99.9%)
$n = 4$	195520 (43%)	48880 (53%)

Table 4.2: Selection efficiency on the number of stations n and on the number of events with the cuts used for the data measured by the SD.

the energy and above $10^{19.5}$ eV the distributions of the $\overline{\text{ToD}}$ become narrower than those in simulations done with iron as the primary cosmic ray.

3.2 The method of splitting

We present the results obtained from the difference between the $\overline{\text{ToD}}_1$ (computed with the stations that are in an odd position when they are sorted by the measured total signal) and $\overline{\text{ToD}}_2$ (computed with the stations that are in an even position).

The distributions of the $\overline{\text{ToD}}$ in each group are very similar to the one using all the available stations, see the left panel of Figure 4.4. The distributions of $\overline{\text{ToD}}_1$ and $\overline{\text{ToD}}_2$ for each energy bin are shown in Figure 2 on page 144. We can also see that the difference $\Delta\overline{\text{ToD}} = \overline{\text{ToD}}_1 - \overline{\text{ToD}}_2$ has a distribution that is similar to a gaussian with wider tails and centred around zero in the right panel of Figure 4.4. It could be expected that the distributions of $\overline{\text{ToD}}_1$ and $\overline{\text{ToD}}_2$ are slightly different, since the stations in the first group have more signal than those in the second group. The distributions of signals in each group for all the energy bins is shown in Figure 3 on page 145. This has a very small effect on the distributions of $\overline{\text{ToD}}_1 - \overline{\text{ToD}}_2$ that are very well centred around zero (see Figure 4 on page 146).

The resolution of the detector σ^2_{det} is obtained from the width of the distributions of $\overline{\text{ToD}}_1 - \overline{\text{ToD}}_2$ (see Equation 4.4) and its statistical uncertainty is obtained from a fit of a gaussian function. Figure 4.5 is a plot of σ^2_{det} as a function of the reconstructed energy for

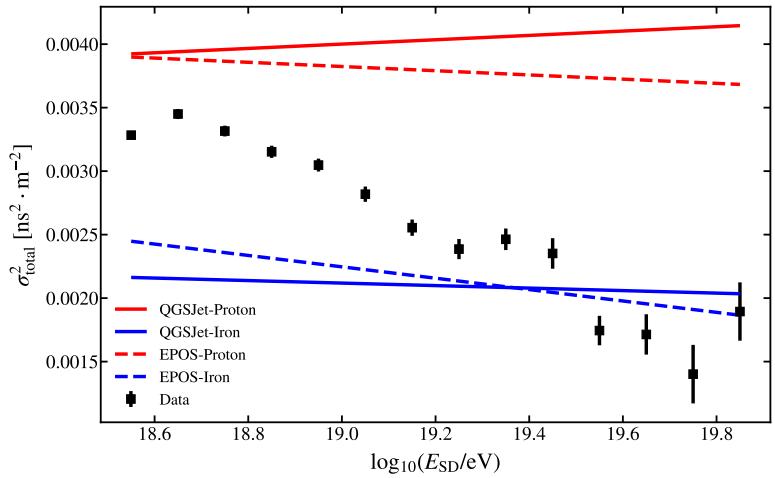

Figure 4.3: Evolution of the total variance of the distribution of the $\overline{\text{ToD}}$ with the energy.

both simulations and data. The resolution is better as the energy increases because stations measure a larger signal, this is also true for many physical observables. This decrease is faster for data than it is for simulations, which could be a signal of a change in composition towards heavier nuclei. Since the resolution improves with higher signal and because the signal size for heavier nuclei of the same energy is larger, the resolution for iron is lower than it is for protons.

When σ^2_{det} is subtracted from σ^2_{total} for each bin of energy, the results shown in Figure 4.6 are obtained. Comparing the scale of σ^2_{det} in Figure 4.5 with the fluctuations in Figure 4.6 we can see that the resolution of the detector is up to an order of magnitude larger than the fluctuations. It comes as no surprise, then, that the uncertainties are large when measuring the fluctuations. Even though the uncertainties are large, fluctuations are found to be significantly above zero for most energy bins and show a dependence with the energy.

To test whether the fluctuations depend on the energy or not we made fits of a constant and a straight line to the results obtained with the method of splitting. The constant in the first fit is 0.00045 ± 0.00004 ns^2 m^{-2} and for the straight line the intercept is 0.005 ± 0.003 ns^2 m^{-2} and the slope is -0.0003 ± 0.0001 ns^2 m^{-2}. A negative slope means that the average composition moves to heavier nuclei as the slopes of the simulations are of the order of -10^{-5}, so that fluctuations of data are steeper than the fluctuations of simulations. A maximum likelihood ratio test comparing the fit of the constant and the straight line yields a p-value $p = 6.7 \cdot 10^{-3}$ (2.7σ) for rejecting the constant in favour of the straight line. A constant value of the fluctuations with the energy can not be rejected but, with almost 3σ, the results point to a decrease of the fluctuations with the energy.

The results obtained in Figure 4.6 can be transformed to the average logarithm of the

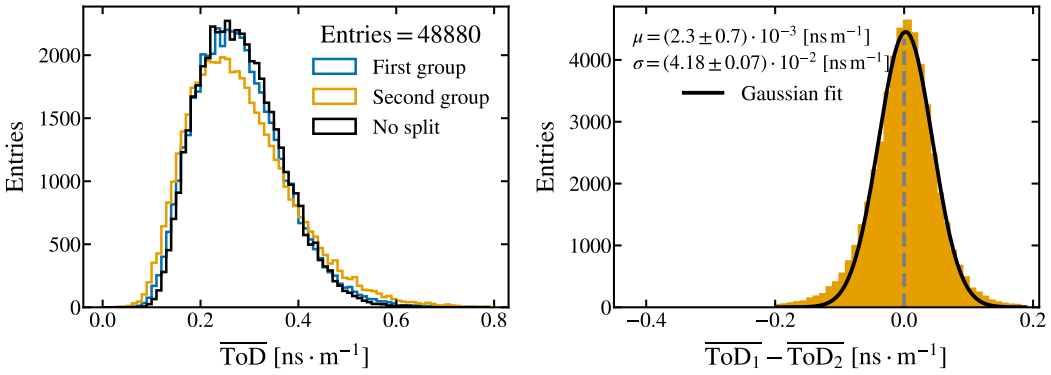

Figure 4.4: Left: Distribution of the values of the $\overline{\text{ToD}}$ when the events are separated in two groups and the distribution when no separation is done. Right: Distribution of the differences in the values of the $\overline{\text{ToD}}$ between the two groups. The distributions are obtained with all the available events (all energies and all zenith angles) for data.

mass with the following expression (the same way as it was done in Equation 3.15 in Chapter 3):

$$\langle \ln A \rangle = \ln 56 \frac{\sigma_f^2(\text{proton}) - \sigma_f^2(\text{data})}{\sigma_f^2(\text{proton}) - \sigma_f^2(\text{iron})} \tag{4.10}$$

The evolution of $\langle \ln A \rangle$ with the energy has been plotted in Figure 4.7. The last points, comprising all events with energy $E_{SD} > 10^{19.8}$ eV are above and below the value for proton and iron, respectively. When uncertainties are taken into account, however, it is not a significant deviation.

3.3 Analysis of variance

The results obtained with the analysis of variance are presented in this section. The values of σ^2_{det} found are in accordance with those obtained when splitting the event. Because of this, and since σ^2_{total} is the same for both methods, the fluctuations obtained with this method agree with the ones shown in the previous section. This comparison will be made clear later, when comparing both results in Figure 4.12.

In Figure 4.8 the evolution with energy of σ^2_{det} is shown and in Figure 4.9 the evolution with energy when σ^2_{det} is subtracted from σ^2_{total} is shown. Repeating what happened with the method of splitting we can see that the detector resolution is greater than the fluctuations and that fluctuations are found to be significantly above zero for most energy bins. The fits of a constant and a straight line have also been repeated and the p-value obtained is $8.6 \cdot 10^{-7}$ (4.9σ) in favour of the straight line.

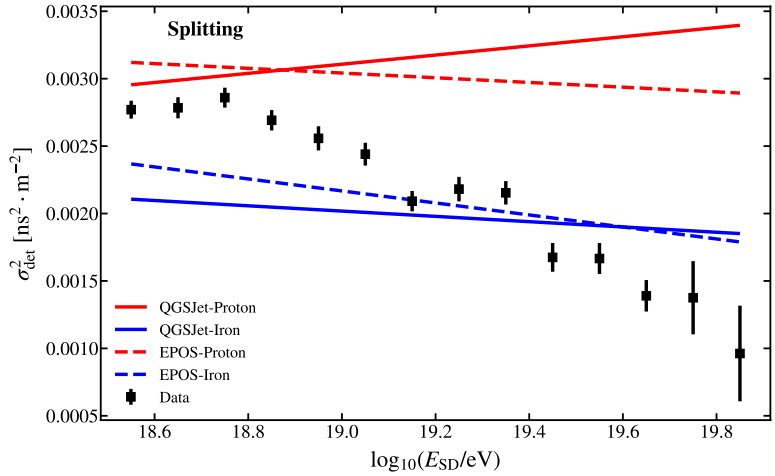

Figure 4.5: Evolution with energy of σ^2_{det} obtained with the method of splitting. The values for the simulations have been fitted with a straight line.

Using Equation 4.10, the results obtained by transforming to $\langle \ln A \rangle$ are shown in Figure 4.10. We see the same behaviour than with the method of splitting: the average mass increases with the energy and the last point is below the line for proton although compatible within the uncertainties.

Comparison with the uncertainty of the risetime

Using the method from the analysis of variance, we have made a comparison with the uncertainty of the risetime, obtained from [43] and σ^2_{det}. This is a good cross check to show that what we call σ^2_{det} is indeed a measurement of the resolution of the detector. We have made bins of energy, of sec θ and of distance to the core. We only pick events that have at least two stations that are in the same bin. We use Equation 4.8 to compute σ^2_{det} as we did before, with the different groups being the events.

In Figure 4.11 a comparison between the values of σ^2_{det} and the uncertainty of the risetime is made. The values of the uncertainty with ANOVA (black squares) agree qualitatively well with the uncertainty of the risetime and follow the same trend with the distance in a wide range of distances.

3.4 Comparison of both results

We have compared the results obtained with the two methods that have been explained. In Figure 4.12 both results are plotted in the same graph. We can see that not only they are compatible within the statistical uncertainties but also the trends are very similar. The

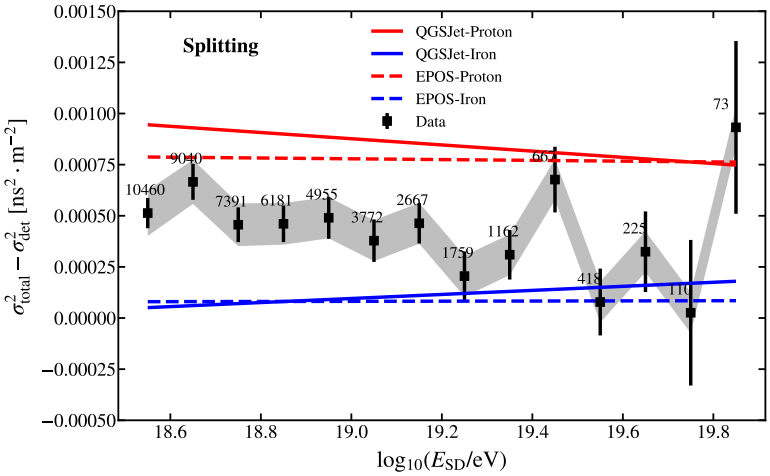

Figure 4.6: Evolution with energy of $\sigma^2_{\text{total}} - \sigma^2_{\text{det}}$ obtained with the method of splitting. The error bars correspond to the statistical uncertainty propagated quadratically from σ^2_{det} and σ^2_{total}. The shaded area corresponds to the systematic uncertainty and the numbers are the number of events in each energy bin for the data. The systematic uncertainties are explained later, on page 69.

fact that the two methods agree very well should not come as surprise since σ^2_{det} is very similar for both of them as it can be seen in Figures 4.5 and 4.8.

3.5 Systematic uncertainties

Both the method of splitting and the method from the analysis of variance have a very good property regarding the systematic uncertainties. Since they are obtained from differences coming from risetimes within the same event, many systematic uncertainties cancel out. This is true for anything that has the same effect on all the risetimes of the same event. We can see that in the equations from which σ^2_{det} is obtained: Equation 4.4 for the method of splitting and Equation 4.8 for the analysis of variance. Thus σ^2_{det} has no contributions to its systematic uncertainty from ageing of the detectors, dependence on the season of the year or on the time of the day, for example.

The most important contribution is the uncertainty on the absolute scale of the energy. We have shifted the energy of the data by $\pm 14\%$. The difference between the values with the energy shifted and those that do not have the energy shifted is taken as systematic uncertainty. The differences found with the two methods are very similar so the systematic uncertainty is taken to be the same in both cases and equal to 0.0001 ns^2 m^{-2}.

Another contribution comes from a bias in the selection. As we have seen in Table 4.1, there is a small difference between the percentage of events selected for proton and iron

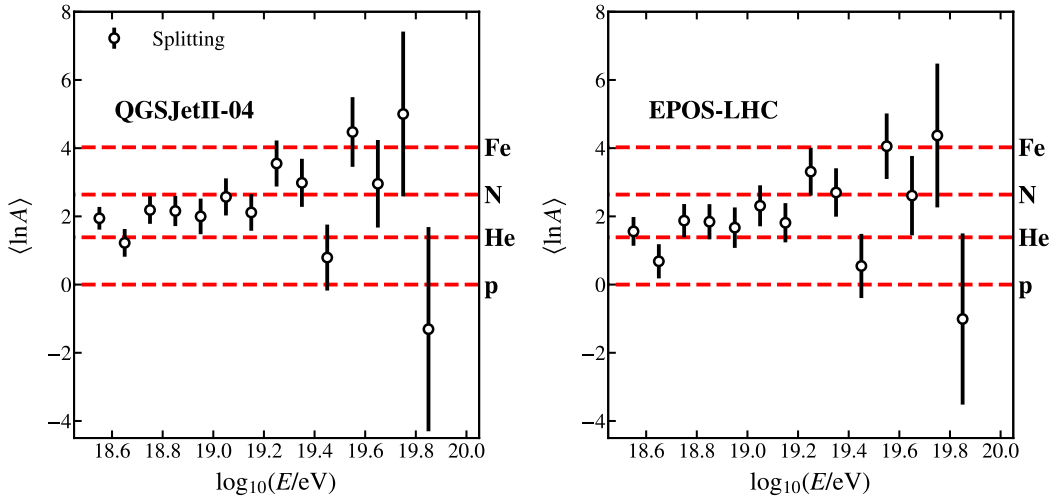

Figure 4.7: Evolution with energy of $\langle \ln A \rangle$ for QGSJetII-04 and EPOS-LHC, obtained with the method of splitting. The error bars have been obtained propagating the statistical uncertainty shown in Figure 4.6.

simulations. To study whether this can have an effect on the results, the worst case has been studied: a composition of 50% proton and 50% iron has been assumed. The value of $\sigma^2_{\text{total}} - \sigma^2_{\text{det}}$ has been computed with this composition and the composition obtained using the percentages of Table 4.1. The difference between the two values of $\sigma^2_{\text{total}} - \sigma^2_{\text{det}}$ is the corresponding systematic uncertainty. It is quite model independent since the bias is similar for both hadronic models. We found that it decreases with energy as expected since the bias in the selection is reduced with increasing energy. Only for the lowest energies the difference is above 10^{-5} ns^2 m^{-2}, which is an order of magnitude below the other contributions to the systematic uncertainty. The final uncertainty that we take is a linear function that goes from $5 \cdot 10^{-5}$ ns^2 m^{-2} at $10^{18.5}$ eV to 10^{-6} at 10^{19} eV and is 0 for energies above 10^{19} eV.

There are other sources of systematic uncertainty. Ageing was studied by computing σ^2_{total} before and after correcting by the linear change on the $\overline{\text{ToD}}$ as a function of the year in the bins with the lowest energies and the highest statistics. The change in the variance is less than one tenth of the contribution from the energy systematic uncertainty. The dependence on the seasons has also been studied but since every bin of energy has to be divided in four parts corresponding to each season and the statistical uncertainties are large, it is hard to establish a conclusion. σ^2_{total} is different for each season but it has not been found that the maximum and minimum value occur at the same season for every bin of energy, as a seasonal change would suggest. No contribution is taking from the seasonal dependence.

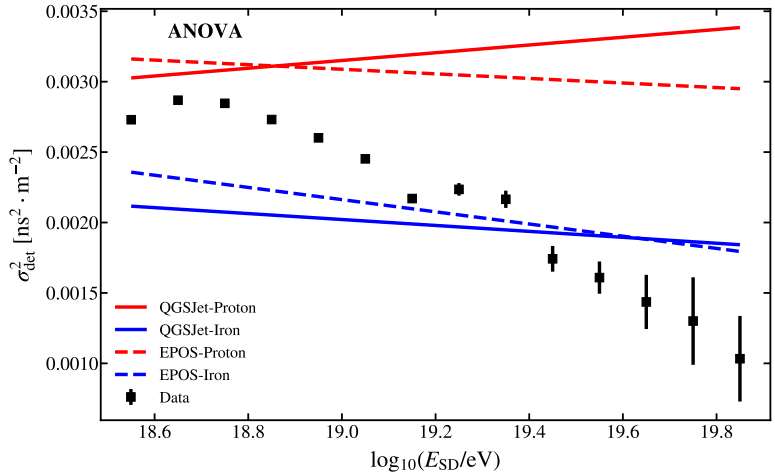

Figure 4.8: Evolution with energy of $\sigma^2_{\rm det}$ obtained with ANOVA. The values for the simulations have been fitted with a straight line.

The final systematic uncertainty is:

$$\sqrt{0.0001^2 + f(E)^2}\ \ {\rm ns}^2\,{\rm m}^{-2} \tag{4.11}$$

where $f(E)$ is a linear function that goes from $5\cdot10^{-5}$ ns^2 m^{-2} at $10^{18.5}$ eV to 10^{-6} at 10^{19} eV and is 0 for energies above 10^{19} eV.

4 Summary and conclusions

In this chapter we have developed two methods for measuring the extensive air shower fluctuations with the risetime divided by the distance to the core. We have found that there are fluctuations due to physics and a possible spread in mass and that fluctuations decrease with the energy. A change in composition towards heavier nuclei can explain this decrease and this is compatible with what the FD measures in the same energy range. Although the resolution of the detector dominates by comparison with the fluctuations, we demonstrate that it is possible to measure the fluctuations using information from the SD only.

We found that uncertainties are large[3]. This is due to the demanding cuts that were needed and also to the size of the detector. The resolution improves with the signal size

[3]The fact that the uncertainties are large should not be surprising. From a purely statistical point of view, to do a precise estimation of the variance of a sample, many samples (>500) are needed. Above $10^{19.5}$ eV all the energy bins for the data have less than 450 events.

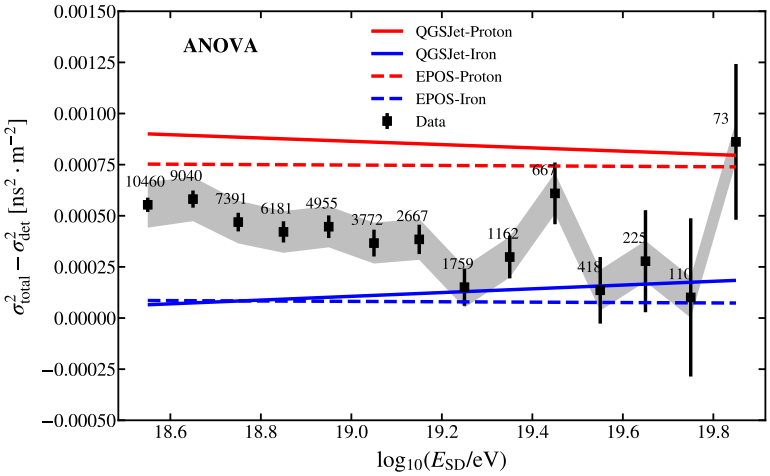

Figure 4.9: Evolution with energy of $\sigma^2_{\text{total}} - \sigma^2_{\text{det}}$ obtained with ANOVA. The error bars correspond to the statistical uncertainty propagated quadratically from σ^2_{det} and σ^2_{total}. The shaded area corresponds to the systematic uncertainty and the numbers are the number of events in each energy bin for the data.

so a detector with an area greater than the current 10 m^2 would help to obtain a better estimation of the fluctuations. However, the most limiting factor is the number of events that can be used in this study.

Both methods give results that are in very good agreement as it can be seen in Figure 4.12. It is also encouraging that the method from ANOVA can reproduce the uncertainty of the risetime, see Figure 4.11. That means that σ^2_{det} is a reliable measurement of the resolution of the detector, because the parameterization of the uncertainty of the risetime comes from the different values measured by twins and pairs of detectors [43].

This approach would need more statistics to be able to make stronger claims. We can see in the results obtained in Figure 4.7 and in Figure 4.10 that above $10^{19.3}$ eV the statistical uncertainties become larger than the systematic uncertainties. It would be necessary to have of the order of a thousand events in each bin to be able to distinguish between a heavy or light composition.

A denser array would be beneficial to increase the event selection efficiency at the lowest energies. However, since the selection efficiencies are already high (more than 80%) above 10^{19} eV, the number of events would remain similar. The quality of the measurement of σ^2_{det} and the resolution of $\overline{\text{ToD}}$, on the other side, would improve as $1/\sqrt{n}$, with n being the number of stations in each event. We reach the same conclusion as before; statistics should be much larger since the resolution increases slowly with both the number of events and the number of stations in each event.

This work was published as an internal note of the Pierre Auger Collaboration [98].

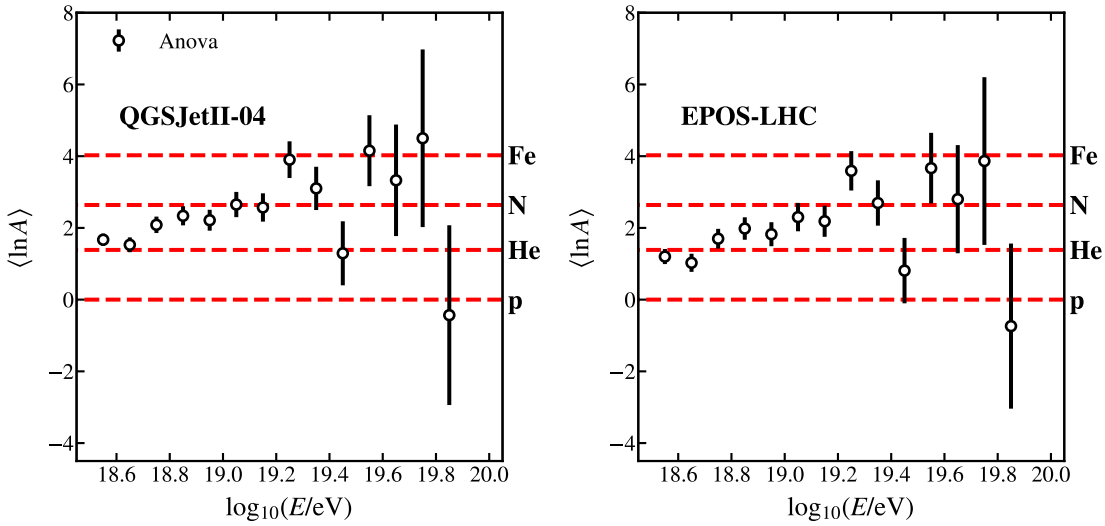

Figure 4.10: Evolution with energy of $\langle \ln A \rangle$ for QGSJetII-04 and EPOS-LHC, obtained with ANOVA.

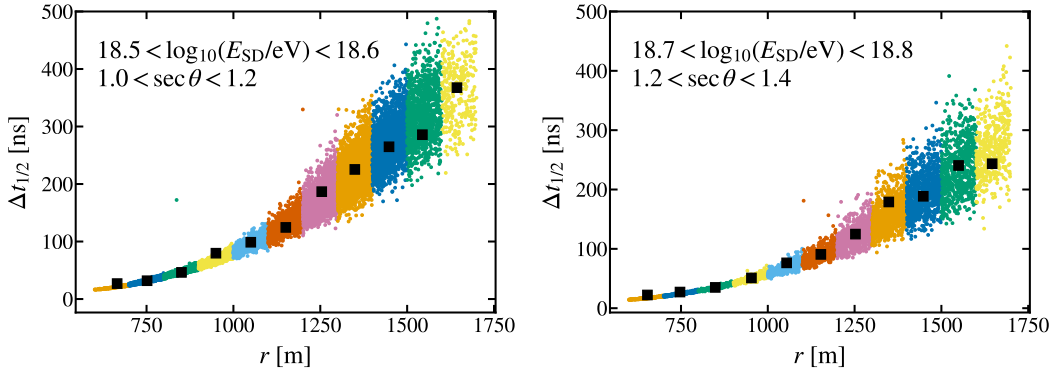

Figure 4.11: Uncertainty of the risetime as a function of the distance for two bins of energy and zenith angle. Coloured circles correspond to individual values of $\Delta t_{1/2}$ from the parameterization done in [43] while the black squares are the values obtained with the ANOVA method. Only events that have two or more stations in one of the bins of 100 m have been considered and Equation 4.8 has been used to compute the values of the black squares. No cuts have been applied to data to increase the statistics for the plot.

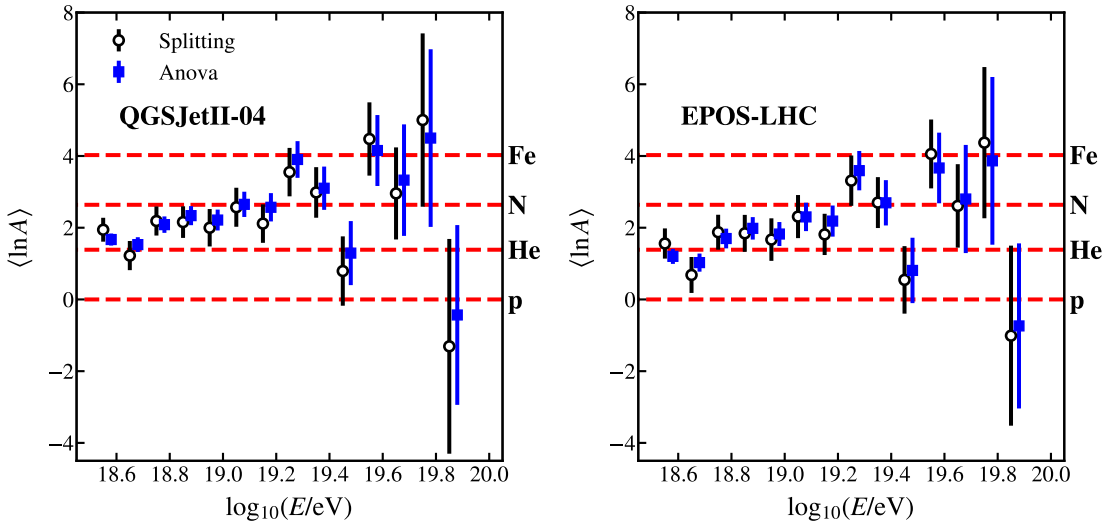

Figure 4.12: $\langle \ln A \rangle$ for the methods used in this analysis.

5

Introduction to Machine Learning

Machine learning is a field that studies algorithms or models that rely on pattern inference from data. These algorithms focus on statistical techniques using computers for the estimation of complicated functions. Over the last few years, machine learning methods have proven to be better than many other state-of-the-art techniques at their respective fields. Together with advances in computing power and increasing amounts of data available, the interest in machine learning has been growing.

Machine learning has also had an impact in physics. For a recent review of the applications of machine learning in different fields of physics see ref. [99]. In particular, experiments in high-energy physics are very well suited for the application of these algorithms. In these experiments a lot of data is collected and made readily available in digital form. Also, there are frequently simulations modelling the data and the experiments that can be used to train the machine learning algorithms.

This chapter is an introduction to machine learning with the objective of providing some background knowledge to understand the core results of this thesis. It begins with a short introduction of basic machine learning concepts and some of its applications in Section 1. Afterwards, one of the main topics of this thesis is used to make a simple dataset in Section 2. Three kinds of machine learning algorithms are studied and applied to this dataset as examples: linear regression in Section 3, tree methods in Section 4 and neural networks in Section 5.

1 Machine learning

In machine learning, algorithms perform tasks using some data. The following is a list of some of the tasks that machine learning algorithms can do, although there are many more [100]:

- Classification. The machine learning algorithm outputs to which class an input object belongs to from several classes. Usually, the output of the algorithm is a vector of probabilities of the object to belong to each class.
- Regression. A numerical value is predicted as a function of the input.
- Transcription. Given some input, transform it into text. One classical example is optical character recognition.
- Synthesis and sampling. The machine learning algorithm generates new samples that are similar to those in the training data.
- Feature extraction. Given some input, output features of the data that are useful for a defined task. One example is the Principal Component Analysis (PCA).

There are many algorithms to accomplish these tasks and they are chosen based on the objective and the properties of the algorithm. Algorithms in machine learning are generally divided in two classes: supervised and unsupervised algorithms. In supervised algorithms the data has some features and some label or target associated. The main task is usually to make a map between the features and the label. One example of this is when an algorithm is trained to make a prediction based on what it has seen for the training data, where there is a label for each input sample. The difference with unsupervised algorithms is that in the latter there is no label in the training data and the algorithms have to learn structures or patterns in data by themselves. One example of unsupervised algorithms is clustering: similar examples are grouped based on a well defined measure.

Associated with the algorithms and the tasks, a performance measure has to be chosen. This measure can assess how well the algorithm is doing. For example, in a classification task the measure could be the accuracy, the number of correctly classified samples. In a regression task, the measure could be the mean squared error between the target and the prediction of the model.

The objective of machine learning algorithms is to perform well on unseen data. This is usually known as generalization. There are two key points regarding the performance of algorithms when generalizing: how well an algorithm does for the training data and the difference between its performance when generalizing to unseen data with respect to the

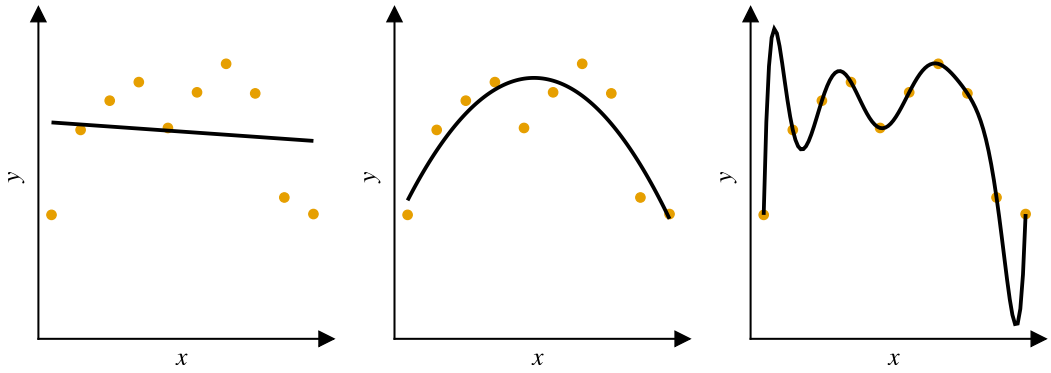

Figure 5.1: Comparison of polynomial models of degree n fitting data generated using a quadratic polynomial with random gaussian noise added. Left: $n = 1$. Middle: $n = 2$. Right: $n = 9$.

training data. When the model is not able to fit the data we say there is underfitting (bad performance on the training data). Underfitting can happen when the model employed is simple and can not capture the patterns in the data. When the model performs much better for the training set than for unseen data there is overfitting. Overfitting can happen when the amount of training data is low and the model does not learn patterns in data but focuses on specific examples. Note that there can be underfitting and overfitting at the same time.

Examples of underfitting and overfitting are shown in Figure 5.1. A comparison of polynomial fits to data generated using a second degree polynomial has been made. In the left panel the model chosen is too simple to capture the quadratic dependence of y on x and there is underfitting. In the right panel the model fits perfectly all the points in the training data but it is unlikely to generalize well for points that follow the same distribution as the training data; there is overfitting. The model used for the middle panel is the one used to generate the data. There is no underfitting or overfitting as the performance for the training data is good and it is likely to be a good fit if other points were generated in the same way.

To control under and overfitting and assess the performance of the algorithm both on training data and on unseen data, available data is usually split on several parts: a training set, used to fit the algorithm; a validation or development set, used to measure the performance of the method while training; and a test set, used to show the performance on unseen data after the algorithm or model has been fixed. The additional split to obtain the test set (instead of only having one set for training and another one for validation or testing) is done to avoid overfitting to the development set, which could happen as the best model or hyperparameters are chosen based on their performance on this set.

2 Data for predicting the muon signal

One of the main topics of this thesis is the problem of predicting the amount of muon signal that would be measured at each station. We are going to use this problem to explain several algorithms in machine learning. Our objective will be to train a model with some variables to predict another one.

The dataset that we are using is a set of simulations of cosmic rays events measured by the SD of the Pierre Auger Observatory where the muon signal S^μ is known. We will use the following variables that can be measured in data to predict S^μ:

- The reconstructed energy E_{SD}
- The reconstructed zenith angle θ
- The total signal measured by each station S
- The distance to the reconstructed core of each station r

Training, development and test sets

Regarding the data, we follow the standard procedure in machine learning of dividing the dataset in several parts. One part is the training dataset: this is the data that will be given to the algorithm that we use to make the model. Another part is called the validation or development set and is reserved to evaluate the performance of the method while the training is going on. There is a third part called the test set that is used to evaluate the performance of the method on unseen data once the method has been trained and fixed.

The training set has a total of 20000 stations and the distribution of the variables that will be used as input and the target are shown in Figure 5.2.

Data scaling

It is also a standard procedure to scale the data before feeding it to our methods. This is done to avoid numerical problems, such as adding a very small quantity to a very large quantity. There are different ways of doing data scaling: each variable can be shifted to the range $[0, 1)$ by subtracting the minimum value and dividing by the new largest value. Similarly, it can be scaled to the range $[-1, 1]$. We have scaled all the input variables to have zero mean and variance equal to one by subtracting the mean and dividing by the standard deviation of its distribution. The same scaling is applied to the development and test sets using the mean and standard deviation obtained for the training set.

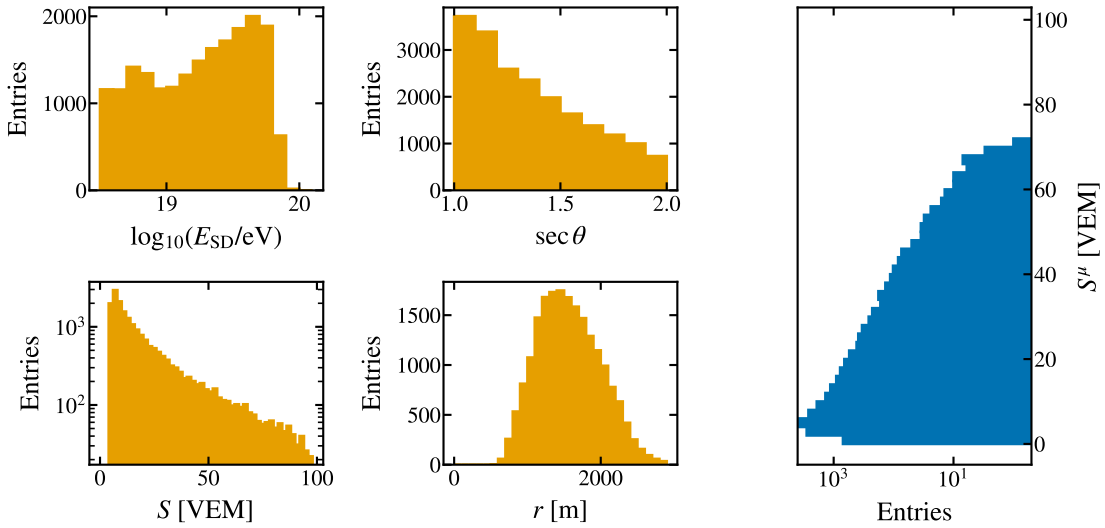

Figure 5.2: Distributions of the available features for the training data used for the examples. For the four plots of the left, corresponding to the input features, from left to right and top to bottom: Logarithm of the reconstructed energy of the event, secant of the reconstructed zenith angle, total signal measured by the station and distance to the core of the station. Right: Distribution of muon signal, the target of the predictions.

3 Example I: Linear regression

Linear regression is one of simplest methods in machine learning. It is fast to execute, easy to implement and can give good results despite its simplicity. It can also be used as an initial baseline benchmark when using other more sophisticated methods. See ref. [101] for a lecture on linear regression and the methods used to solve it.

Linear regression is defined as a linear model of its parameters:

$$\hat{y} = \tilde{w}\tilde{X} + b \tag{5.1}$$

where $\hat{y}$ is the output of the linear model, $\tilde{w}$ is a vector of free parameters, $\tilde{X}$ is the input matrix and b is real number called intercept. Both $\hat{y}$ and $\tilde{w}$ are row vectors while $\tilde{X}$ is a matrix:

$$\begin{aligned} \hat{y} &= \begin{pmatrix} \hat{y}_1 & \hat{y}_2 & \cdots & \hat{y}_m \end{pmatrix} \\ \tilde{w} &= \begin{pmatrix} w_1 & w_2 & \cdots & w_n \end{pmatrix} \end{aligned} \qquad \tilde{X} = \begin{pmatrix} \vec{x}_1 \\ \vec{x}_2 \\ \vdots \\ \vec{x}_m \end{pmatrix} = \begin{pmatrix} x_{11} & x_{12} & \cdots & x_{1n} \\ x_{21} & x_{22} & \cdots & x_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ x_{m1} & x_{m2} & \cdots & x_{mn} \end{pmatrix} \tag{5.2}$$

where m is the number of samples and n is the number of features. That means that each row of the matrix $\tilde{X}$ is one sample or observation and each column has the values of the same feature for all the samples. There are as many predictions as samples and as many free parameters as features the data has. To make the notation easier b can be absorbed into $\tilde{w}$ if a column of ones is added to $\tilde{X}$ and we denote the result as w and X respectively:

$$w = \begin{pmatrix} w_1 & w_2 & \cdots & w_n & b \end{pmatrix} \qquad X = \begin{pmatrix} x_{11} & x_{12} & \cdots & x_{1n} & 1 \\ x_{21} & x_{22} & \cdots & x_{2n} & 1 \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ x_{m1} & x_{m2} & \cdots & x_{mn} & 1 \end{pmatrix} \tag{5.3}$$

This simplifies the computations of the derivatives and the implementation since it is not necessary to have derivatives with respect to both w and b. Equation 5.1 transforms to:

$$\hat{y} = wX \tag{5.4}$$

To obtain the optimal values of w a function is going to be minimized. This function is usually called cost function and has to take into account the differences between the predicted and true values, such that lower values of the function mean that the model is making better predictions. There are several possible choices and the model obtained with each choice will make different predictions and have different properties. For this example, the difference squared between the predictions and the true values is being minimized and the optimal values of w, w_{opt}, are obtained as follows:

$$C = ||\hat{y} - y||^2 = ||wX - y||^2 \tag{5.5}$$

$$w_{\text{opt}} = \min_w C \tag{5.6}$$

We are going to minimize C and to do so we need the derivative of C with respect to each of the components of w. We define $\partial/\partial w$ as a row vector whose components are the derivatives with respect to each of the components of w. $\partial/\partial w$ is applied on scalars such as C. With this notation and arbitrary row vectors a and a symmetric square matrix A:

$$\frac{\partial w a^T}{\partial w} = a^T, \quad \frac{\partial a w^T}{\partial w} = a, \quad \frac{\partial w A w^T}{\partial w} = wA + wA^T = 2wA \tag{5.7}$$

where ^{T} is the transpose operation. To obtain the derivatives of C, first we expand C:

$$C = ||wX - y||^2 = (wX - y)(X^T w^T - y^T) = wXX^T w - wXy^T - yX^T w^T + yy^T \tag{5.8}$$

and now we compute the derivatives using Equation 5.7:

$$\frac{\partial C}{\partial w} = 2wXX^T - Xy^T - yX^T \tag{5.9}$$

3.1 Analytical solution

Linear regression is a convex problem and has a single global minimum [102]. From Equation 5.9, a linear system of equations can be made:

$$\frac{\partial C}{\partial w} = 0 \;\Rightarrow\; 2wXX^T = Xy^T + yX^T \tag{5.10}$$

This is a system of $m+1$ equations and $m+1$ unknowns, the number of features plus one.

For the example applied to predict the muon signal S^μ, w_{opt} is found to be:

$$w_{\text{opt}} = \begin{pmatrix} 2.00 & 1.87 & 8.26 & -2.00 & 11.65 \end{pmatrix} \tag{5.11}$$

which means that the predicted muon signal $\widehat{S_\mu}$ will be obtained as

$$\widehat{S_\mu} = 2.00\,\overline{E_{\text{SD}}} + 1.87\,\overline{\sec\theta} + 8.26\,\overline{S} - 2.00\,\bar{r} + 11.65 \tag{5.12}$$

where the var on top of the variables means that the variables have been normalized as explained before.

In general, there is not an analytic solution for other methods in machine learning. Although it exists for linear regression, it does not scale well with the problem size as the matrix XX^T is a matrix with dimensions $(m+1, m+1)$ which is approximately m^2 for large problems. Inverting the matrix to solve the system would require of the order of m^3 operations, while the number of multiplications in the matrix product XX^T scales as $m^2 n$. Gradient based methods scale better and are better suited for large problems.

3.2 Gradient descent

We can find w_{opt} by *gradient descent*, which is an iterative process. First, w is given an initial value. In each iteration, $\partial C/\partial w$ is computed using Equation 5.9 and w is updated as follows:

$$w := w - \alpha \frac{\partial C}{\partial w} \tag{5.13}$$

where α is called the *learning rate* and is a parameter[1] that can be tuned. The choice of α is important; it should not be very large because then w is modified by a large amount at each iteration and can overshoot the minimum and if α is too small it can take many iterations to reach the minimum.

[1] Instead of α being a single number it is also possible to have a different α_i for each component of $\partial C/\partial w$.

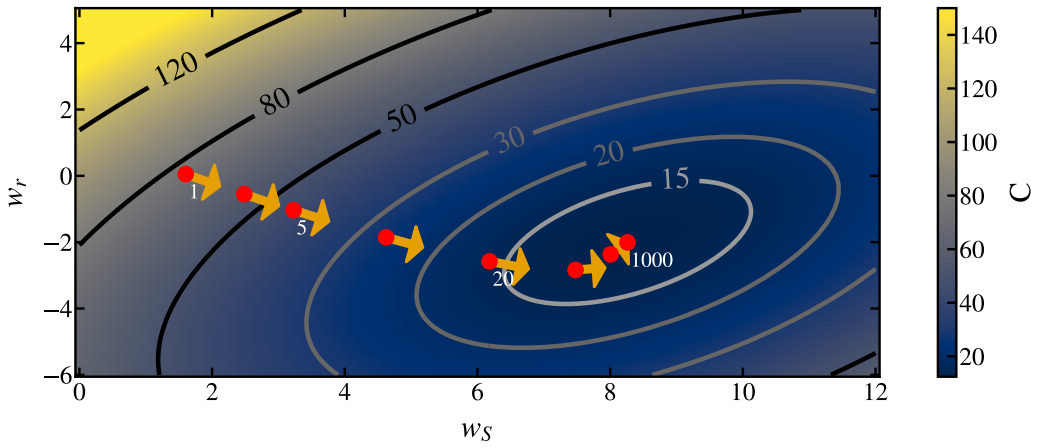

Figure 5.3: Plot of the cost function as a function of two of the parameters in w, the ones that multiply r and S in Equation 5.4, when the rest are set to the values given by its solution in Equation 5.11. The red points correspond to some of the different values of w obtained with gradient descent. The arrows represent the direction of the negative gradient for each value of w chosen. The number of the current iteration is also shown for some points.

The process can be repeated until C stops changing under a certain tolerance or a number of iterations has been reached, for example. Using this method with $\alpha = 0.03$ and 2000 iterations it was found:

$$w_{\text{opt}} = \begin{pmatrix} 2.00 & 1.87 & 8.26 & -2.00 & 11.65 \end{pmatrix} \tag{5.14}$$

When compared to what was found with the analytical solution in Equation 5.11, differences in the values of the parameters start to appear in the fifth or sixth decimal place. In Figure 5.3 the parameters at positions three and four in w have been left free while the others have been fixed to its solution value to illustrate how w changes during the iterative process.

3.3 Adding features

The linear model that we have used is simple but it can be improved with a small modification. Polynomial combinations of the features can be added to the input. For example, if polynomial features up to second degree are added, the new input matrix would be:

$$X = \begin{pmatrix} x_{11}^2 & x_{12}^2 & \cdots & x_{1n}^2 & x_{11}x_{12} & \cdots & x_{1n-1}x_{1n} & x_{11} & x_{12} & \cdots & x_{1n} & 1 \\ x_{21}^2 & x_{22}^2 & \cdots & x_{2n}^2 & x_{21}x_{22} & \cdots & x_{2n-1}x_{2n} & x_{21} & x_{22} & \cdots & x_{2n} & 1 \\ \vdots & \vdots & \ddots & \vdots & \vdots & \ddots & \vdots & \vdots & \vdots & \ddots & \vdots & \vdots \\ x_{m1}^2 & x_{m2}^2 & \cdots & x_{mn}^2 & x_{m1}x_{m2} & \cdots & x_{mn-1}x_{mn} & x_{m1} & x_{m2} & \cdots & x_{mn} & 1 \end{pmatrix} \tag{5.15}$$

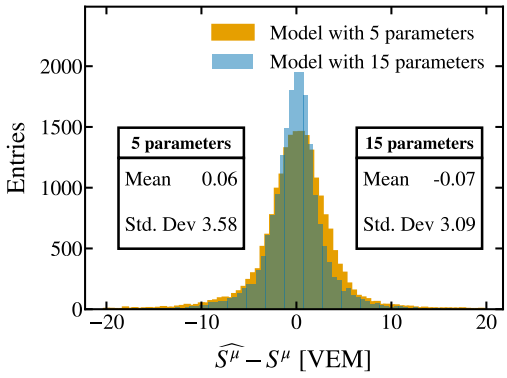

Figure 5.4: Distribution of the differences between the true value and the predicted value of the muon signal for the validation set. The model with 5 parameters is the standard linear regression model while for the model with 15 parameters polynomial features up to second order have been added.

After the polynomial features have been added the solution can be found using the analytical technique or gradient descent.

For this example polynomial features up to second order have been included. The model changes from 5 free parameters to 15. In Figure 5.4 the distribution of the difference between the predicted and true value of the muon signal is plotted. There is an improvement over the model without polynomial features; this new model has more parameters and is able to model better the data.

4 Example II: Tree methods

Tree methods are another set of algorithms widely used in machine learning. The main building block of these models are decision trees. Since decision trees are very simple models that usually do not give the desired accuracy, it is necessary to use ensembling or other grouping techniques to obtain good performance. See ref. [102] for more information about decision trees.

4.1 Decision tree

Decision trees are a very simple but powerful method. The basic idea is to perform a split on each feature X of the data, defining two regions $x \leq X$ and $x > X$. Then, a simple model is fitted to each split or region R. In its most basic form, the model is a constant for

each region. The phase space can be thought as being divided in rectangles and a different constant will be predicted for each rectangle. If we name each feature of our data with $X_1, X_2, \ldots, X_n$, and the data is partitioned into r regions $R_1, R_2, \ldots, R_r$ the output of the decision tree will be:

$$\hat{f}(X) = \sum_{i=1}^{r} c_i I\{(X_1, X_2, \ldots, X_n) \in R_i\} \tag{5.16}$$

where I is an indicator function: 1 if the input with features $(X_1, X_2, \ldots, X_n)$ is in the region R_i and 0 if it is not; R_i are the regions defined by the splits and c_i is the constant obtained for each region.

How is a decision tree built? It turns out that finding the optimal tree is a problem that does not scale well so a greedy approach is followed. In an iterative process, the best split is found and chosen given the previous splits. To build the tree, first the metric or the function that will be minimized has to be chosen. For this example we chose the Mean Squared Error or MSE. It is easy to see that the constant that minimizes the MSE for each region is the mean:

$$c_i = \text{mean}(y_j | x_j \in R) \tag{5.17}$$

The split s for the feature j that will define the regions $X_j \leq s$ and $X_j > s$ is found requiring that the sum of the MSE in each region is minimized:

$$\min_{j,s} \left[\min_{c_1} \frac{1}{n_1} \sum_{x_i \in R(j,s)} (y_i - c_1)^2 + \min_{c_2} \frac{1}{n_2} \sum_{x_i \in R(j,s)} (y_i - c_2)^2 \right] \tag{5.18}$$

The process to build a decision tree is summarized as follows: for each feature X_j the data is sorted along that feature and the best split is found. One way of doing so is by trying all the possible splits. For each region, the constant is found using Equation 5.17. The value of the split s and the total MSE is saved for each feature and the process repeated for all the features. Then, the feature that gives the lower MSE is chosen for the split. The process can be repeated for the data in each split until some stopping condition is met, such as the maximum depth of the tree allowed. See the left panel of Figure 5.5 for an example of a decision tree of depth two for predicting the muon signal.

In the right panel of Figure 5.5, it can be seen how the MSE evolves for different depths of the decision tree for the training and development sets. For depths greater than four the decision tree starts to overfit since the performance for the development becomes worse than for the training set.

Another problem of the decision trees can be seen in the left panel of Figure 5.5. What this tree does is assign lower values of the muon signal for lower values of the total signal S and vice versa. It is trying to model a linear relationship but decision trees are notoriously bad at this task since they predict a constant value for each region.

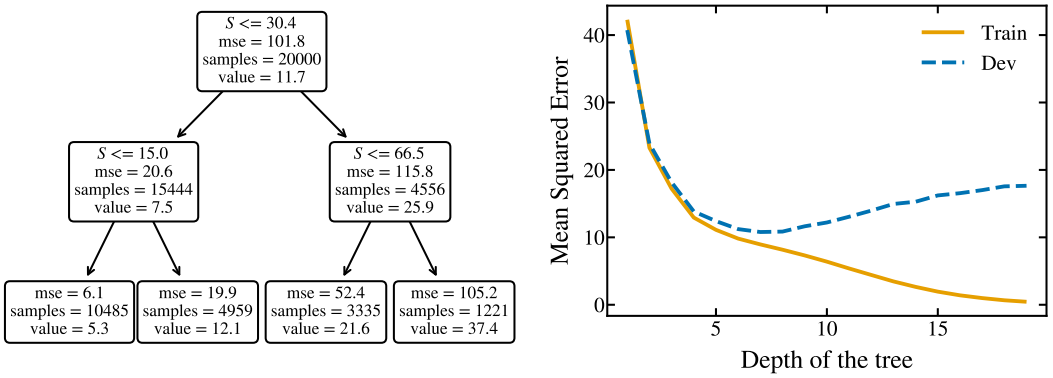

Figure 5.5: Left: Decision tree obtained when the maximum depth is set to two. For each leaf of the tree there is information about the MSE before doing the splits, the number of samples and the value of the constant for that leaf. Going to the left means that the condition in the current leaf is true and going to the right means it is false. Right: Mean Squared Error as a function of the maximum depth of the tree for the train and development datasets.

However, decision trees have some good properties. Since finding the splits only requires ranking the values, they are invariant under data rescaling. This means that usually little preprocessing is needed when using decision trees. Another good property is that the model is not a black box; it can be inspected as we have done with Figure 5.5. In practice, more complicated models based on decision trees are used, such as random forests or boosted decision trees.

4.2 Random forest

Random forest is an example of the more general technique of bagging. Many decision trees are trained on a subset of the training data and their predictions are averaged. The subset of the data can be obtained by picking only a subset of all the features or a subset of the samples belonging to the training data or a combination of both for each decision tree. With this technique sensitivity to the training dataset is reduced, overfitting decreases and it is possible to obtain better predictions than with a single decision tree.

In the left panel of Figure 5.6, random forests with different number of trees have been fitted to the training data and evaluated on both the training and validation data. There is a slight overfitting even though only a random 50% of the data has been used to fit each decision tree. However, this overfitting is not as severe as with a single decision tree. The MSE improves as the number of trees used increases.

4.3 Boosted decision tree

Boosting consists on having several weak (simple and not accurate) learners whose predictions are combined to have a better prediction. One example of a weak learner that we have studied is a decision tree. Algorithms for boosting vary; the classic AdaBoost is an iterative algorithm in which weights are given to the points in the training set so that each learner can focus on the samples for which the performance was worse with previous learners [102].

We denote the output of a single decision tree for a single sample x by $T(x;\Theta_m)$, where $\Theta_m = \{R_m, \gamma_m\}$, that is, Θ_m corresponds to the free parameters of the tree: the set of regions that it defines, R_m, and the constants that predicts for each region, γ_m. The boosted tree model is a sum of decision trees:

$$f_M(x) = \sum_{i=1}^{M} T(x;\Theta_i) \tag{5.19}$$

and the parameters Θ_m are found by minimizing the loss or cost function which is usually introduced in a stage wise form:

$$\Theta = \arg\min \sum_{i=1}^{m} L(y_i, f_{i-1} + T(x_i;\Theta_i)) \tag{5.20}$$

Finding the optimal solution of Equation 5.20 is not computationally feasible. What is usually done is a greedy iterative process: first f_0 is initialized and fixed to some value γ, then f_1 is obtained and fixed, then f_2 and so on. Even then, Equation 5.20 is only easy to solve for a few loss functions L. One way of solving it in the greedy approach is using gradient boosting, giving the name of gradient boosted decision trees. The trees are fitted to the negative gradient values of the cost function. When L is the MSE, fitting to the negative gradient is equivalent to fitting each tree to the residuals between the true value and the previous prediction since $(y - (f_{m-1} + T))^2 = ((y - f_{m-1}) - T)^2$. Each successive tree focus on the examples for which the previous trees gave a worse prediction.

In the right panel of Figure 5.6, a gradient boosted decision tree has been trained to predict the muon signal and its performance is shown as the training develops. The performance obtained for the validation set is better than using a single decision tree or a random forest.

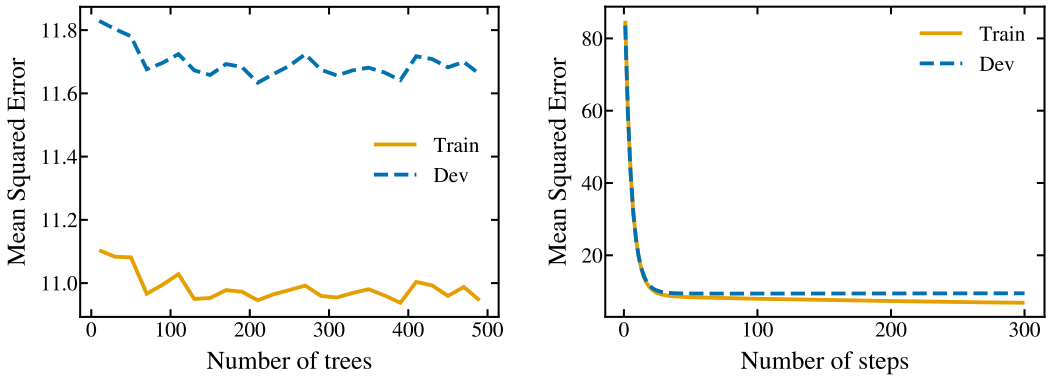

Figure 5.6: Left: MSE as a function of the number of trees using random forests. Starting from 10 trees and with steps of 20 trees until 500 trees, a new random forest has been trained for each step with maximum depth equal to four and using a random 50% of the training data for each tree. Right: MSE as a function of the step while training a single gradient boosted decision tree.

5 Example III: Neural Networks

Neural Networks is one of the techniques within machine learning that has received the most attention lately. They have proven to perform very well for many different tasks such as object recognition in images or natural language processing. Even though they were invented in the 60s [100], recent advances both in software but particularly in hardware, have allowed for their resurgence.

We have studied linear regression and tree methods already. A neural network is also a model that has free parameters that have to be tuned. These parameters are usually organized in layers, the building blocks of neural networks. Each layer performs some operation on its input and gives and output. Layers can be connected to other layers in different ways and do not necessarily have to have free parameters but often do. See ref. [103] for an overview and implementation of a simple neural network and refs. [104,105] for reviews on deep learning, the field that studies deep neural networks.

5.1 Fully connected layer

For this example we are going to build a neural network from scratch using the simplest architecture: a sequential model of fully connected layers. After defining it and describing the training process, we apply the neural network that we build to the problem of predicting the muon signal. Even though for this example of a neural network a particular choice of

architecture has been made, it is quite general as most neural networks work in a similar way.

A fully connected layer is one that applies a linear transformation on its input and a function called activation function afterwards. For a single observation x:

$$y = g(Wx + b) \tag{5.21}$$

where x is the input (a vector with as many entries as the number of features present in the input), y is the output, W and b contain the free parameters of this layer, usually called weights, and g is an activation function that has to be chosen. W is a matrix of free parameters, Wx is the regular matrix product, b is a vector and the addition is an addition of two vectorrs since Wx produces a vector. The function g is applied element wise, returning also a vector. Fully connected layers are defined by their number of units or their output dimension. If x has n features, then W has dimensions (d, n) so that Wx has dimensions $(d, 1)$.

Equation 5.21 is valid for a single observation. In practice, instead of using a single observation many are stacked together as this allows to obtain the output for all of them at the same time. The modifications done to Equation 5.21 are to change x to a matrix, b to a matrix which is a repetition of the original vector b and the output y becomes a vector with the labels corresponding to each observation in the input.

5.2 Neural Network

The neural network that we are using has L fully connected layers, $l = 1, \dots, L$, and the index 0 is reserved for the input. The output of each layer is obtained as follows:

$$a^l = g(W^l a^{l-1} + b^l), \quad l = 1, \dots, L \tag{5.22}$$

and a^L is the output of the neural network. Each layer l has a number of units d_l and d_0 is the number of features of our data. We use the same activation function g for all the layers. It is common in the literature to introduce z^l to make the calculations clearer and rewrite Equation 5.22:

$$z^l = W^l a^{l-1} + b^l \implies a^l = g(z^l) \tag{5.23}$$

The input X is a matrix with dimensions (m, n) (m samples and n features), see Equation 5.15. We can set $a_0 = X^T$ so that the following rule for dimensions holds for each l:

$$\underset{(d_l,n)}{a^l} = g(\underset{(d_l,d_{l-1})}{W^l}\ \underset{(d_{l-1},n)}{a^{l-1}} + \underset{(d_l,n)}{b^l}) \tag{5.24}$$

Note that d_L is the dimension of the output of the neural network. Since in this case we will want a single value, the muon signal, we will set $d_L = 1$.

We use as activation function the Rectified Linear Unit (ReLU) that has the following definition and derivative:

$$g(x) = \max(0, x); \quad g'(x) = \theta(x) \tag{5.25}$$

where $\theta(x)$ is the Heaviside step function [106]. The cost function (the function that we want to minimize) C that we use for this example the Mean Squared Error (MSE):

$$C = \frac{1}{n}||a^L - y||^2 \tag{5.26}$$

The neural network is trained as follows. There is a forward pass that given some inputs will give us some outputs, see Equation 5.22. Then, the cost function C is computed using Equation 5.26. After this, there is a step called backpropagation: derivatives of C with respect to each of the free parameters are computed and the weights are modified in the direction of the negative gradient to minimize C.

5.3 Backpropagation derivation

For backpropagation the analytical derivatives of C are needed, see ref. [107] for an overview and derivation of the backpropagation algorithm. Large neural networks can have millions or even more parameters so finding the gradients numerically is not feasible, even for small neural networks. First, some direct derivatives are computed:

$$\frac{\partial C}{\partial a_i^L} = \frac{2}{n}(a_i^L - y_i) \tag{5.27}$$

$$\frac{\partial a_k^l}{\partial z_i^l} = \delta_{ik} g'(z_i^l) \tag{5.28}$$

$$\frac{\partial z_k^l}{\partial W_{ij}^l} = \delta_{ik} a_j^{l-1}, \quad \frac{\partial z_k^l}{\partial b_i^l} = \delta_{ki}, \quad \frac{\partial z_k^l}{\partial a_i^{l-1}} = W_{ki}^l \tag{5.29}$$

where δ_{ij} is the Kronecker delta [108].

Let's derive the equations for the backward pass. We apply the chain rule and use the convention that repeated indexes are summed:

$$\frac{\partial C}{\partial W_{ij}^l} = \frac{\partial C}{\partial a_k^L} \frac{\partial a_k^L}{\partial z_m^L} \frac{\partial z_m^L}{\partial W_{ij}^l} \tag{5.30}$$

$$\frac{\partial C}{\partial b_i^l} = \frac{\partial C}{\partial a_k^L} \frac{\partial a_k^L}{\partial z_m^L} \frac{\partial z_m^L}{\partial b_i^l} \tag{5.31}$$

When $l = L$ we can use the results obtained in Equation 5.29 to have:

$$\frac{\partial C}{\partial W_{ij}^L} = \frac{2}{n}(a_i^L - y_i)g'(z_i^L)a_j^{L-1} \tag{5.32}$$

$$\frac{\partial C}{\partial b_i^L} = \frac{2}{n}(a_i^L - y_i)g'(z_i^L) \tag{5.33}$$

If $l \neq L$ then we have to keep using the chain rule. Let's suppose that $l = L - 1$, we need the following result:

$$\frac{\partial z_m^L}{\partial W_{ij}^L} = \frac{\partial z_m^L}{\partial a_n^{L-1}} \frac{\partial a_n^{L-1}}{\partial z_p^{L-1}} \frac{\partial z_p^{L-1}}{\partial W_{ij}^{L-1}} = W_{mi}^L g'(z_i^{L-1})a_j^{L-2} \tag{5.34}$$

so that the complete formulas for the case $l = L - 1$ are:

$$\frac{\partial C}{\partial W_{ij}^{L-1}} = \frac{2}{n}(a_m^L - y_m)g'(z_m^L)W_{mi}^L g'(z_i^{L-1})a_j^{L-2} \tag{5.35}$$

$$\frac{\partial C}{\partial b_i^{L-1}} = \frac{2}{n}(a_m^L - y_m)g'(z_m^L)W_{mi}^L g'(z_i^{L-1}) \tag{5.36}$$

Comparing Equation 5.33 and Equation 5.36 a pattern can be extracted. For each layer starting from the last a term $W^l g'(z^{l-1})$ is added. We have all the tools to describe in detail the training algorithm.

5.4 Training algorithm

First, we define a baseline term for the gradients that will appear in each of the derivatives:

$$B = \frac{2}{m}(a^L - y)g'(z^L) \tag{5.37}$$

The algorithm is the following. First, the forward pass is done using Equation 5.22. Then, compute B using Equation 5.37. The derivatives of C with respect to W and b, ∇W^l and ∇b^l, are obtained for each layer and the weights are modified with $W := W - \alpha \nabla W$ and $b := b - \alpha \nabla b$, where α is the learning rate. The complete algorithm in pseudocode is given in Algorithm 1.

Last, the update rule for B (line 12), is obtained noting that in Equation 5.35 there is a summation over the index m that can be obtained with the dot product $(W^l)^T \cdot B$. The product with $g'(z^{l-1})$ is done element-wise.

Algorithm 1 One step of the training of a feedforward neural network

$a^0 \leftarrow X^T$
Choose a learning rate α
for all l in $1, \dots, L$ **do**:
    Forward propagation using $a^l = g(W^l a^{l-1} + b^l)$ ▷ Remember to save a^l and z^l
$B \leftarrow \frac{2}{m}(a^L - y)g'(z^L)$
for all l in $L, L-1, \dots, 1$ **do**:
    $\nabla W^l = B \cdot (a^{l-1})^T$
    $\nabla b^l =$ sum of B ▷ This sum is computed over the samples
    $W^l \leftarrow W^l - \alpha \nabla W^l$
    $b^l \leftarrow b^l - \alpha \nabla b^l$
    if $l \neq 1$ **then**
        $B \leftarrow (W^l)^T \cdot B g'(z^{l-1})$

5.5 Results

A neural network has been implemented to predict the muon signal with $4, 8, 12, 1$ units in each layer, see the left panel of Figure 5.7. The learning rate used is $\alpha = 0.01$ and the number of iterations or epochs is 500. The first layer has four units which is the number of features in the data. The last layer has one unit because a single quantity is being predicted. The results can be seen in the right panel of Figure 5.7 and the distribution of the differences between the true and predicted value in Figure 5.8 As it can be seen the performance for the training and development sets is very similar, likely because the neural network is a small one with not many free parameters and the amount of data used in the training set is large compared to the size of the network.

5.6 Real-world neural networks

There are a number of free and open source frameworks [109–111] that free the user from having to do most of the work done here. Depending on the usage, the user can only define the architecture, set the hyperparameters and let the frameworks do the rest of the work. These frameworks also allow the user to customize practically everything and have also implemented automatic differentiation, which allows to differentiate any function. Another benefit of using these frameworks is that they can be GPU-accelerated or parallelized on multiple machines, yielding speed ups of many orders of magnitude over a hand-made approach.

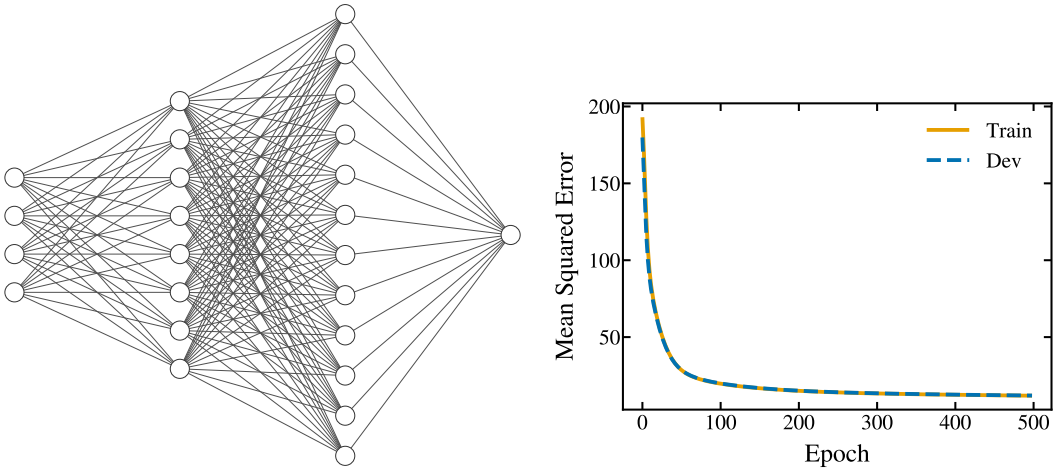

Figure 5.7: Left: Architecture of the neural network used for the example. Right: Cost computed for both the training and development sets while training is going on. Each epoch corresponds to a forward pass and a backward pass.

6 Machine learning in this thesis

The next chapters deal with the problem of predicting the muon signal using a neural network. The main ideas studied in this chapter are useful to understand how the neural network works, although the focus is on the physics problem. The process followed in the next chapters is very similar to the one followed in this chapter: data from simulations is split in training, development or validation and test sets; a function to be minimized is chosen for the fit and the training algorithm is performed on the training data and its performance evaluated on the development and test set. No equations of backpropagation are obtained since we use one of the open source frameworks mentioned above.

This chapter has another purpose. Sometimes neural networks are treated or thought as a black box that given some input will output something. While the exact reason why a complicated method from machine learning makes a prediction over another is hard to know, we have seen that the methods follow well defined mathematical rules. We implemented a simple neural network from scratch in this chapter to predict the muon signal. In practice, when using one of the open source frameworks, it is not necessary to define the operations as there are predefined layers, favouring the black box approach.

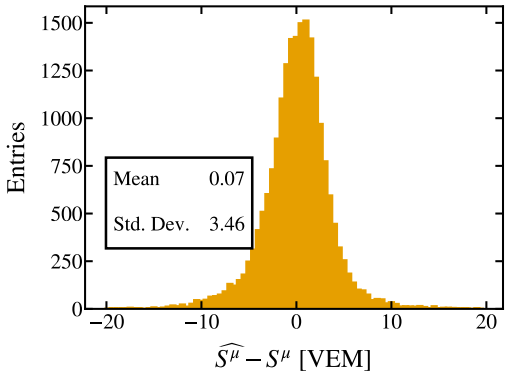

Figure 5.8: Difference between the predicted and true values of the muon signal using a neural network.

6

Extracting the muon component with neural networks:

Total muon signal

The collision of a UHECR with a molecule of air at the top of the atmosphere produces a shower of secondary particles. When the shower reaches the ground, that shower is mainly composed of photons, electrons, positrons and muons. The SD of the Pierre Auger Observatory measures the time of arrival and signal left by those particles. However, it can not distinguish which part of the signal comes from the electromagnetic cascade and which part from the muon cascade.

Determining the signal left by muons is a very powerful tool to study the mass composition of UHECR. Using both the muon signal together with the position of the shower maximum X_{max} it could be possible to infer the mass composition on an event-by-event basis [112]. We have developed a method to separate the integral of the muon component of the signal registered by the SD using Neural Networks.

This chapter begins with an introduction to the problem of extracting the muon signal from the physics point of view in Section 1. Afterwards, the method is described in Section 2 and the data used in Section 3. The performance of the neural network trained on simulations is shown in Section 4. We conclude with a short summary and the conclusions of this study in Section 5.

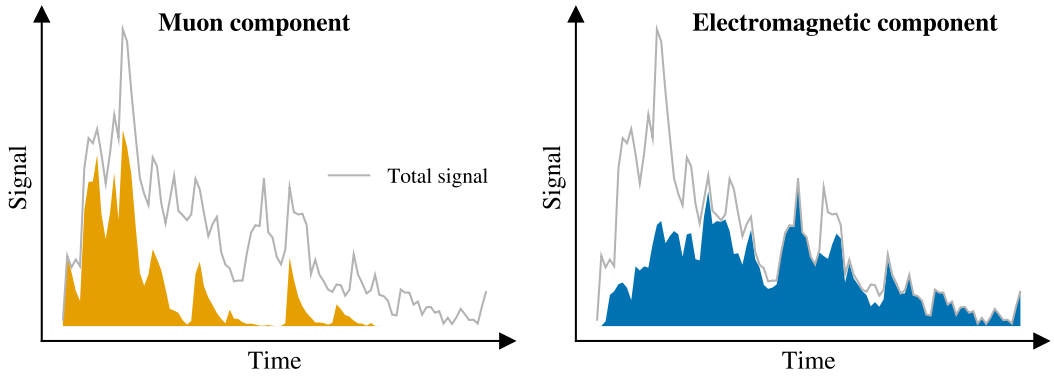

Figure 6.1: Signal in time measured by the one of the stations of the SD for a simulated event. The total signal corresponds to the signal measured by the SD and is the sum of the electromagnetic and muon components. The muon (left) and electromagnetic (right) components are shown independently; this information is available in simulations. The total trace is the sum of the muon component and the electromagnetic component. Our goal in this chapter is to obtain the integral of the muon component.

1 Introduction

When charged particles traverse the stations of the SD, they leave a signal which is a mixture of electromagnetic particles (photons, electrons and positrons) and muons, see Figure 6.1. The determination of the muon component in the cascade is crucial to infer, on an event by event basis, the mass composition of the recorded data. This is the main motivation behind the upgrade of the Pierre Auger Observatory, dubbed AugerPrime [112].

As it can be seen in Figure 6.1, the electromagnetic and muon signals are different. Muons usually arrive earlier and leave spikier signals than the electromagnetic particles. The signal that they produce also extends less in time than the electromagnetic signal. However, this is not enough to have an accurate prediction of the muon signal. Only when studying inclined events, where electromagnetic particles are absorbed and the signal is almost purely muonic, the muon signal can be accurately measured with the current design of the SD [53].

In this chapter we use the information available in simulations to build and fit a model that can predict the integral of the muon signal. The model that we use is a neural network with fully connected layers, such as the one studied in the previous chapter on page 87. The large amount of simulations available allows to efficiently train this method and extract how the muon signal behaves as a function of the different variables. The goal of this study is to estimate, for every single triggered detector used in the reconstruction of an event,

the amount of signal that corresponds to the energy deposited by muons.

2 Neural Network: technical details

We explain the technical aspects related to the Deep Neural Network (DNN) that we have built, comprised of fully connected layers. There are several possible choices when building a DNN: the number of neurons in each layer, the number of layers, the activation function, etc. We employ a genetic algorithm to chose these hyperparameters and, in the end, we give an overview of the final architecture of the DNN that we used. The code has been implemented using Numpy and Matplotlib from the SciPy ecosystem [113], Scikit-learn [114] and Keras [115] using Tensorflow [110]. The programming language used is Python [116] (version 3.6).

2.1 Data preprocessing

Each feature of the data used as input and the output, the muon signal S^{μ}, have been normalized each one independently so that their distributions have mean $\mu = 0$ and standard deviation $\sigma = 1$. Normalizing data is a usual practice in machine learning and it helps to avoid computational problems associated dealing with large and small numbers at the same time. Before normalizing, some variables have been transformed to help the DNN learn. The details are explained in Section 3.

2.2 Activation function and weight initialization

As it has been explained in the previous chapter, in a fully connected layer each neuron computes a function of the linear combination of its inputs. There is a wide variety of functions that can be chosen, although following the recommendations in refs. [117,118], the Rectifier Linear Unit (ReLU, see Equation 5.25) is a good choice and one of the most used activation functions. Nonetheless, instead of choosing it ourselves we have let the genetic algorithm chose the one that gave the best performance among the ones we tested. The list of functions that we have included is the following: linear or $f(x) = x$, tanh, softmax [119], ReLU and SELU [120].

Regarding the initialization of the weights of the neural network, we have obtained the bests results using a random weight initialization that draws the values of the weights from a normal distribution with mean $\mu = 0$ and standard deviation $\sigma = 0.05$.

2.3 Loss function and optimization algorithm

We use the Mean Squared Error (MSE) to determine if the model has a good performance:

$$\mathrm{MSE} = \frac{1}{n}\sum_{i=1}^{n}\left(\widehat{S^{\mu}}_i - S^{\mu}_i\right)^2 \tag{6.1}$$

where n is the number of samples, S^{μ}_i is the true value that we want to predict for the i-th example and $\widehat{S^{\mu}}_i$ is the prediction done by the neural network for the same example. This is the function that will be minimized during the training process.

The optimization algorithm used is Adam [121], which is a stochastic gradient-based method. It is the recommended algorithm in many deep learning applications. We followed the common usage of running the algorithm using the default parameters recommended in ref. [121]. The gradients are computed and then corrected using a first and second raw moment estimate, leading to the new value of the parameters to be optimized.

2.4 Genetic algorithm

Genetic algorithms (GAs) are global optimization algorithms that have been widely used in many areas as an improvement over a random search [122]. The idea beneath these algorithms is to imitate the behaviour of nature by evolving populations of individuals through time, in a way that only the characteristics, also called genes, of the fittest individuals are propagated into the next generation of the population, see Figure 6.2. We have let the number of neurons in each layer, depth or number of hidden layers and choice of activation function free for the genetic algorithm to pick the best combination.

The first step in a GA is to decide what to choose as the individual that will be evolved over time. We have defined our individual as a DNN, and represented it as a vector of a certain length, where each element corresponds to the number of neurons in each layer in the allowed interval $[0, 100]$. This vector also has an integer that maps the activation function (see the list in subsection 2.2), and therefore, the algorithm can check different combinations of them. The number of hidden layers was limited to a maximum of ten.

We generated randomly 50 individuals with a random number of neurons in each layer, a random number of layers and random activation functions from the possible choices. We have repeated the following process for 100 generations. To compute the fitness of each individual, networks are trained over ten epochs to have an estimation of their potential performance. The validation is done over a sample independent of the one used for training. Once they are evaluated, a binary tournament selection is carried out to obtain the best individuals, based on their performance on the validation sample. Then, there is a

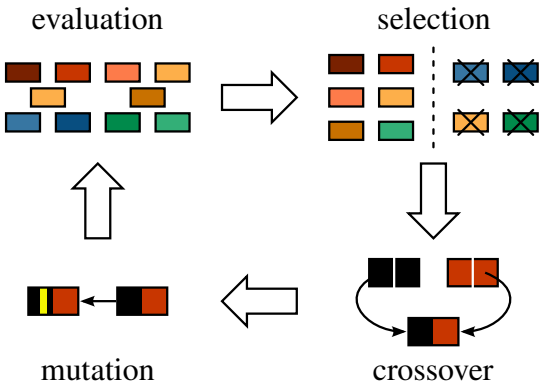

Figure 6.2: Diagram of the genetic algorithm used in this work. DNNs are built with random numbers of hidden layers and neurons per layer and trained. The DNNs with better performance are selected in binary tournaments. Only for the selected DNN, we cross the number of neurons in some layers between individuals and we introduce mutations (random changes of the number of neurons in certain layers). This process provides a new generation of DNN ready to be trained.

two-point crossover, that is, the number of neurons in each layer from a certain start layer l_0 up to an end layer l_1 are exchanged between two individuals with probability 0.8. The last step is to mutate or change randomly, with probability 0.1, the number of neurons in each layer of an individual. When the new population is available, we have used an elitism mechanism where the best individual in the current population is included into the next one. All these values were set up according to the literature, following the same design principles discussed in refs. [123–126] and after checking, by doing experiments, that other values did not produce a significant improvement.

When any layer was found to have less than five neurons, it was discarded. In this way, we make sure that the layer is really needed and there is no need to add dropout or apply regularization afterwards. The last layer only has one neuron and is used to obtain the output.

2.5 Final DNN structure

The final DNN obtained is represented in Figure 6.3. It is the outcome of running over a mix of equal fractions of proton, helium, nitrogen and iron nuclei generated with QGSJetII-04. The network is made up of six layers: five fully connected layers using ReLU as activation function and a final layer that combines the outputs from the fifth layer linearly. In this neural network, complexity starts increasing from low to high (from 9 to 56 neurons) and then it decreases to the point where it started. The first runs of the GA, using a maximum of ten layers, always decreased the number of them to six or seven, so we fixed the

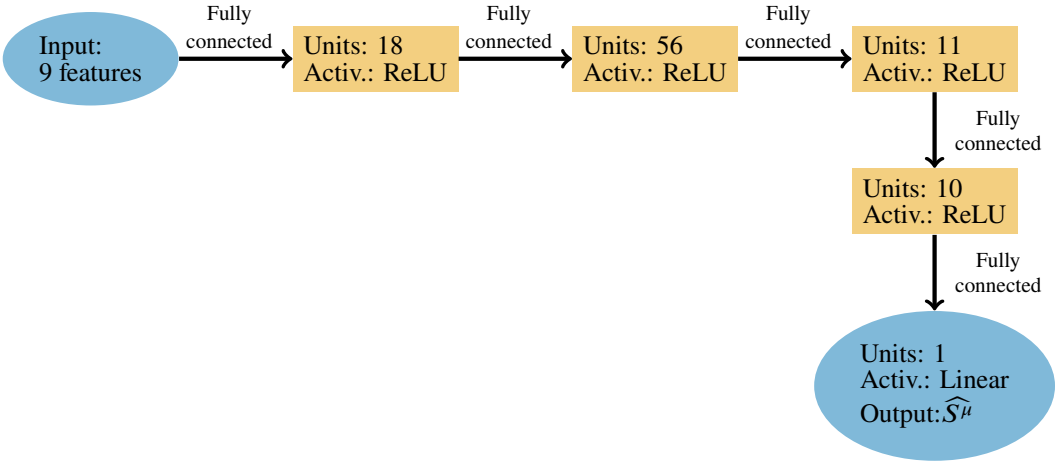

Figure 6.3: Final structure of the DNN coming from the GA optimization. It has five fully connected layers using ReLU as activation function, except for the last layer, for which a linear activation function ($f(x) = x$) is used. The numbers of neurons in each layer are indicated.

number of layers to six to simplify the structure of the DNN without losing performance. We executed the GA again (with a maximum number of layers set to six) obtaining the current configuration. In this way, once we explored complex solutions regarding the network depth, we exploited the reduced solution space to determine more precisely how many neurons per layer should be included.

3 Input variables and data

We feed the neural network with the set of variables described below. Notice that some of them pertain to the global features of the event (items 1 and 2 in the list below), while others refer specifically to the information at the station level. The total trace used in our analysis is computed as the average of the traces recorded by each of the active PMTs of a station. In our analysis we do not use detectors with signs of saturated traces. We disregard as well stations whose measured total signal is below 10 VEM (above this value the probability of single detector triggering is close to 100%). We do not work with events whose energy lies below $3 \cdot 10^{18}$ eV since this is the minimum energy at which the Auger trigger efficiency reaches 100% for the events recorded by the 1500 m array. The event selection efficiency is above 95% for the energy interval of interest ($\geq 3 \cdot 10^{18}$ eV). The chosen list of variables reads as follows:

1. Monte Carlo energy: E_{MC}. This is the only Monte Carlo variable used. Since this study deals only with simulated events, we did not use reconstructed energies to

avoid all the discussion associated to the energy calibration of SD events. We use as input variable the decimal logarithm of energy, with the energy expressed in eV.

2. Zenith angle: θ. It is measured in degrees. For this study and due to the lack of simulated horizontal events, we limit ourselves to angles below 45 degrees.

3. Distance of the station to the reconstructed core of the air-shower r. We measure it in meters.

4. Total signal registered by the station S. We measure it in VEMs.

5. Trace length. Number of 25 ns bins between the Start and End bins of the trace [127].

6. Polar angle of the detector with respect to the direction of the shower axis projected on to the ground ζ. It is measured in radians.

7. Risetime, $t_{1/2}$ (see Chapter 3).

8. Falltime. Similarly to the risetime, it is the time for the integrated signal to go from 50% to 90% of its total value and it is measured in ns.

9. Area over peak. It is defined as the ratio of the integral of the trace to its peak value.

The set of mixed global and local (station level) variables listed above performs well. Enlarging it with new variables to obtain a better estimation of the muon signal just represents an increase of the training time but does not improve our results. In particular, one can think that feeding the net with the whole temporal series of bins that form the trace would substantially improve the precision with which the muon signal is extracted. We observed that in that case the gain is almost negligible, while the computational cost in terms of time increases by a sizeable factor. We understood that to fully exploit the time information contained in the recorded traces we have to consider new classes of artificial neural networks. In the next chapter we study the use of recurrent neural networks to try and extract not only the total muon signal but also its time structure.

We decided to train the network with events generated using QGSJetII-04. The events have been simulated with a distribution that is uniform in energy. Once the network is trained and its performance has been evaluated with QGSJetII-04 events, we use EPOS-LHC to show how our results depend on the different assumptions made to model hadronic interactions at ultra-high energies. Another important decision is related to the nature of the nucleus or set of nuclei used to train the neural network. We observed that training the network with a single species is a far from optimal decision. We show the values of the MSE, our evaluation metric, and the mean values of the distributions obtained as the difference between the predicted and the true muonic signals in Table 6.1. For example, when iron nuclei are used to build the network, the number of predicted muons in events

Training sample	MSE	Mean (p, VEM)	Mean (Fe, VEM)
Pure p	6.8	0.16	−0.26
Pure Fe	7.5	0.33	−0.18
Mix 25% (p, He, N, Fe)	6.4	0.08	−0.16

Table 6.1: Performance of the neural network when different compositions are used for the training sample. We use QGSJetII-04 as the model to simulate hadronic interactions. The third and fourth columns show the mean of the distributions of predicted minus true muon signals, measured in VEM, for proton and iron nuclei, respectively.

QGSJetII-04				
		Training	Validation	Test
Primary	Number of events	Number of traces		
Proton	19362	16088	4022	57522
Helium	12341	15960	3989	34314
Nitrogen	12201	16071	4017	33739
Iron	19478	16076	4018	65231
EPOS-LHC				
		Training	Validation	Test
Primary	Number of events	Number of traces		
Proton	18456	–	–	78063
Iron	18779	–	–	86862

Table 6.2: Summary of the number of simulated events and traces used in this work. Notice that the batches of stations used for training, validation and test correspond to different sets of detectors.

generated by protons is overestimated. Similar conclusions are drawn when a pure sample of protons is used to train the network. In this case we underestimate the muonic signals produced by Fe. The situation improves when the network is trained with a mix of iron nuclei and protons in equal amounts. However, we observed that our estimations of the muonic signal improve even more when a mix of equal fractions of proton, helium, nitrogen and iron nuclei is used in the final training sample. As shown in Table 6.1, this is the combination that offers the best performance.

The numbers of events generated and the stations used are shown in Table 6.2 for two models of hadronic interactions. QGSJetII-04 events are used to train, validate and test the neural network. EPOS-LHC events are used for testing purposes only. The validation sample is used to choose the model that works best and also to study whether the learning process shows signs of overfitting, something that does not occur in the case under study. The events have been simulated using the CORSIKA package [128] version 74004 and reconstructed using an official version of Offline [127] .

4 Results

Once the model described in the previous sections was optimized to extract an estimation of the muon signal recorded by the stations, we show to it samples of simulated events generated with QGSJetII-04 and EPOS-LHC. Before discussing the outcome of this procedure, we note that this method can be readily applied to experimental data. The event selection efficiency is, as for the case of simulated events, close to one and we observed that the performance of the method is similar when the reconstructed energy is used. The only problem comes when interpreting the results, given the known inconsistencies between data and simulations (see subsection 3.4 on page 14).

4.1 QGSJetII-04 simulations

The distribution of muonic signals for a set of events generated with QGSJetII-04 is shown in Figure 6.4. They have energies higher than $10^{18.5}$ eV and zenith angles up to 45 degrees. The figures in this section show results at the single station level. For each species, we find that the distributions of predicted signals reproduce reasonable well the true signal distributions. This is illustrated in Figure 6.5 where the difference between predicted and true signals is plotted for every nuclei. We obtain Gaussian distributions with means very close to zero and standard deviations around to 2.5 VEM. The accuracy in the prediction of the muonic signal is shown, this time as a scatter plot, in Figure 6.6. The Pearson correlation coefficient is 0.98 for p, He, N and Fe.

We have checked whether potential biases arise in the estimation of the muonic signal as a function of the following variables: distance of the station to the position of the core at the ground (Figure 6.7), simulated energy of the event (Figure 6.8), $\sec\theta$ (Figure 6.9) and the total signal recorded by every water-Cherenkov detector (Figure 6.10). For this set of variables, the mean of the differences (in absolute value) between true and predicted signals are most of the time below 2 VEM. Relative errors are typically below 10%. Our predictive power does not depend on the energy or the zenith angle of the air shower since the differences between predicted and measured signals are flat as a function of those two variables. At distances close to the core the spread in differences is wider. In addition to the fact that the number of events is smaller at short and very large distances to the shower core, we attribute part of this behaviour to the presence of a stronger contribution of the electromagnetic component.

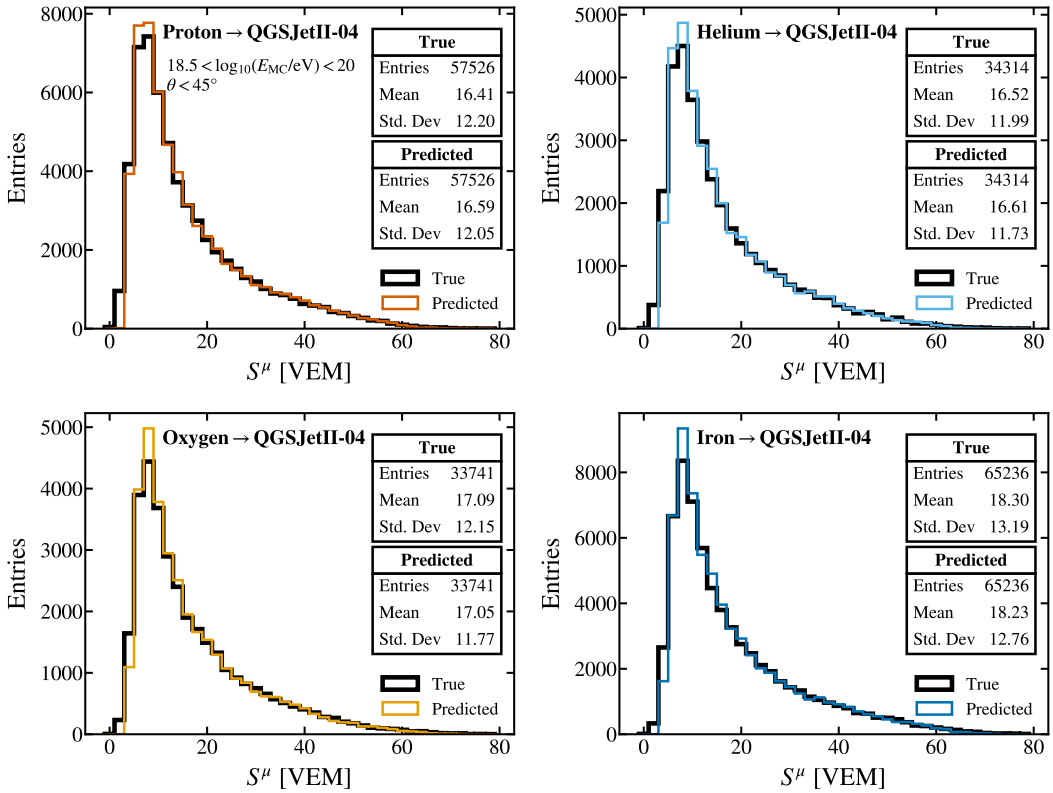

Figure 6.4: True and predicted muon signals for four different species of primary nuclei. Each entry corresponds to the muonic trace recorded by each individual detector. The events have been generated in an energy interval that spans from $\log_{10}(E/\text{eV})$=18.5 up to $\log_{10}(E/\text{eV})$=20. Simulations use QGSJetII-04 to model hadronic interactions.

4.2 EPOS-LHC simulations

A crucial test our approach must overcome is to prove that the neural network is capable of accurately estimating the muon signal in a detector independently of the hadronic model used. With this goal in mind, we generated a sample of events that used EPOS-LHC as the model for hadronic interactions, see Table 6.2. The results of this exercise are shown in Figure 6.11–Figure 6.16. These plots illustrate the robustness of the final DNN to a change in the hadronic models used for testing its performance. The relative error stays below 10%, and the absolute difference between predicted and true signal does not exceed 2 VEM units. In addition, no sensible bias occurs as a function of the set of variables previously checked. We interpret this as a sign that the correlations between the variables used in the models under consideration are similar and therefore show a high degree of

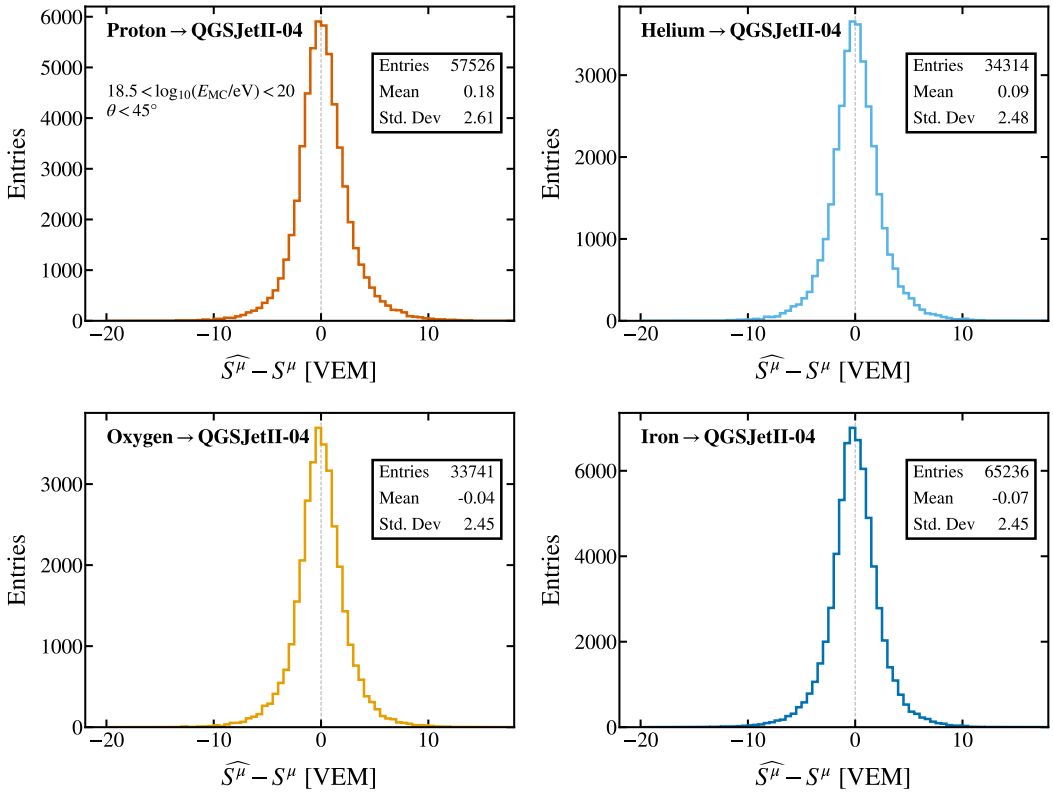

Figure 6.5: Difference between predicted and true muon signals for four different kinds of primaries. Every entry corresponds to the information provided by a single detector. Events have been generated using QGSJetII-04 to model hadronic interactions.

universality.

5 Summary and Conclusions

We have demonstrated that deep learning methods are a powerful tool to estimate the muon signal fraction in the traces registered by the water-Cherenkov detectors of the Auger Surface Detector Array. Based on Monte Carlo studies, we have proven that, for each individual station, we obtain accuracies that are typically better than 10%. Our method can be applied to a wide spectrum of primary nuclei, energies, zenith angles and distance ranges. It has also been proven that it is independent of the model for hadronic interactions used in simulations.

This work was published in ref. [129].

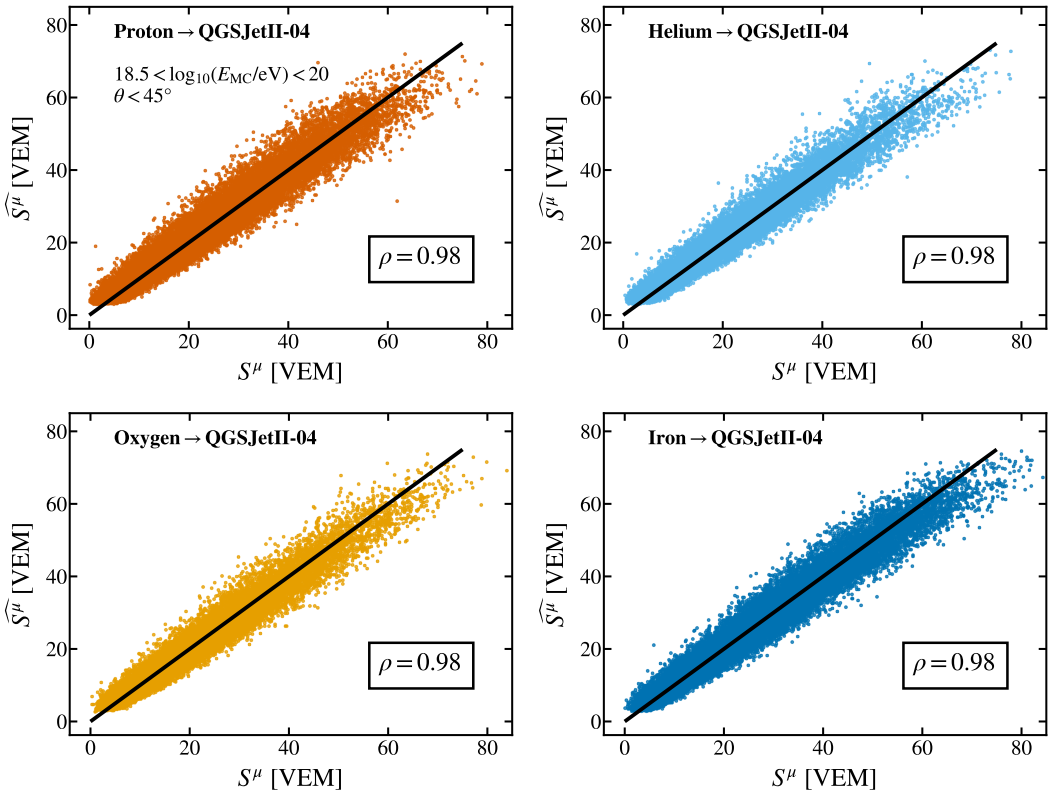

Figure 6.6: Correlation between the predicted and the true muon signals. Events have been generated using QGSJetII-04 to model hadronic interactions.

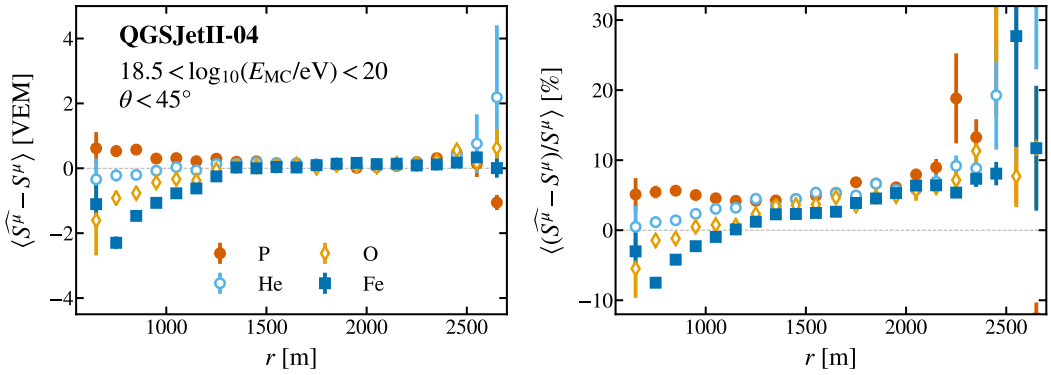

Figure 6.7: Mean of the distribution of differences between predicted and true muon signals as a function of the distance to the core (left) and its associated relative error (right). Events have been generated using QGSJetII-04 to model hadronic interactions.

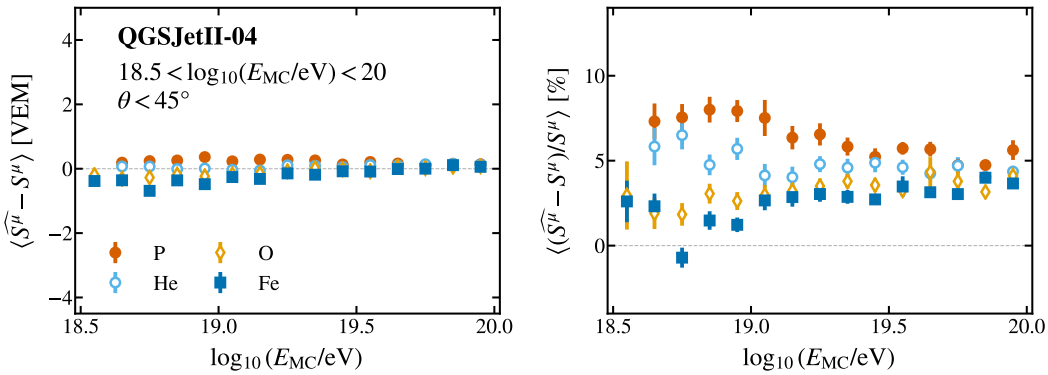

Figure 6.8: Mean of the distribution of differences between predicted and true muon signals as a function of the event Monte Carlo energy (left) and its associated relative error (right). Events have been generated using QGSJetII-04 to model hadronic interactions.

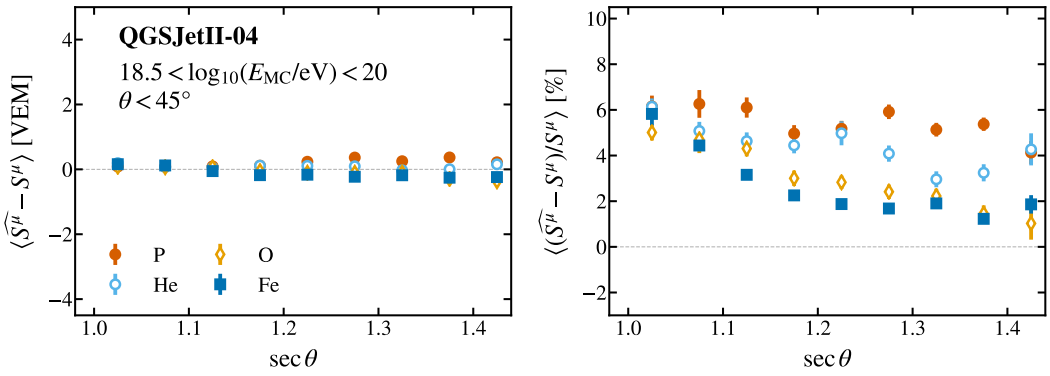

Figure 6.9: Mean of the distribution of differences between predicted and true muon signals as a function of $\sec\theta$ (left) and its associated relative error (right). Events have been generated using QGSJetII-04 to model hadronic interactions.

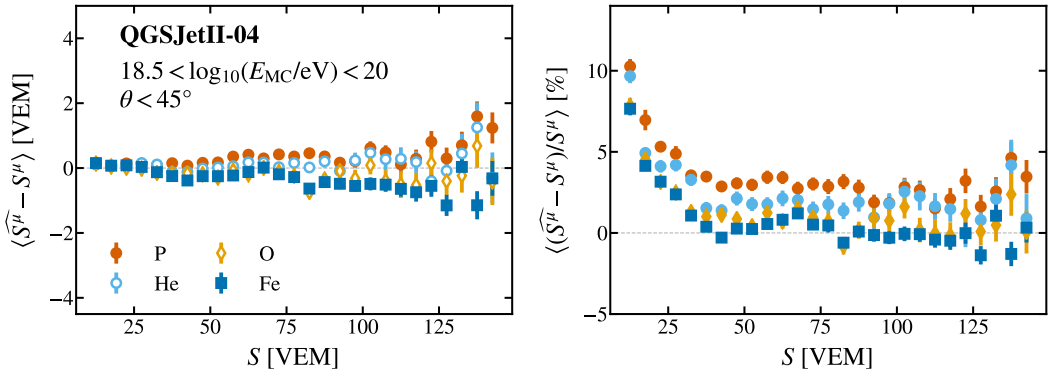

Figure 6.10: Mean of the distribution of differences between predicted and true muon signals as a function of the total signal registered in individual stations (left) and its associated relative error (right). Events have been generated using QGSJetII-04 to model hadronic interactions.

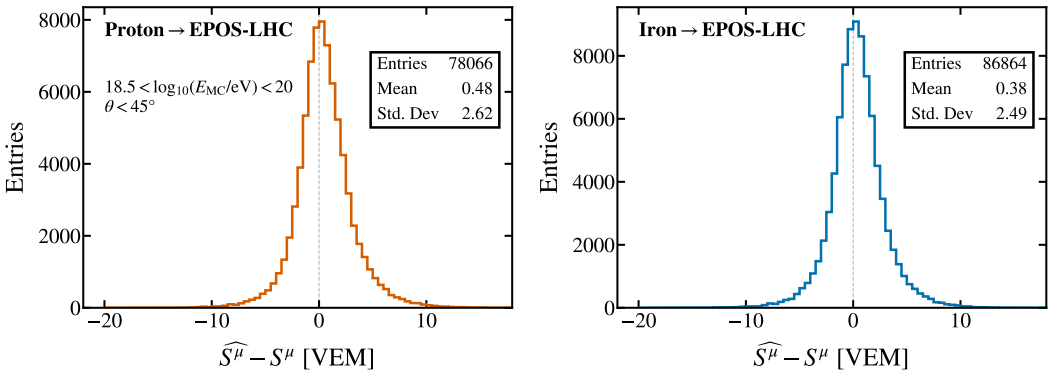

Figure 6.11: Difference between predicted and true muonic signals at detector level for two different kinds of primaries. Events have been generated using EPOS-LHC to model hadronic interactions.

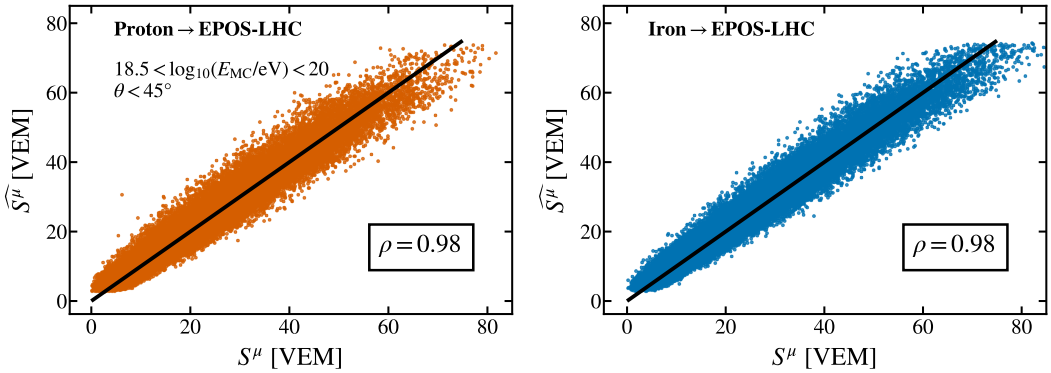

Figure 6.12: Correlation between the predicted muon signal and the true muon signal. Events have been generated using EPOS-LHC to model hadronic interactions.

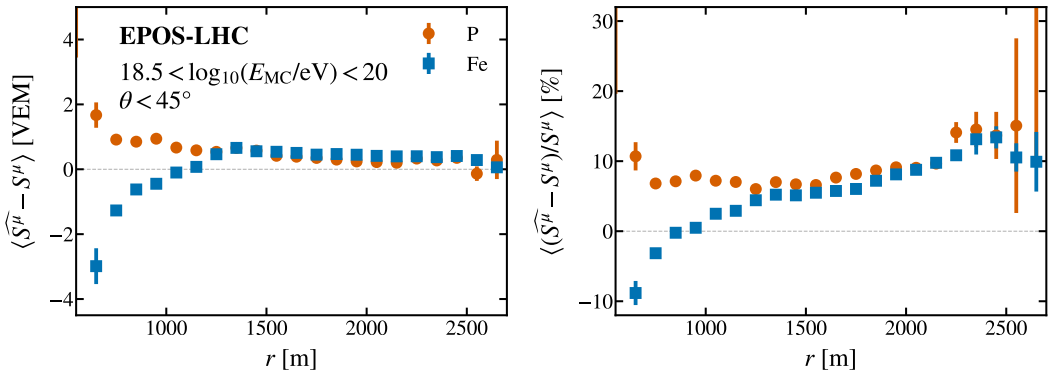

Figure 6.13: Mean of the distribution of differences between predicted and true muon signals as a function of the distance to the core (left) and its associated relative error (right). Events have been generated using EPOS-LHC to model hadronic interactions.

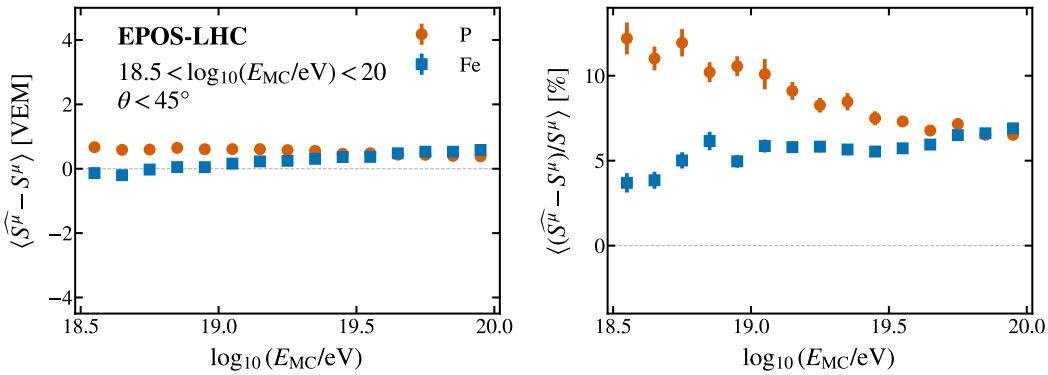

Figure 6.14: Mean of the distribution of differences between predicted and true muon signals as a function of the Monte Carlo event energy (left) and its associated relative error (right). Events have been generated using EPOS-LHC to model hadronic interactions.

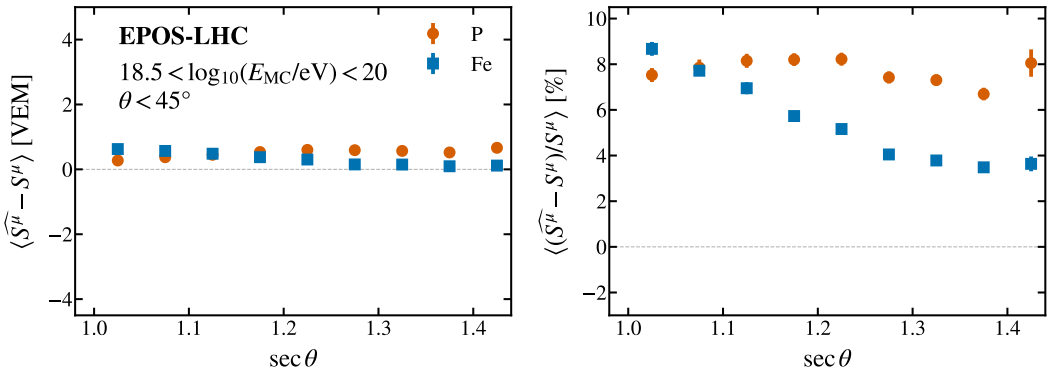

Figure 6.15: Mean of the distribution of differences between true and predicted muon signals as a function of $\sec\theta$ (left) and its associated relative error (right). Events have been generated using EPOS-LHC to model hadronic interactions.

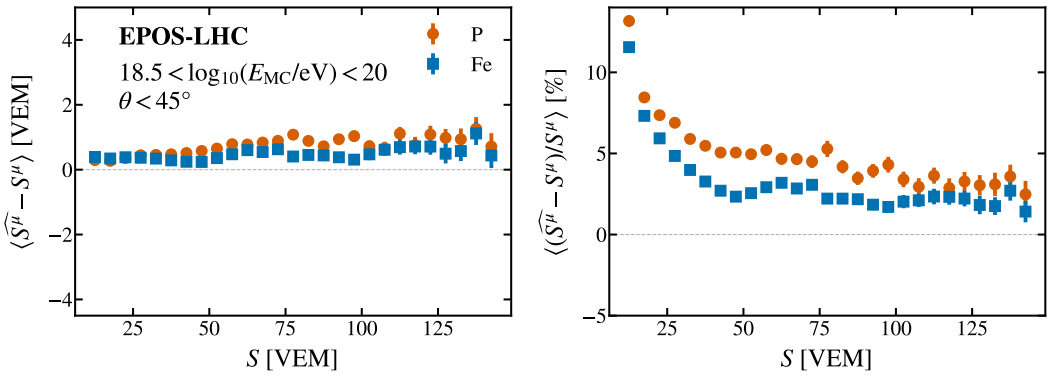

Figure 6.16: Mean of the distribution of differences between true and predicted muon signals as a function of the total signal registered in individual stations (left) and its associated relative error (right). Events have been generated using EPOS-LHC to model hadronic interactions.

7

Extracting the muon component with neural networks:

Temporal muon signal

In the previous chapter we presented a method to extract the total muon signal recorded by individual stations of the SD. We now present a method to obtain the full muon time signal using modern techniques from the field of machine learning: neural networks. This method is much more powerful since both the temporal distribution and the total signal can be obtained from its predictions, by integrating the predicted temporal series.

This chapter is organized as follows. In Section 1 we explain the method. Results on simulations are given in Section 2 and some preliminary results on data are given in Section 3. Then, the predictions for data are fitted in Section 4 using functions obtained from measurements done in different UHECR experiments. The chapter ends with a short summary and conclusions in Section 5.

1 Method

Our approach is based on a kind of Neural Network known as Recurrent Neural Network (RNN). RNNs are specially well suited for time series because they have a memory mechanism. As the computing process develops, for each step of the temporal series the RNN stores information from a preceding time slot that can be used at a later stage. Nowadays, RNNs are used successfully in different fields such as natural language processing or machine translation, see ref. [130]. There are several kinds of RNNs and one of the most common is called Long Short-Term Memory (LSTM). Our neural network is based on this kind of RNN.

1.1 Long Short-Term Memory layer

Long Short-Term Memory (LSTM) layers were proposed in 1997 [131]. At those times it was hard to make use of these in neural networks, as they require a lot of computing power. The basic idea is that they operate on sequences, such as temporal sequences and also other sequences like language fragments. They keep information from previous steps in the sequences. Even though RNNs also do this, LSTM are well suited to keep information from many previous steps, allowing for the treatment of long sequences. A detailed description of the operations and the parameters of a LSTM layer is given.

LSTM layers have four matrices and four vectors of free parameters, denoted as W_f, W_i, W_C and W_o and b_f, b_i, b_C and b_o respectively. These layers work on sequences of data, see ref. [132]. For a detailed description of the operations see Figure 7.1. The layers keep some information in a cell state C_t and a hidden state h_t, where t is the index that represents the steps in the sequence. The index f is associated to a *forget* gate. This gate decides what to keep or forget from the previous cell state C_{t-1}. The index i is associated to an *input* gate. This gate decides what to keep from a candidate cell state $\tilde{C}_t$. The index o is associated to an *output* gate that selects what to output from the cell state C_t. The final result of this gate is two vectors: the vector of cell states C_t and another vector having the hidden states h_t.

1.2 Input

For the input of our neural network we use the traces recorded by the SD. Each station is equipped with 3 PMTs and the traces used in this study are obtained after averaging the

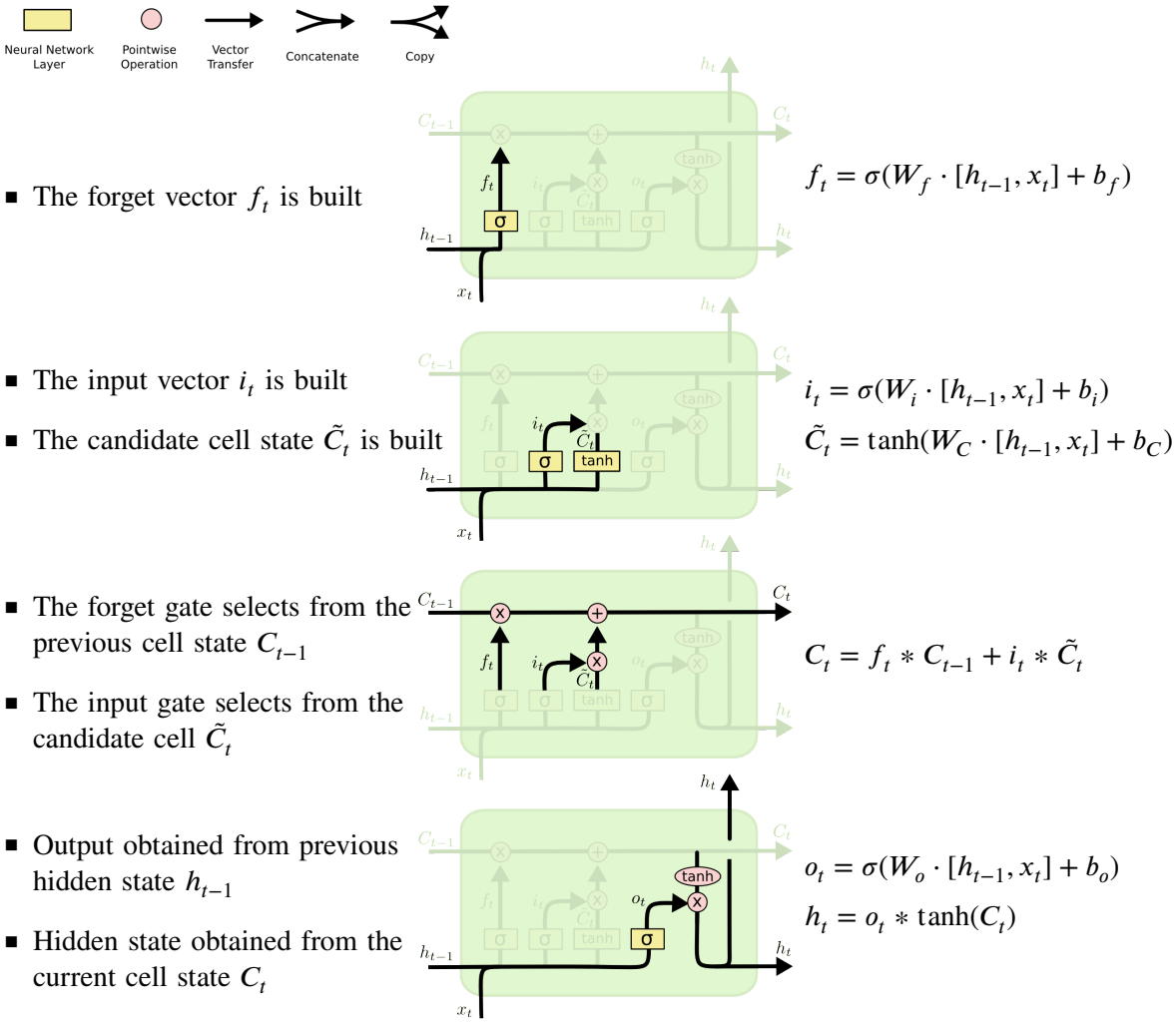

Figure 7.1: Description of the operations performed in a LSTM layer, with x_t being the input of the layer (a temporal sequence) and o_t being the output of the layer (another temporal sequence). [] means concatenation, $*$ means elementwise product and $\cdot$ is the regular matrix product. The functions σ and tanh are applied elementwise and σ is the sigmoid function.

trace of each active PMT. A study conducted using the traces provided by each individual PMT did not produced better results than those presented in the next sections.

We use only the first 200 bins of each trace. Traces that are shorter than those 200 bins are padded with zeros at the end. Most of the relevant information is encapsulated in the first bins of each trace and this is specially true for muons, which arrive earlier than the electromagnetic component. From simulations we know that for $E < 10^{19}$ eV, around 90% of the stations have the complete muon signal in the first 200 bins and the rest have more than 99% of the muon signal in those 200 bins. For $E > 10^{19}$ eV around 70% of the

stations have the complete muon signal in the first 200 bins and the rest have around 99% of the muon signal in those 200 bins. The fraction of muon signal outside the first 200 bins is then negligible and can be ignored.

Besides the traces, we also use as input the secant of the zenith angle of the event $\sec\theta$ and the distance to the core of each station r. We saw that only the trace information was not enough to determine the muon component without bias. These two variables are related to the amount of atmosphere that the particles go through, which plays an important role, since the electromagnetic and muonic component are attenuated differently in the atmosphere. Particles from the electromagnetic cascade are attenuated much more quickly than muons. In fact, muons are very penetrating particles that traverse the atmosphere practically unaffected. This is why they also arrive earlier than eletromagnetic particles. Muons are typically minimum ionizing particles, consequently, as the attenuation of electrons, positrons and photons increases, the traces richer in muons become spikier and shorter in time.

Both variables take into account the amount of atmosphere crossed by the particles. As $\sec\theta$ increases, the shower becomes more inclined and the amount of traversed atmosphere is larger because the thickness of the atmosphere is proportional to $\sec\theta$. As r increases, the station is further away from the shower core and, therefore, the larger the distance travelled by particles.

1.3 Neural Network (NN) architecture

The neural network architecture is as follows. The input is a vector of 202 components: r, $\sec\theta$ and the 200 values of each trace $S_1, S_2, \ldots, S_{200}$, where S_i is the value of the signal measured at the time bin t_i. This input is split at the start into two sets. One of them is the set of the time-independent variables, i.e. r and $\sec\theta$. These variables are fed into two identical sets of dense or fully connected layers (explained on page 87) that will compute the initial values of the parameters for the first layer of LSTMs. The outputs of the dense layers together with the 200 time values of each trace form the input to the LSTMs. The LSTMs produce 70, 32 and 32 sequences of 200 numbers and the last of these sequences is fed to a final dense layer. This architecture is depicted in Figure 7.2. The model has a total of 87 212 trainable or free parameters.

The set of dense layers computes the vector of initial parameters that encodes information about the amount of atmosphere crossed by the particles, which in turn changes the shape of the traces. Without this block, the neural network does not have enough information to distinguish between traces with high or low fractions of muons and will be biased: it will underestimate the muon component for showers with a large value of $\sec\theta$ and overestimate it for small values of $\sec\theta$. The block of LSTMs is responsible for computing the traces and using the temporal information of the input.

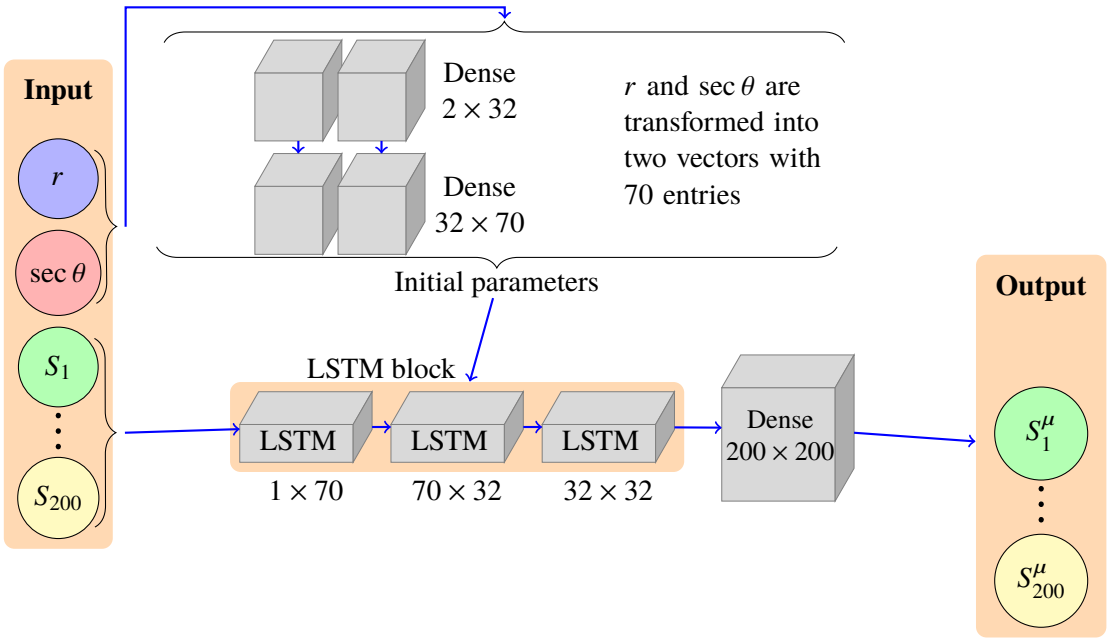

Figure 7.2: Schematic drawing of the input, architecture and output of the neural network and output. See the text for details.

1.4 Data selection and training

The neural network is trained with simulated showers. The energy range covered by simulations spans from $10^{18.5}$ eV to $10^{20.2}$ eV for zenith angles up to 60°. All events selected are 6T5. The method is applied to traces whose signals do not show any sign of saturation and whose integral is more than 5 VEM.

The training was done with simulations using EPOS-LHC. Simulations were done with the CORSIKA version 7.3700 and reconstructed using the official software of the Pierre Auger Collaboration: Offline version v3r3p4-icrc2017-preprod-v3. The simulated showers are initiated by proton, helium, oxygen and iron and the library of simulations that we use has the same number of showers inintiated by each nuclei. A total of 450 434 events were used. The events were sampled randomly and assigned to the training, validation or test datasets, using a uniform distribution in energy and $\sec\theta$ for the validation and test sets and the rest of the events for the training set. The training dataset does not require special care regarding the energy and zenith angle distributions since the dependence on the energy is mild and the zenith angle is given as input. The whole event sample was split as follows: 393 994 in the training set, 22440 in the development set and 34000 in the test set.

Before training, both r and $\sec\theta$ values are scaled to be between 0 and 1 and all the

traces are scaled individually to be between 0 and 1. For each station, the true muon trace is scaled by a unique factor. The output of the neural network is also between 0 and 1 so the same factor is used to rescale back the predicted trace. The function that is minimized in the process of training is the mean square error, defined for a single trace as

$$L = \frac{1}{200} \sum_{t=1}^{200} \left(\widehat{S_t^\mu} - S_t^\mu \right)^2 , \tag{7.1}$$

that is, the average of the squares of the differences between the predicted muon trace $\widehat{S_t^\mu}$ and the true muon trace S_t^μ, for each time bin of 25 ns[1]. The neural network is trained in batches and for each batch the value of L is computed for each trace and averaged over all the samples in the batch.

The training was done with the optimizer ADAM with a fixed learning rate of 10^{-4} and the default values for the rest of the parameters, see [121]. Using a batch size of 512 and 150 epochs on a Nvidia Titan V, the training takes around 8 hours. The loss as a function of the epoch is shown in Figure 7.3 for both the training and validation sets. We can see that it decreases as the epoch increases. The curve for validation is below the one for training because, as explained before, a uniform distribution has been used for the validation dataset and there are more events for which the performance is worse (lower zenith angles) in the training set. As an additional measure of goodness, the difference between the integral of the predicted and true muon trace is computed after each epoch. In the right panel of Figure 7.3, we show the mean and the standard deviation of the distribution of the difference. We use this measure because it has a straightforward physics interpretation. The mean controls the bias: i.e., if the mean is above zero the neural network is consistently overestimating and vice versa. The standard deviation tells us how well we are predicting the muon signal, although this depends on several factors such as the zenith angle θ, as we will see in Section 2. All the pipeline was implemented in Python 3.8 using numpy, scipy, pandas and Pytorch 1.5.0.

2 Results

We show some examples of the NN predictions in Figure 7.4 with more examples on page 148. We observe that, qualitatively, the prediction follows the shape and peaks of the total signal. The network has learnt to reproduce the main features of the muon trace:

[1] We use a hat $\widehat{\ }$ for all the quantities that are predicted or computed from a prediction from the neural network. For example, the integral of the true muon signal in simulations is S^μ, while the integral of the predicted muon signal is $\widehat{S^\mu}$.

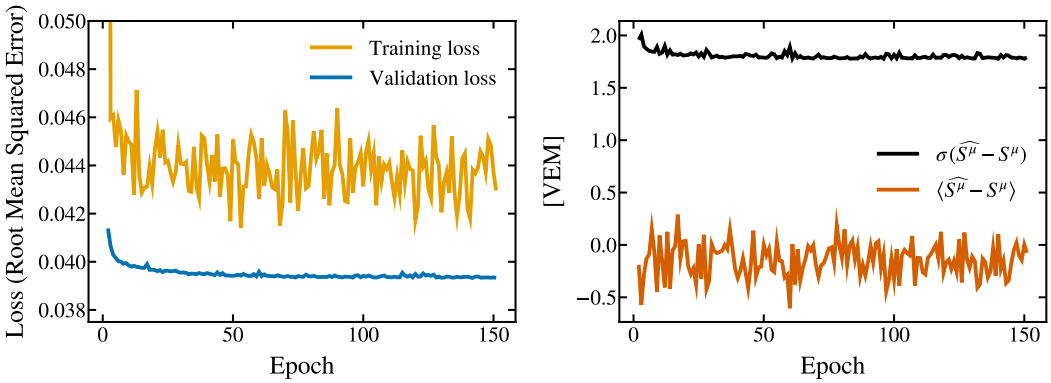

Figure 7.3: Left: Loss as a function of the epoch, see Equation 7.1. Right: Mean value and standard deviation of the difference between the integral of the true muon signal and the predicted muon signal for the validation set.

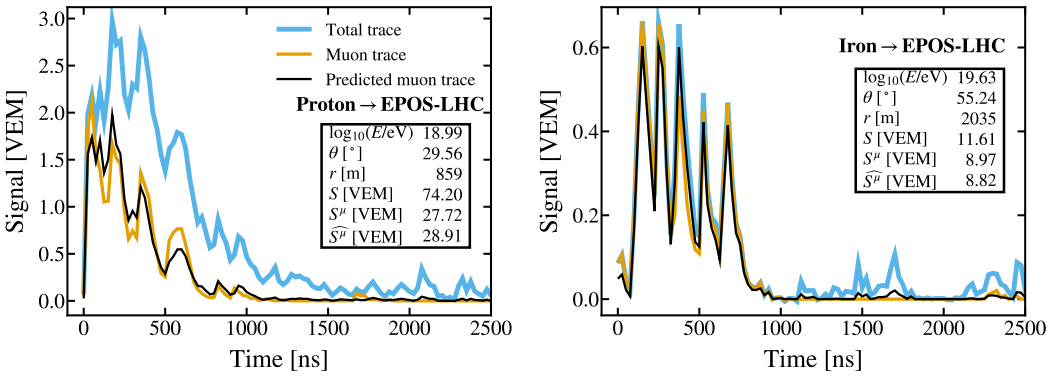

Figure 7.4: Examples of predicted muon traces for two simulated events done with EPOS-LHC from a proton primary for the plot on the left and an iron nucleus primary for the plot on the right. The prediction (black line) agrees well with the shape of the true muon trace (orange line) for a majority of the time bins. The blue thicker line corresponds to the total trace, the one measured by the stations of the SD.

its spiky shape and the fact that most muons arrive earlier. A thorough discussion of the results and their dependence on several variables are given in the next sections.

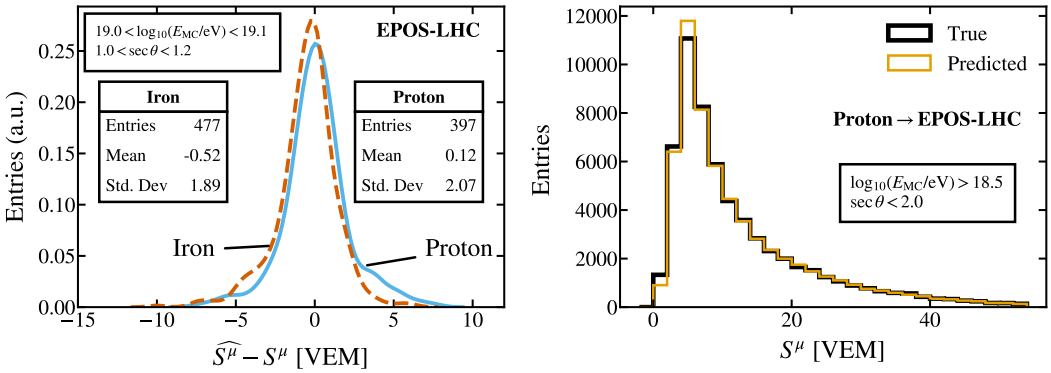

Figure 7.5: Left: Distribution of the difference between the integral of the predicted muon signal $\widehat{S^\mu}$ and the integral of the true muon signal S^μ. Right: Distribution of the predicted and true muon signals for all the stations used.

2.1 Integrals of the trace

One way to assess the performance of the method is to compute the intregal of the predicted muon trace, $\widehat{S^\mu} = \sum_{t=1}^{200} \widehat{S_t^\mu}$ and compare it to the integral of the true muon trace $S^\mu = \sum_{t=1}^{200} S_t^\mu$. The integral of the muon trace is an interesting physical observable, since it relates to the total number of muons that reach the ground. In the left panel of Figure 7.5, we show a distribution with the difference between $\widehat{S^\mu}$ and S^μ for a particular bin of energy and zenith angle. The difference is compatible with zero and does not show a strong dependence with the value of the true muon signal. In the right panel of Figure 7.5, we show the distribution of the $\widehat{S^\mu}$ and S^μ for all the stations in the test set for showers initiated by a proton nucleus. The distributions are very similar with only small differences at lower values of S^μ.

In Figure 7.6 we plot the mean and standard deviation of the distribution of $\widehat{S^\mu} - S^\mu$ as a function of S^μ. We have a mean bias that is close to zero, even for large values of S^μ. The bias exceeds 2 VEM only in the rare cases of large zenith angle and low muon signal. For vertical events, we show that the standard deviation increases as S^μ increases. These results depend heavily on the zenith angle. We can see how the performance for larger zenith angles improves comparing the left and the right panels of Figure 7.6. This happens because for large zenith angles, the total signal is dominated by the muon signal, therefore it is easier to predict it. More performance plots can be found on pages 149 and 150.

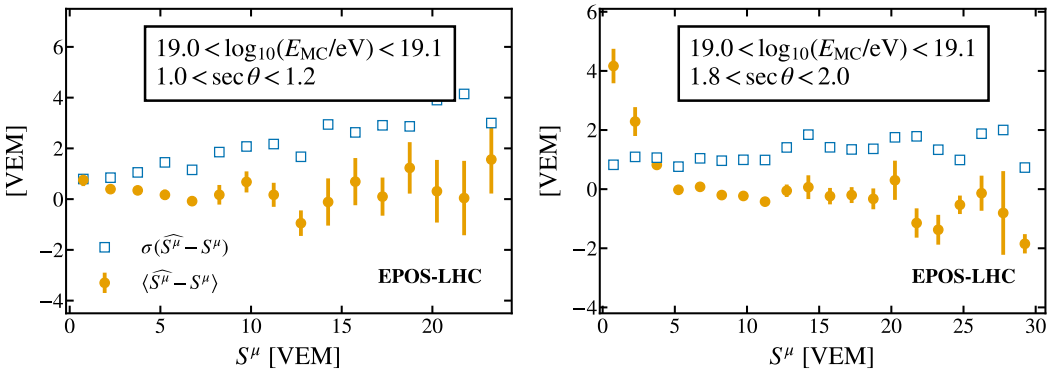

Figure 7.6: Mean and standard deviation of the difference between the integral of the predicted muon signal $\widehat{S^\mu}$ and the integral of the true muon signal S^μ for all the stations from events with the energies and zenith angles specified in the boxes.

2.2 Time distribution

Another physical observable worth to analyze is the risetime of the signals. The risetime gives us information about the shape of the trace: traces where the signal is concentrated in a few bins will have a shorter risetime, while traces where the signal is spread over time will have larger risetimes. In particular, the muon component has a smaller risetime than the electromagnetic component, since muons arrive earlier and in a shorter window of time.

In Figure 7.7 we compare the risetime of the predicted muon trace $\widehat{t^\mu_{1/2}}$ with the risetime of the true muon trace $t^\mu_{1/2}$. We can see that the standard deviation is less than 100 ns for most values of the risetime. This is a very small time compared to the considered trace length (200 bins corresponding to 5 μs). Note that the pulse generated by a single muon has a risetime of around 15 ns and a decay constant of around 60 ns [43]. The mean value is close to zero and the performance improves with the zenith angle. This means that we can successfully predict not only the integral of the muon trace but also the shape of the muon trace.

2.3 Hadronic model

As we have explained before, our neural network has been trained on simulations done with EPOS-LHC. We test our method now using simulations done with QGSJetII-04 and Sibyll 2.3. That is, we predict for simulations that are not only unknown to the NN but also have been generated using a different hadronic model.

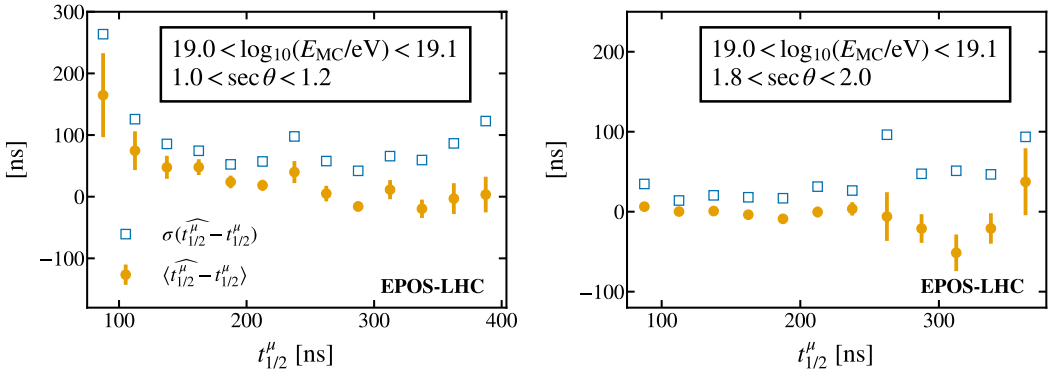

Figure 7.7: Mean and standard deviation of the difference between the risetime of the predicted muon signal $\widehat{t^{\mu}_{1/2}}$ and the risetime of the true muon signal $t^{\mu}_{1/2}$ for all the stations from events with the energies and zenith angles specified in the boxes.

QGSJetII-04

When studying the differences between the predicted and true muon signals, results are similar to those shown in Figures 7.6 and 7.7 for simulations done with QGSJetII-04. In Figure 7.8 an example of a trace obtained with simulations produced using QGSJetII-04 is shown. The prediction follows the shape of the peaks, predicting quite accurately the muon signal. The difference between true and predicted muon signals does not show a strong deviance from zero.

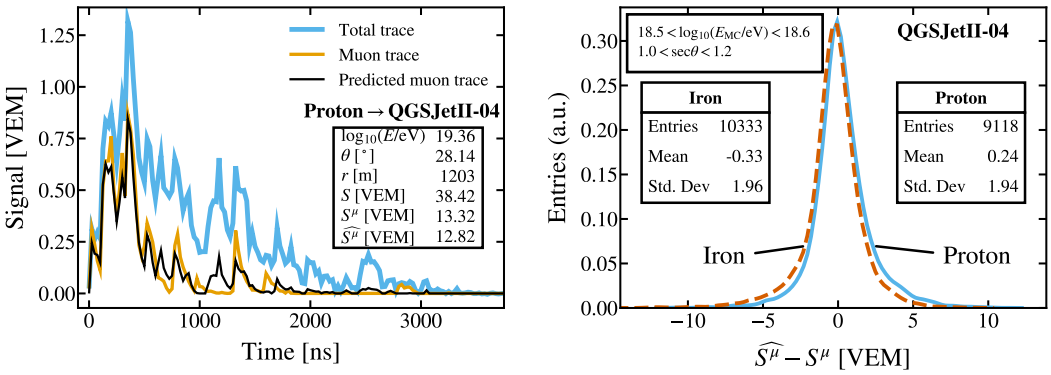

Figure 7.8: Left: Example of a predicted trace for a simulation of a proton generated air-shower done with QGSJetII-04. Right: Distribution of $\widehat{S^{\mu}} - S^{\mu}$ for all the stations in the bin specified for simulations using proton and iron nuclei.

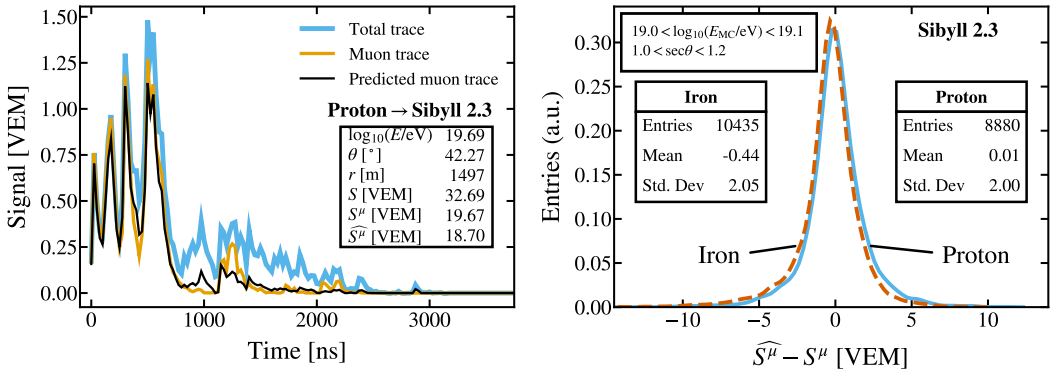

Figure 7.9: Left: Example of a predicted trace for a simulation done with Sibyll 2.3 with a proton as primary cosmic ray. Right: Distribution of $\widehat{S^\mu} - S^\mu$ for all the stations in the bin specified for simulations using proton and iron nuclei.

Sibyll 2.3

The result of predicting the muon traces for simulations done with Sibyll 2.3 is shown in Figure 7.9. By comparing the values of the bias and resolution, we can see that the performance is similar to the case where the predictions were based on simulations done with QGSJetII-04.

With these results we have proven that the NN predictions are independent of the hadronic model used to simulate extensive air-showers. For completeness, we have also carried out the opposite exercise: train the neural network with QGSJetII-04 and Sibyll 2.3 and predict for the other hadronic models. The outcome is in good agreement with what has been discussed for the case where the NN learns from events simulated with EPOS-LHC.

3 Comparison with data

In this section we have a preliminary look at how the neural network performs when applied to experimental data. In Figure 7.10 we show examples of the muon trace predicted for two typical traces recorded by the SD. We can see similar features to those shown by the simulated traces of Figure 7.4: the predicted muon fraction is larger at earlier times since muons arrive earlier; in addition, the predicted muon trace is spikier.

After looking at some of the most relevant physics observables, the conclusion is that we reproduce the behaviour observed in other analyses, based on the use of conventional tools. In particular, we find a muon deficit like the one discussed in refs. [52,53,88]. In the

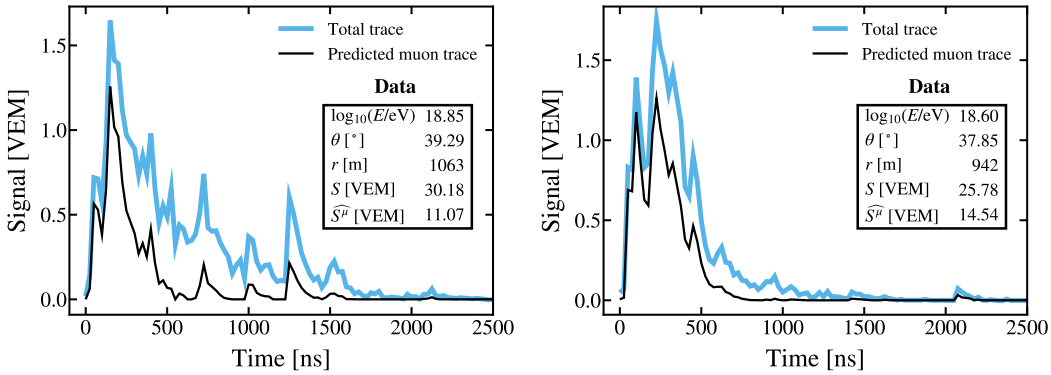

Figure 7.10: Examples of the predicted muon traces for two stations that belong to two different events recorded by the SD.

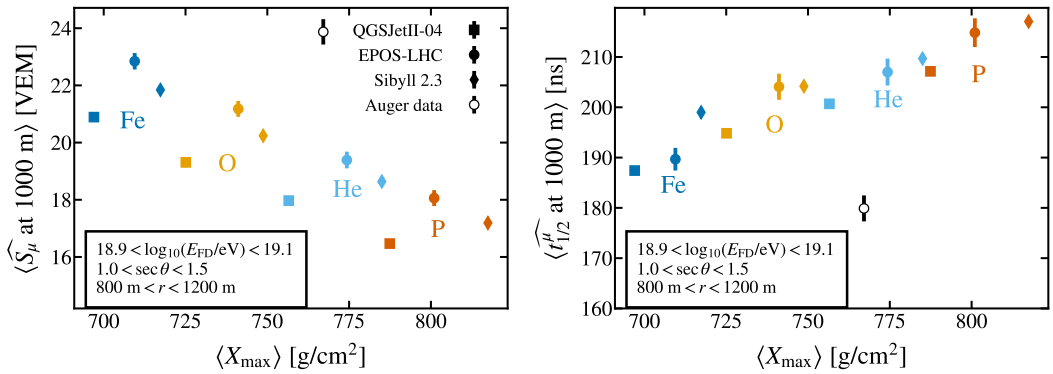

Figure 7.11: Left: Average value of the integral of the predicted muon trace as a function of the averange value of X_{max}. Right: Average value of the risetime for the predicted muon trace as a function of the average value of X_{max}.

left panel of Figure 7.11, we compare distributions of muon signals at 1000 m from the shower core for simulations and hybrid data, that is, data that are measured simultaneously by the SD and the FD. We clearly see a difference in the amount of the recorded muon signal. With this plot we reproduce the results shown in Figure 5 of ref. [53] for the first time for vertical events, since in ref. [53] only inclined events were used.

Turning our attention to time signals, we can compare data and simulations using observables related with the time distribution of the muons rather than with the integral of the trace. In the right panel of Figure 7.11, we show the distributions for the predicted risetime, both for hybrid data and simulations. The differences observed between hybrid data and simulations are striking. This is the first time that the risetime from the muon signal is obtained for data and compared to the one in simulations. With the plots presented,

we have shown that not only there is a problem with the size of the signals but also there is a problem with the temporal distribution of the muons.

4 Comparison to other experiments

Thanks to the predictions of the neural networks, we can extract the lateral distributions of the muon and electromagnetic signals. We compare our findings for the data collected with the Surface Detector to parameterizations obtained decades ago by the Akeno and Volcano Ranch experiments. This is an elegant way of proving that the predictions of the neural network for data are sensible. Currently, the design of the SD does not allow to measure the muon signal in an independent way but other experiments have focused on measuring the muon signal with other types of detectors.

At both Akeno and Volcano Ranch, plastic scintillators of 1 m^2 and 3.3 m^2 respectively were used to record air-showers. The scintillators respond to both electrons and muons and photons to a lesser extent. A measurement of the electromagnetic component with the Auger detectors is expected to have a similar lateral distribution (with regard to shape) as that observed at Volcano Ranch and Akeno. The photon/electron ratio is known from direct measurements to change only rather slowly with distance [133].

The atmospheric depths of the Volcano Ranch and Akeno arrays are 834 and 920 g cm^{-2} respectively, conveniently straddling that at the Pierre Auger Observatory (875 g cm^{-2}). It is not anticipated that changes of LDFs, particularly in the case of the muons, will depend so strongly on depth as to invalidate our qualitative conclusions.

4.1 Akeno measurement I: J. Phys. G. Nucl. Part. Phys 21 1101 (1995)

This paper [134] studies the properties of muons with energies $\geq$ 1 GeV. They fit the lateral distribution of muons (LDM) with the Greisen formula [135]:

$$\rho_\mu(r) = N_\mu(C_\mu/R_0^2)R^{-\alpha}(1+R)^{-\beta} \tag{7.2}$$

where:

- ρ_μ is the muon density
- N_μ is the total number of muons
- $R = r/R_0$

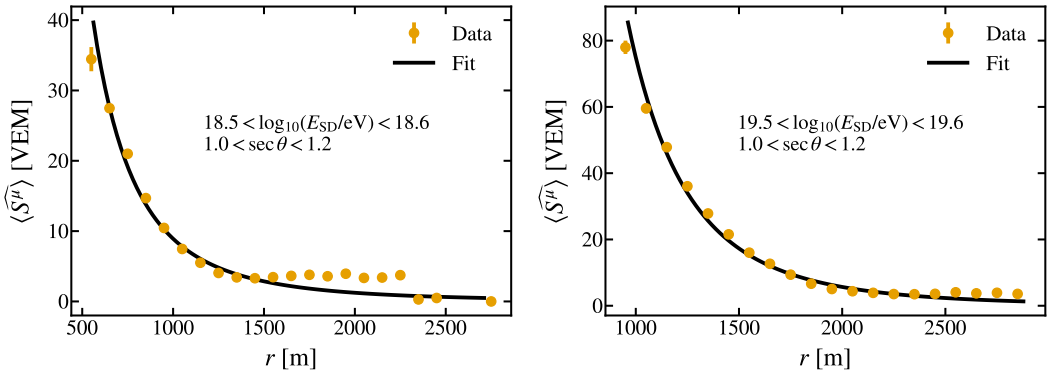

Figure 7.12: Average lateral distribution of muons fit to the Akeno parameterization. Stations with signal above 5 VEM are considered.

- $\log R_0 = (0.58 \pm 0.04)(\sec\theta - 1) + (2.39 \pm 0.05)\,\mathrm{m}$
- $\langle \sec\theta \rangle = 1.09$, since the fit is restricted to what they define as vertical events ($\sec\theta < 1.2$).
- $C_{\mu} = \dfrac{\Gamma(\beta)}{2\pi\Gamma(2-\alpha)\Gamma(\alpha+\beta-2)}$
- $\alpha = 0.75$
- $\beta = 2.52 \pm 0.04$

We use Equation 7.2 to fit our data. We consider only vertical events with $\sec\theta < 1.2$. In our fit we have a single free parameter, N_{μ}, since we are using different units that affect the scaling. The rest of the parameters are fixed to the values given above. The results for the energy bins $18.5 < \log_{10}(\mathrm{E_{SD}/eV}) < 18.6$ and $19.5 < \log_{10}(\mathrm{E_{SD}/eV}) < 19.6$ can be seen in Figure 7.12 . The Akeno parameterization and our data show a reasonable agreement. Our LDM flattens out at large distances. This unnatural behaviour is due to the cut on the total signal at 5 VEM that we have used to train the neural network that removes very low muon signals. It was studied and found that this flattening occurs when signals start being close to 5 VEM and the percentage of stations that pass the cuts is decreasing rapidly with distance.

For the energy bin $19.5 < \log_{10}(\mathrm{E_{SD}/eV}) < 19.6$, our data starts at distances around 1000 m. For distances as large as those, the Akeno paper suggests a slight modification of Equation 7.2 above. We now fit our data with the following formula:

$$\rho_{\mu}(r) = N_{\mu}(C_{\mu}/R_0^2)R^{-\alpha}(1+R)^{-\beta}[1+(r/800\mathrm{m})^3]^{-\delta} \tag{7.3}$$

The parameter δ was fixed to a value of 0.6 in the Akeno paper to reproduce the fall at large distances to the core. We leave free N_μ and δ as well in order to have a better description of the tail of the LDM. The result of the fit is shown in the right panel of Figure 7.12. For a value of $\delta = 0.35 \pm 0.04$, we obtain a good level of agreement.

4.2 Akeno measurement II: J. Phys. G. Nucl. Part. Phys 18 423 (1992)

In this paper [84] the methods used to measure the Akeno energy spectrum above 10^{17} eV are described. The parameterization of the density of electromagnetic signal as a function of the core distance is expressed as follows:

$$\rho_e = N_e C_e R^{-\alpha}(1+R)^{-\eta+\alpha}\left(1+\frac{r}{2000}\right)^{-0.5} \tag{7.4}$$

where

- N_e is the number of electrons
- C_e is a normalization factor
- $R = r/R_M$
- $R_M = 91.6\,\text{m}$ is the Moliére length
- $\alpha = 1.2$
- $\eta = (3.80 \pm 0.05) + (0.10 \pm 0.05)\log_{10}(N_e/10^9)$

We fit the predicted average electromagnetic signal with Equation 7.4. The predicted electromagnetic signal for a station is obtained as the difference between the total signal measured and the muon signal predicted by the neural network $\widehat{S^{EM}} = S - \widehat{S^{\mu}}$. For data from the Pierre Auger Observatory, we take R_M equal to 80 m. The fit result does not vary sensibly if we modify this value by $\pm 10\,\text{m}$. Since they do not offer a numerical value for C_e in the paper, we take the product $N_e C_e$ as a single free parameter in our fit. Notice that the expression for η depends on the number of electrons, and therefore on the energy. For the typical values of N_e considered in the Akeno paper, η values vary around 4. We let η as a free parameter as well. For the two energy bins we have considered, the fit values of η are 3.5 and 4.2. η values increase with energy and they are similar to the figures reported by the Akeno collaboration. There is a remarkable agreement between the Akeno parameterization and the lateral distribution of the electromagnetic signal (LDE) corresponding to data, see Figure 7.13.

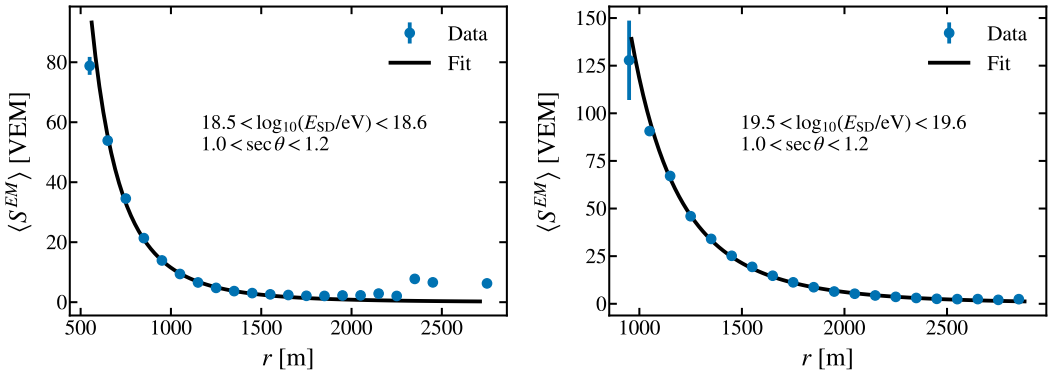

Figure 7.13: Average lateral distribution of electromagnetic signal fit to the Akeno parameterization.

4.3 Volcano Ranch measurement: Phys. Rev. Lett. 10 146 (1963)

This paper [55] reports the first evidence for a cosmic ray particle with energy 10^{20} eV, recorded by the Volcano Ranch experiment. The zenith angle of the event is $10 \pm 5°$. In Figure 2 of that paper, John Linsley shows the particle density of the event as a function of distance from the shower axis. The data of the triggered detectors are fit with the following expression:

$$VR(\alpha, \eta) = \frac{N}{R_0^2} C(\alpha, \eta) \left(\frac{R}{R_0} \right)^{-\alpha} \left(1 + \frac{R}{R_0} \right)^{-\eta+\alpha} \tag{7.5}$$

where the best fit parameters are:

- $N = 5 \cdot 10^{10}$
- R_0 is the Moliére length
- $C_\mu = \dfrac{\Gamma(\eta - \alpha)}{2\pi\Gamma(2-\alpha)\Gamma(\eta - 2)}$
- $\alpha = 2 - s$
- $\eta = 6.5 - 2s$
- $s = 1$

To compare with this parameterization, we now take the predicted electromagnetic signal for vertical events ($\sec\theta < 1.2$) of our highest energy bin $\log_{10}(E_{SD}/eV) > 19.8$. We fit these data with Equation 7.5. We consider N and s as free parameters. The best

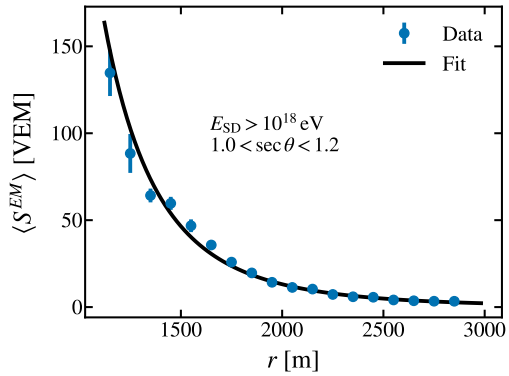

Figure 7.14: Average lateral distribution of electromagnetic signal fit to the Volcano Ranch parameterization.

fit gives a value of $s = 0.98 \pm 0.05$. In Figure 7.14 the parameterization done by Linsley and our data are shown. The observed fluctuations can be explained by the low amount of statistics in the highest energy bin. The level of agreement between the parameterization and data is remarkable.

5 Summary and conclusions

Using Recurrent Neural Networks we are able to predict accurately the muon signal contained in the simulated time traces of the Surface Detector of the Pierre Auger Observatory. The predictions for the signal size and temporal distribution of the muon signal are precise and the predictions do not depend on the model used to simulate hadronic interactions.

We have shown that the lateral distributions of muons and electromagnetic particles, extracted from the SD data, are well reproduced by published parameterizations, which are based on the data collected by the Akeno and Volcano Ranch experiments. We conclude that neural networks are a powerful tool to extract reliable estimations of the muonic component of extensive air-showers.

However, simulations are not able to reproduce all the features of the recorded SD data (see the next chapter). It is then hard to quantify the precision with which the muon component is extracted from the data. Even if the predicted muon component was the real one, simulations are needed to make a comparison and extract mass composition information, thus making it hard to make inferences about mass composition. The data collected with AugerPrime will be paramount to improve the models that simulate hadronic interactions at extreme energies. Thanks to this improvement it will be hopefully possible to obtain a very reliable estimate of the muon signal for the large amount of data collected by the SD

since 2004. This will represent a major step forward in the capability of the Pierre Auger Observatory to make mass estimates on an event by event basis.

8

A study on the differences between data and simulations

The results obtained with methods from machine learning depend on the data used for training the models. We have used simulations to train the models. If these simulations can not reproduce correctly the data, the model and its predictions may not be accurate.

We have already given hints in refs. [52,53,88] of the problems that simulations have when modelling the data. It is as important as training the model to study these discrepancies. In this chapter we compare simulations and data with the same FD energy and prove that the signals at the ground are different. We study what rescaling is needed on the electromagnetic and muon components of the simulated signals at the ground such that data and simulations agree.

This chapter is structured as follows. Section 1 is a short introduction motivating this study and introduces the problem with the signal at the ground. In Section 2 the two methods used to compare data and simulations are explained, and the results obtained with these methods are shown in Section 3 and Section 4. The chapter ends with the conclusions of this study in Section 5.

1 Introduction

This chapter is a short and original study on some of the discrepancies between data and simulations. In ref. [52] it has been shown that there are discrepancies between the signals at the ground, but for this analysis simulations have the same longitudinal profile as data, which is somehow similar to fixing the same electromagnetic signal. In ref. [53] there are discrepancies between data and simulations but these are measurements of only the muon signal in inclined events. There are other works that take into account both the electromagnetic and muon signal independently using the FD [136] but this is the first one using the SD with its larger data sample.

We study the distributions of signals at the ground and compare them between data and simulations. Since the electromagnetic and muon signals are known in simulations, we find the best rescaling of these components that makes the distribution of simulated signals match those in data. We will study the distributions of two samples: one is obtained selecting stations with distances to the core around 1000 m. The other sample will be obtained using S_{1000}, explained in Lateral distribution function on page 29, that is used as the energy estimator.

For the comparison between data and simulations, we are going to pick events with the same FD energy. To avoid any discussion associated with the cuts done for the measurements of the FD, we use the Monte Carlo energy (E_{MC}) as a proxy of the FD energy (E_{FD}) for simulated events. In the left panel of Figure 8.1, it is shown how the Monte Carlo energy and the FD energy are practically the same, hence this justifies the previous choice. This is not true, however, when comparing the energy measured from the FD and the energy measured from the SD in simulations, see the right panel of Figure 8.1. For data the energies from the SD and FD are similar because the FD energy is used to calibrate the SD energy. This is related to the problems we have seen in simulations: given the same energy from the FD, the signals at the ground are different.

2 Methods

The two methods that we use are very similar. In both cases a signal is compared between data and simulations. This signal $\tilde{S}$ can be written as a function of the muon and electromagnetic contributions

$$\tilde{S} = \alpha \tilde{S}^{\mu} + \beta \tilde{S}^{EM} \tag{8.1}$$

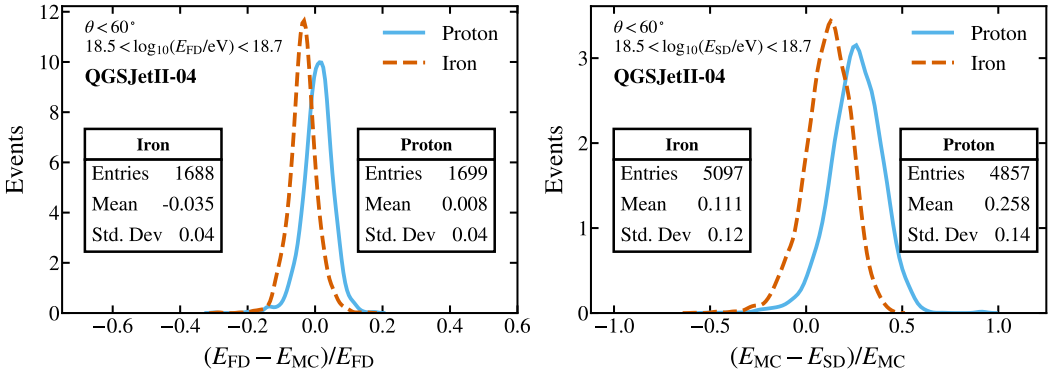

Figure 8.1: Comparison between the distribution of energies for proton and iron simulations. Left: Energy from the FD, E_{FD}, compared to the Monte Carlo energy, E_{MC}. Right: Energy from the SD, E_{SD}, compared to the Monte Carlo energy, E_{MC}.

where α and β are the parameters that rescale the muon and electromagnetic component, respectively. The values of α and β are found by requiring that the signal $\tilde{S}$ in data and simulations are as close as possible. We quantify this distance using the Kolmogorov-Smirnov test. For two normalized cumulative distributions F_1 and F_2, this test is defined as follows:

$$KS = \max_x \left| F_1(x) - F_2(x) \right| \tag{8.2}$$

Intuitively, if the distributions are similar then the cumulative distributions are also similar and KS will be small. Note that since the cumulative distributions only take values between zero and one, the minimum and maximum values of KS are zero for the same distribution and one, respectively.

To make the comparison between data and simulations as fair as possible the distribution of energies and zenith angles of simulations is matched to the one in data by random sampling. The sampling is repeated 100 times for each energy bin of 0.1 in $\log_{10} E$ and 0.2 in $\sec\theta$. The matching is, however, not very important as the comparison is done for narrow bins of energy and zenith angle but it is useful to estimate the statistical uncertainty of the values of α and β obtained. The minimization is done in a brute force way. α and β are given values independently between 0 and 2.5 in steps of 0.01 and the values for which Equation 8.2 is minimized are found and saved. Then, from the list of 100 trials, the mean values of α and β are taken as the final value of α and β for each bin of energy and zenith angle.

For the first method we pick stations with a distance to the core r close to 1000 m, in particular stations with 900 m $< r <$ 1100 m, and compute KS for these samples. For the second method we use $\tilde{S} = S_{1000}$. In this case Equation 8.1 is rewritten as

$$S_{1000} = \alpha S^{\mu}_{1000} + \beta S^{EM}_{1000} \tag{8.3}$$

However, S^{μ}_{1000} and S^{EM}_{1000} is not something known a priori. As a reminder, S_{1000} comes from a fit of the individual signals measured at each station. It is natural then to obtain S^{μ}_{1000} and S^{EM}_{1000} by doing a fit of the muon and electromagnetic signals as a function of the distance and then picking the value at 1000 m. The function chosen for this fit is the same as the one used for the fits of S_{1000}: a NKG, see Equation 2.2 on page 29.

2.1 Data selection

Data selection is simpler than in the other analyses discussed in this thesis. As usual we only include 6T5 events with energies above $10^{18.5}$ eV and $\sec\theta < 2$. The study is done only for energies below $10^{19.4}$ eV since above this energy there are few events measured (less than 1000 events in each bin, with the number of events decreasing with energy) and a reliable comparison of the distributions between data and simulations can not be made.

For the first method stations with $r \simeq 1000$ m have large signals (more than the 5 VEM used for the other analyses) so it is not necessary to worry about the problems associated with stations with low signals. Only about 0.25% of the stations in the range of distances studied had signals below 5 VEM and almost all of them come from events with a reconstructed energy between $10^{18.5}$ eV and $10^{18.6}$ eV. The final cut on the stations is $S > 5$ VEM and not to use stations that have any of the high-gain or low-gain channels saturated. A total of 46497 stations from 40628 events were included in the sample.

For the second method all the events have S_{1000}. However, for simulations a fit of the lateral distribution of the muon and electromagnetic signals is performed and sometimes those fits can not be done or are of poor quality. This occurs mostly when the number of stations available in an event is low. Events with only two or one stations are not included. This only happens at the lowest energies, from $10^{18.5}$ eV to $10^{18.65}$ eV. The other set of events that are not included are those for which it is not possible to reconstruct correctly S^{μ}_{1000} and S^{EM}_{1000}, that is, when:

$$\frac{\left|(S^{\mu}_{1000} + S^{EM}_{1000}) - S_{1000}\right|}{S_{1000}} > 1 \tag{8.4}$$

This ensures that the fits are good, as we will see later. The number of events removed with the previous two cuts is quite small: more than 99% of the previously selected events (6T5, $E > 10^{18.5}$ eV and $\sec\theta < 2$) pass the two new cuts introducted when using S_{1000}.

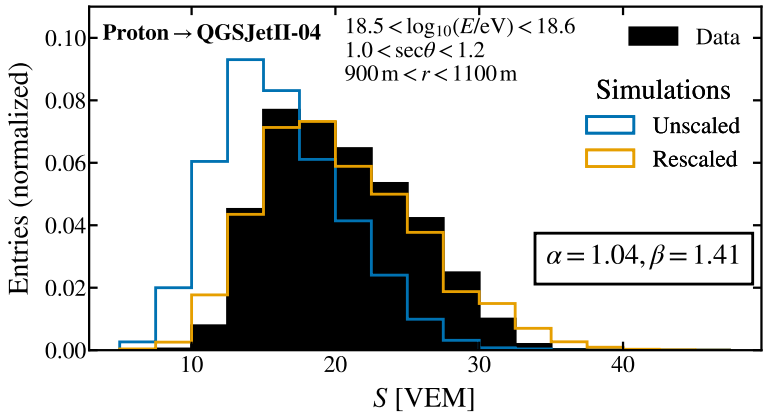

Figure 8.2: Example of the distribution of signals for data and simulations before and after the scaling using Equation 8.1.

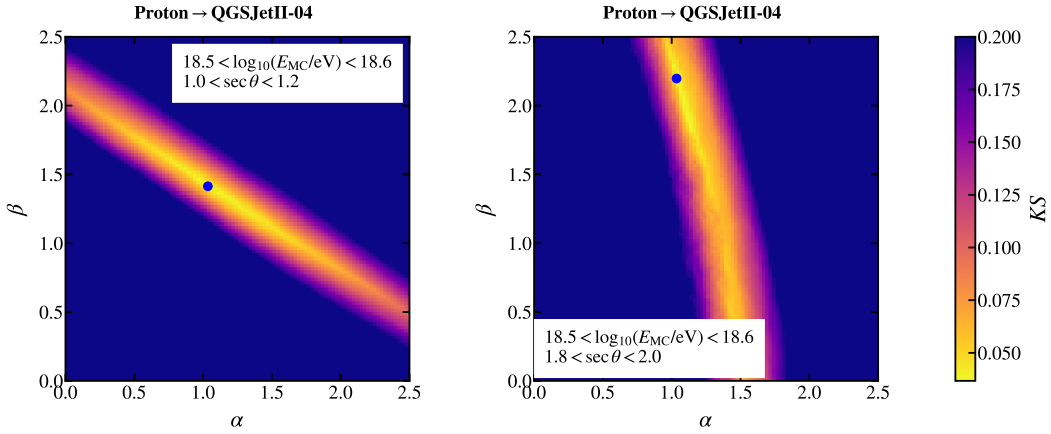

Figure 8.3: Evolution of the values of the Kolmogorov-Smirnov test with $\sec\theta$. The minimum value of KS is obtained when α and β have the values of the blue circle.

3 Method I: Using the signal of stations at 1000 m

In Figure 8.2 the distribution of signals at the ground for stations around 1000 m for data and simulations is shown. The distribution of signals for simulations (in this case protons) has been corrected using $\alpha = 1.04$ and $\beta = 1.41$. It can be seen the corrected distribution is much more similar even though it is still wider than the one in data. This can be expected since the composition at those energies is unlikely to be purely proton.

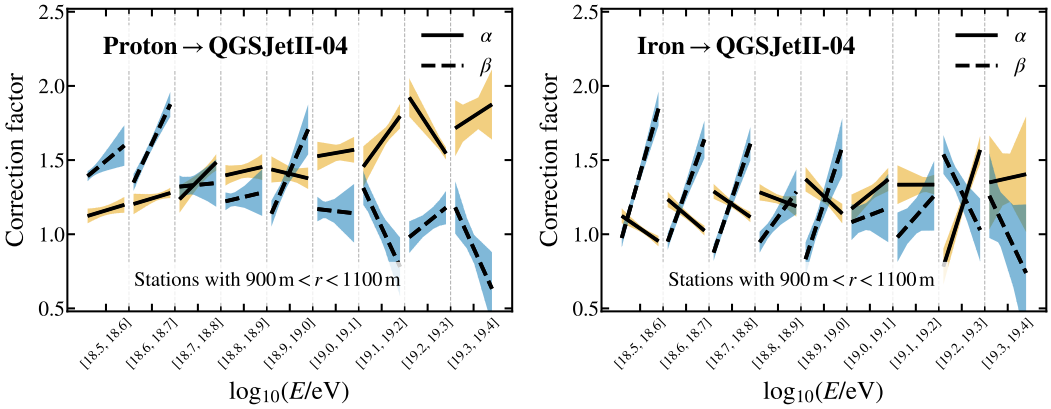

Figure 8.4: Evolution of α and β with the energy and zenith angle for simulations done with QGSJetII-04. For each energy bin, the zenith angle increases linearly in $\sec\theta$ from left to right from 1 to 2. Linear fits (black lines) are shown instead of the points for each energy bin. The shadowed area corresponds to the uncertainty of the fit.

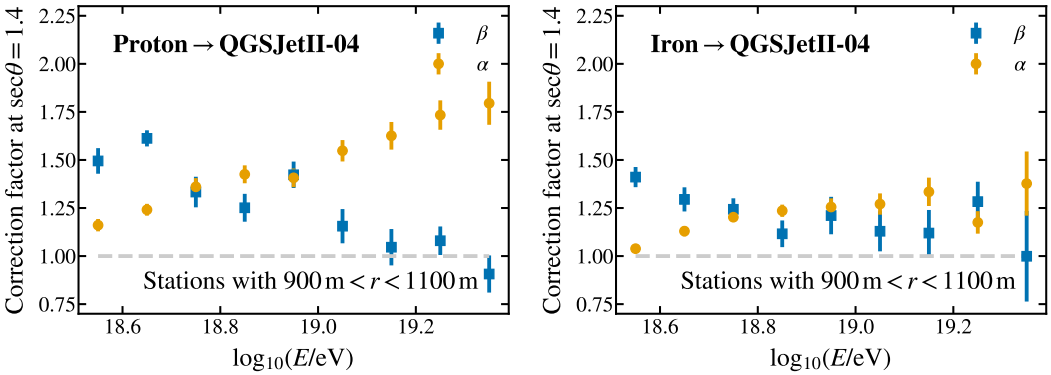

Figure 8.5: Evolution of α and β with the energy for simulations done with QGSJetII-04. The values of the fits in Figure 8.4 have been plotted at $\sec\theta = 1.4$. The error bars correspond to the uncertainty of the fits.

In Figure 8.3 the dependence of α and β with $\sec\theta$ is plotted. We can see how the ellipse of low values of the KS test becomes more and more vertical. That means that the value of β becomes less important. The explanation is simple: as $\sec\theta$ increases, the muon fraction of the signal increases while the fraction of electromagnetic signal decreases. At $\sec\theta \simeq 2$ (60°) almost all of the signal is muon signal and then the value of β is undetermined when minimizing Equation 8.2, since $S^{EM} \approx 0$.

The values of α and β that minimize KS as a function of the energy and zenith angle have been plotted in Figure 8.4 for QGSJetII-04. The same plot for EPOS-LHC can be

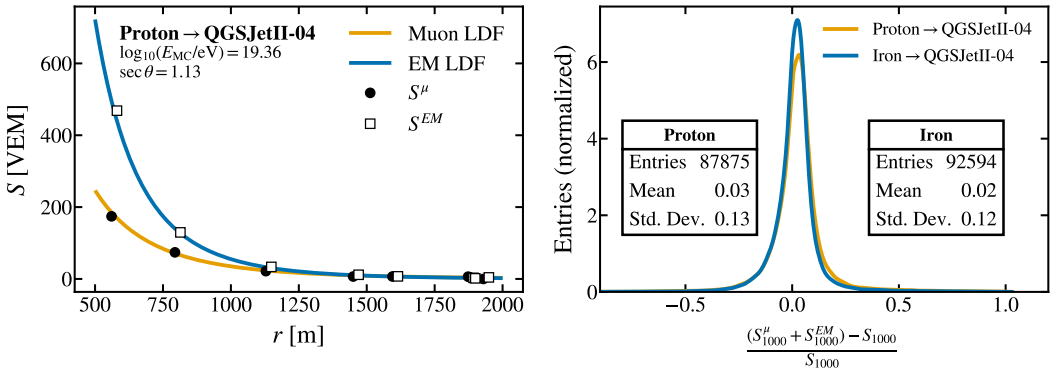

Figure 8.6: Left: Example of fitted LDFs for the muon and electromagnetic components for an event with 8 stations. r has been shifted slightly to the right for the muon component and to the left for the electromagnetic signal to avoid overlapping between the points. Right: Histogram of the comparison between the true value of S_{1000} and the reconstructed value $S^{\mu}_{1000} + S^{EM}_{1000}$, relative to the true value.

found in Figure 12 on page 152. A linear fit of the values of α and β obtained for each bin has been performed to ease the visualization. While there are fluctuations a trend can be extracted from the plots: a scaling greater than 1 is almost always needed. The scaling increases with the energy for the muon component but it decreases for the electromagnetic component.

In Figure 8.5 the evolution of α and β has been plotted as a function of the energy. An arbitrary choice has been made: the values of the fits at $\sec\theta = 1.4$ have been plotted. This value has been chosen because it is close to the median of $\sec\theta$ for the data used. It can be seen that a bigger rescaling is needed for proton than for iron simulations. This can be expected since, independently of the composition assumed for data, the signals at the ground are lower for proton simulations than for iron simulations. However, since the composition for data is expected to be between proton and iron, the rescaling for iron should be lower than 1 if simulations were modelling the data correctly, in contradiction with our results.

4 Method II: Using S_{1000}

Now S_{1000} is used to compare data and simulations instead of the signal of individual stations close to 1000 m. In the left panel of Figure 8.6, an example of the fitted LDF is shown for the electromagnetic and muon components of one simulated event. In the right panel of Figure 8.6, S_{1000} is compared to the value of $S^{EM}_{1000} + S^{\mu}_{1000}$ with both S^{EM}_{1000} and

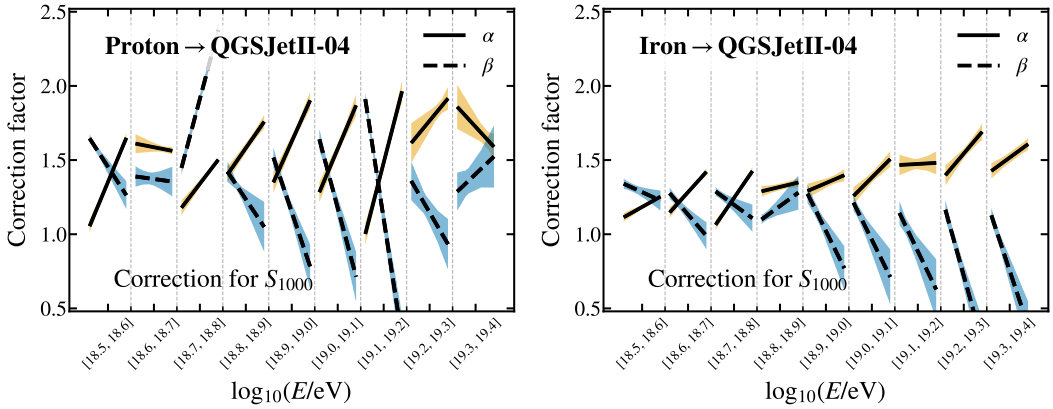

Figure 8.7: Evolution of α and β when using QGSJetII-04 with the energy and zenith angle. For each energy bin, the zenith angle increases linearly in $\sec\theta$ from left to right from 1 to 2. Linear fits (black lines) are shown instead of the points for each energy bin. The shadowed area corresponds to the uncertainty of the fit.

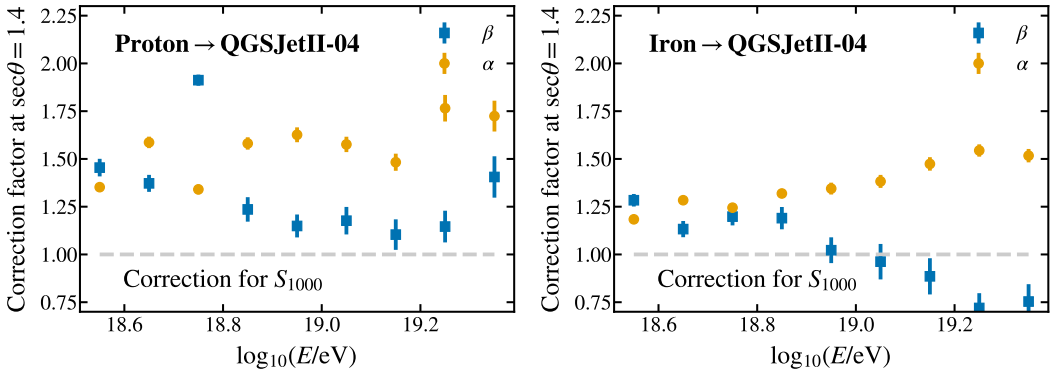

Figure 8.8: Evolution of α and β when using QGSJetII-04 with the energy when picking the values of the fits in Figure 8.4 at $\sec\theta = 1.4$. The error bars correspond to the uncertainty of the fits.

S^{μ}_{1000} obtained from our fits. We can see that the mean value or bias is less than a 5% and the resolution is less than 15%. Most of the points are very close to zero.

With the values of S^{EM}_{1000} and S^{μ}_{1000}, we proceeded in the same way as before and compute α and β for each energy and zenith angle bin. The results are shown in Figure 8.7 for QGSJetII-04 and Figure 12 on page 152 for EPOS-LHC. As it has been done in the previous section, we have picked the value of the fits at $\sec\theta = 1.4$ and plotted them in Figure 8.8 for QGSJetII-04 and Figure 15 on page 153. Again, we see similar trends with the rescaling needed for the muon component increasing with the energy and the rescaling needed for the electromagnetic component decreasing with energy. However, the total

rescaling is still above 1. At high energies the muon signal is rescaled by $1.5 = 3/2$ and the electromagnetic signal by $0.75 = 3/4$. Assuming equal contributions of the muon and electromagnetic signals, the rescaling is above 1 since $\frac{\frac{3}{2}+\frac{3}{4}}{2} > 1$.

5 Summary and conclusions

After using neural networks to predict the muon trace for the measured data in the previous chapter, we focus on differences between data and simulations. We have studied differences in the signals and shown that both the electromagnetic and muon component would need to be rescaled to match the distributions measured in data.

The rescaling needed is larger for proton than for iron, as could be expected since signals for showers initiated by a proton nucleus have lower signals at the ground. The rescaling is consistently above 1 even for iron simulations when the expected composition for data is lighter than pure iron. The scaling is similar for the two hadronic models used: QGSJetII-04 and EPOS-LHC although it is slightly larger for QGSJetII-04, in agreement with what has been found in previous studies [52,53,88].

The results shown here are restricted to using a few stations around 1000 m or S_{1000} and capture only the differences due to the scale of the signals. A complementary study would consist on studying also the differences due to the time distribution of the signals. Systematic uncertainties have not been computed for this study, since its main purpose is to illustrate that simulations are different from data. A thorough evaluation of all the sources of systematic uncertainties falls beyond the scope of this simple and preliminary analysis.

Conclusions and results

We conclude summarizing the most important results of this thesis:

- Two studies that use the risetime of the signals measured by the Surface Detector with the goal of inferring the mass composition of UHECRs in Chapter 3 and Chapter 4. A new observable is introduced and its performance assessed: the average risetime divided by the distance, $\overline{\text{ToD}}$. We find a trend towards heavier masses as energy increases in Figure 3.11 on page 51 and Figure 4.9 on page 72, in agreement with other methods that use the information from the risetime. While the exact interpretation of the mass depends on the hadronic model employed for the simulations, the evolution with the energy does not depend on the hadronic model.

- A method for predicting the integral of the muon signal measured by the Surface Detector in Chapter 6. This method is based on neural networks and is our first attempt at predicting the muon signal. This is the first work achieving precisions at the level of 1-2 VEM. With very few restrictions, it can be applied to the full data sample collected by the Surface Detector of the Pierre Auger Observatory. It led to a publication in Astroparticle Physics [129].

- A method of predicting the muon component of the signal measured by the Surface Detector in Chapter 7. This method is based on neural networks and is the main result of this thesis. It is a very powerful result that allows to extract the temporal component of the muon signal, having both information about the temporal distribution of the muon arrival time and the signal they deposit.

  Obtaining the muon component can enhance the capabilities of the Observatory to do studies about the mass composition of UHECRs, study hadronic interactions or find showers produced by photons that have very few muons, to name a few.

  The method is applied to data, where differences are found with respect to simulations when comparing the amount of muon signal. This finding agrees with previous results. For the first time, it is also shown that the distribution in the particle arrival time is different, using the risetime of the predicted muon traces $\widehat{t^{\mu}_{1/2}}$.

  This work has been approved for publication as a full-author list article by the Pierre Auger Collaboration.

- A comparison between data and simulations in Chapter 8. With this comparison we show that there are some differences between data and simulations that need to be studied and understood as they may affect the results obtained when applying the

method developed in Chapter 7 to data. With most of the previous results pointing towards a rescaling of the muon component, we find that a rescaling of the electromagnetic component is also necessary.

Appendix for Chapter 3

Linear fits

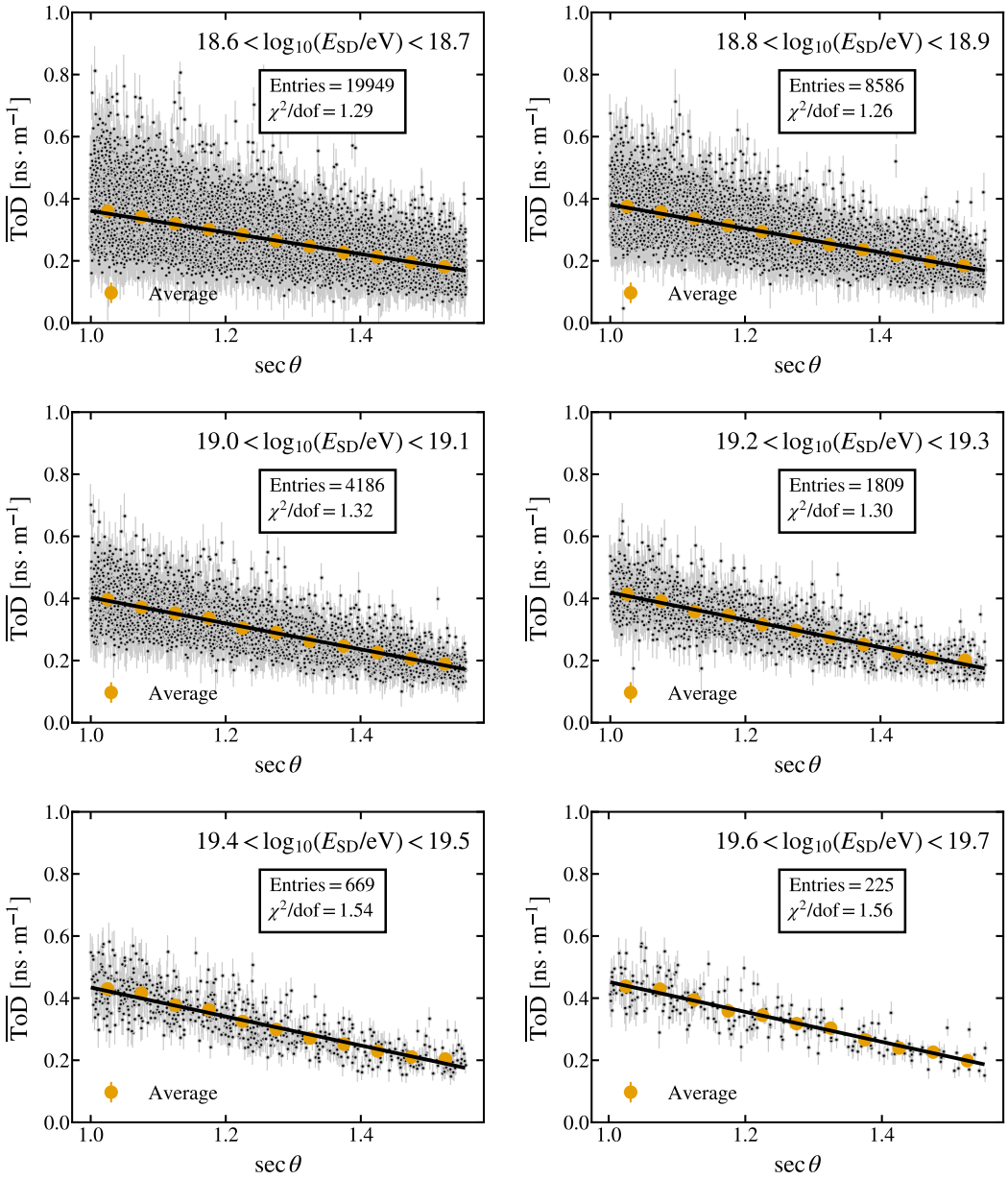

Figure 1: Fits obtained for the $\overline{\text{ToD}}$ as a function of sec θ for some bins of energy.

Appendix for Chapter 4

Method of splitting - Additional plots

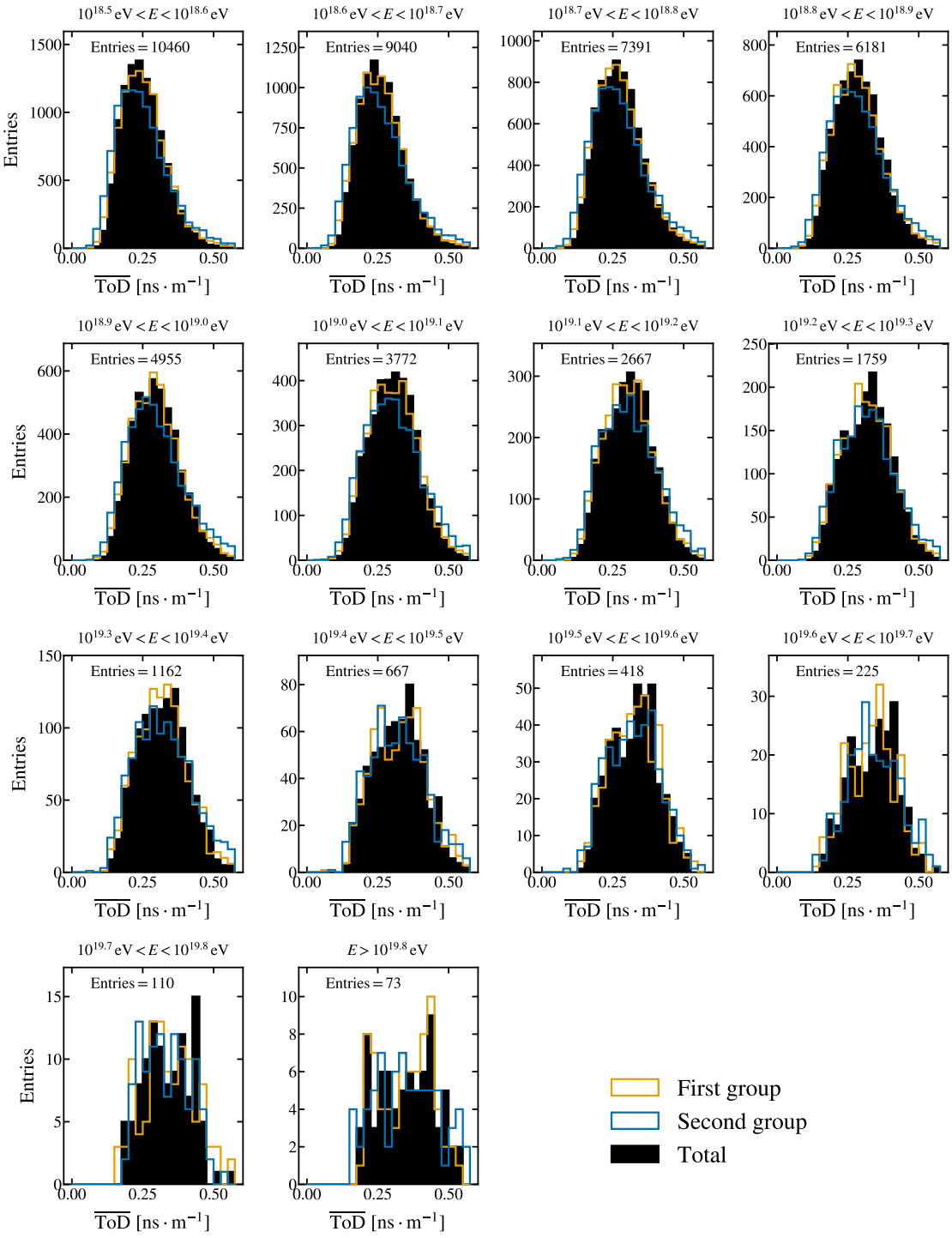

Figure 2: Distributions of the $\overline{\text{ToD}}$ (black, whole event), $\overline{\text{ToD}}_1$ and $\overline{\text{ToD}}_2$, obtained dividing each event for data. Entries is the number of events for each energy bin.

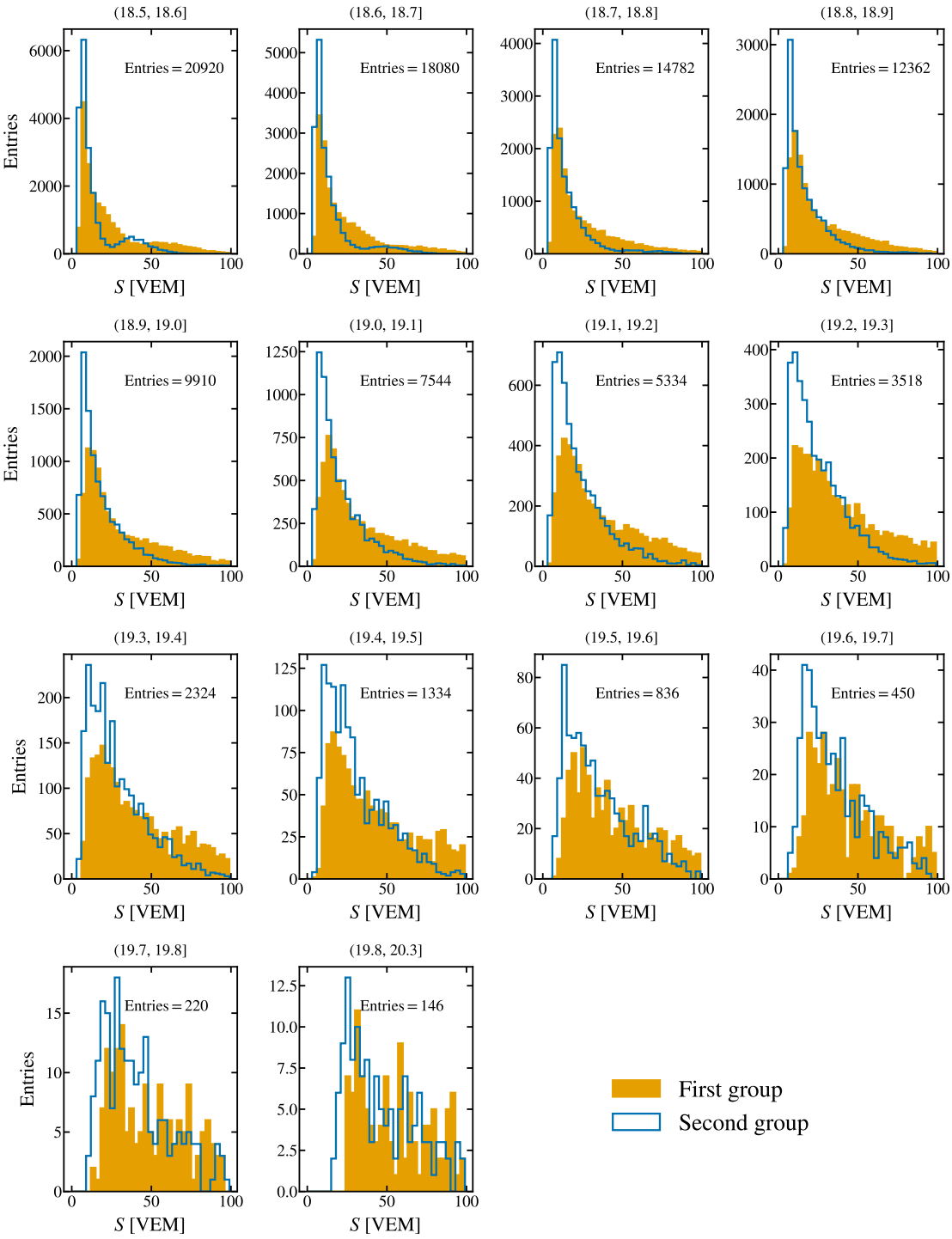

Figure 3: Distributions of the total signal in each of the two groups obtained for the method of splitting. Entries is the number of stations for each of the histograms.

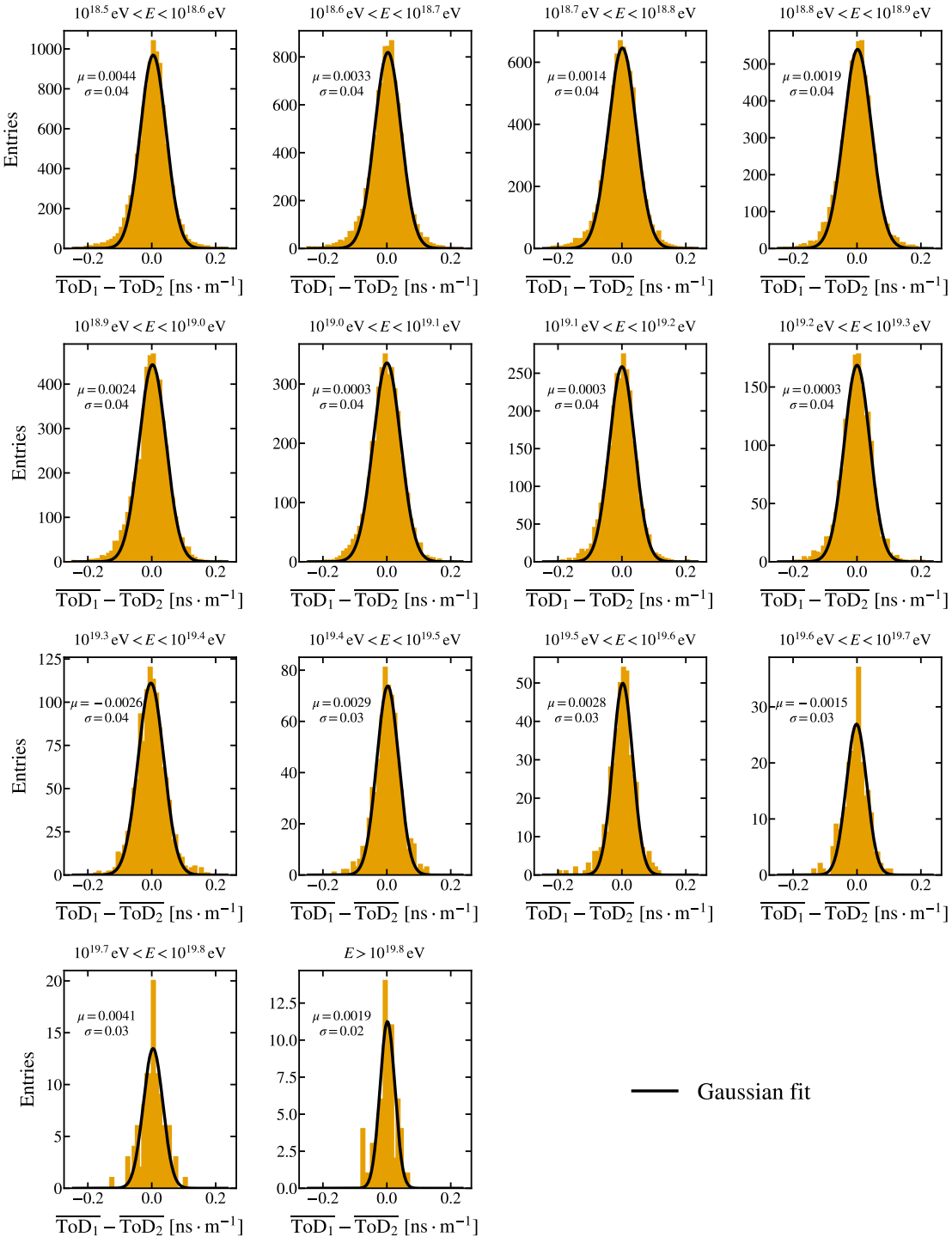

Figure 4: Distributions of the differences between $\overline{\mathrm{ToD}}_1$ and $\overline{\mathrm{ToD}}_2$ for data in each energy bin.

ANOVA - Additional plots

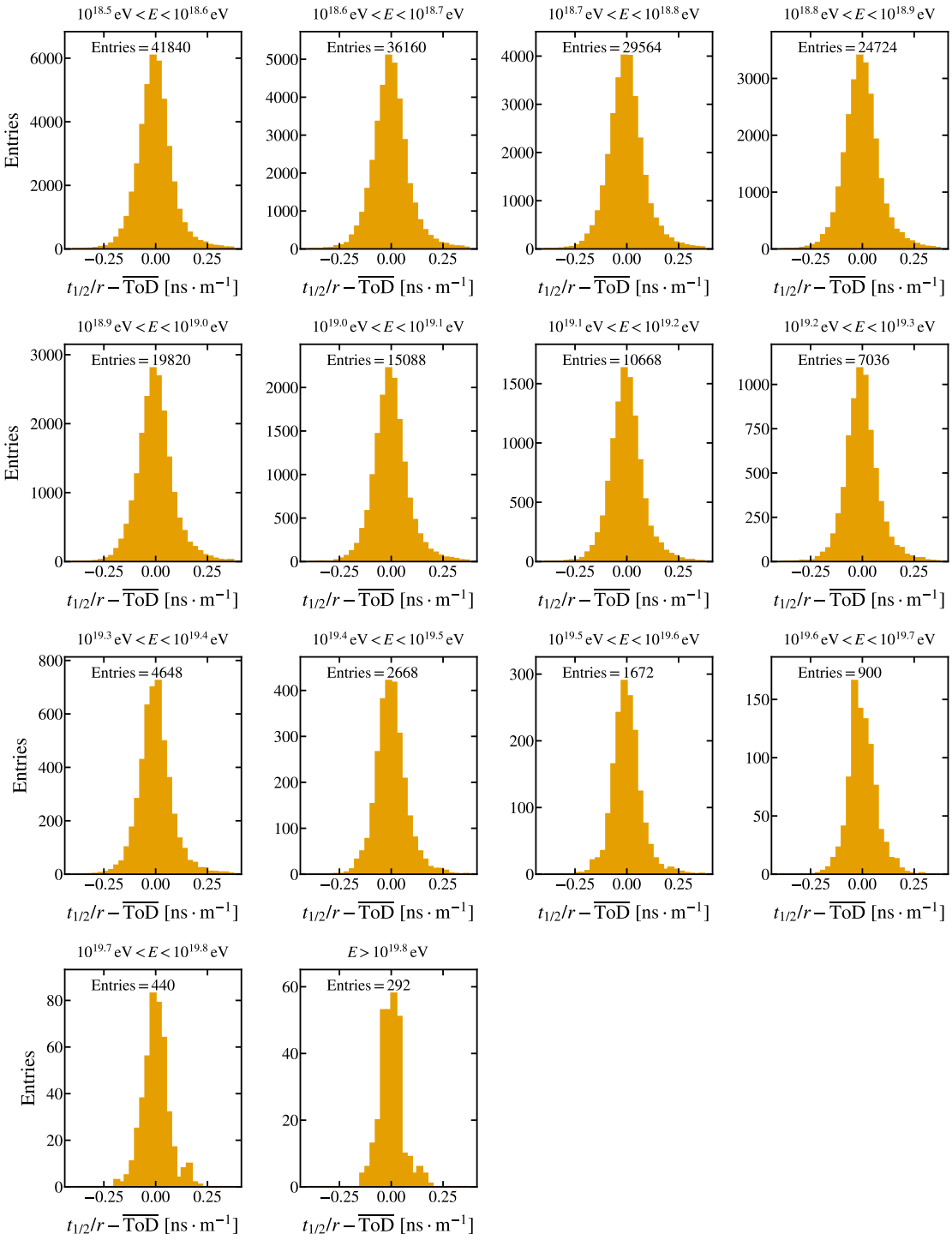

Figure 5: Distributions of the $t_{1/2}/r - \overline{\mathrm{ToD}}$. These values are used to compute σ^2_{det} in Equation 4.8.

Appendix for Chapter 7

Examples of traces

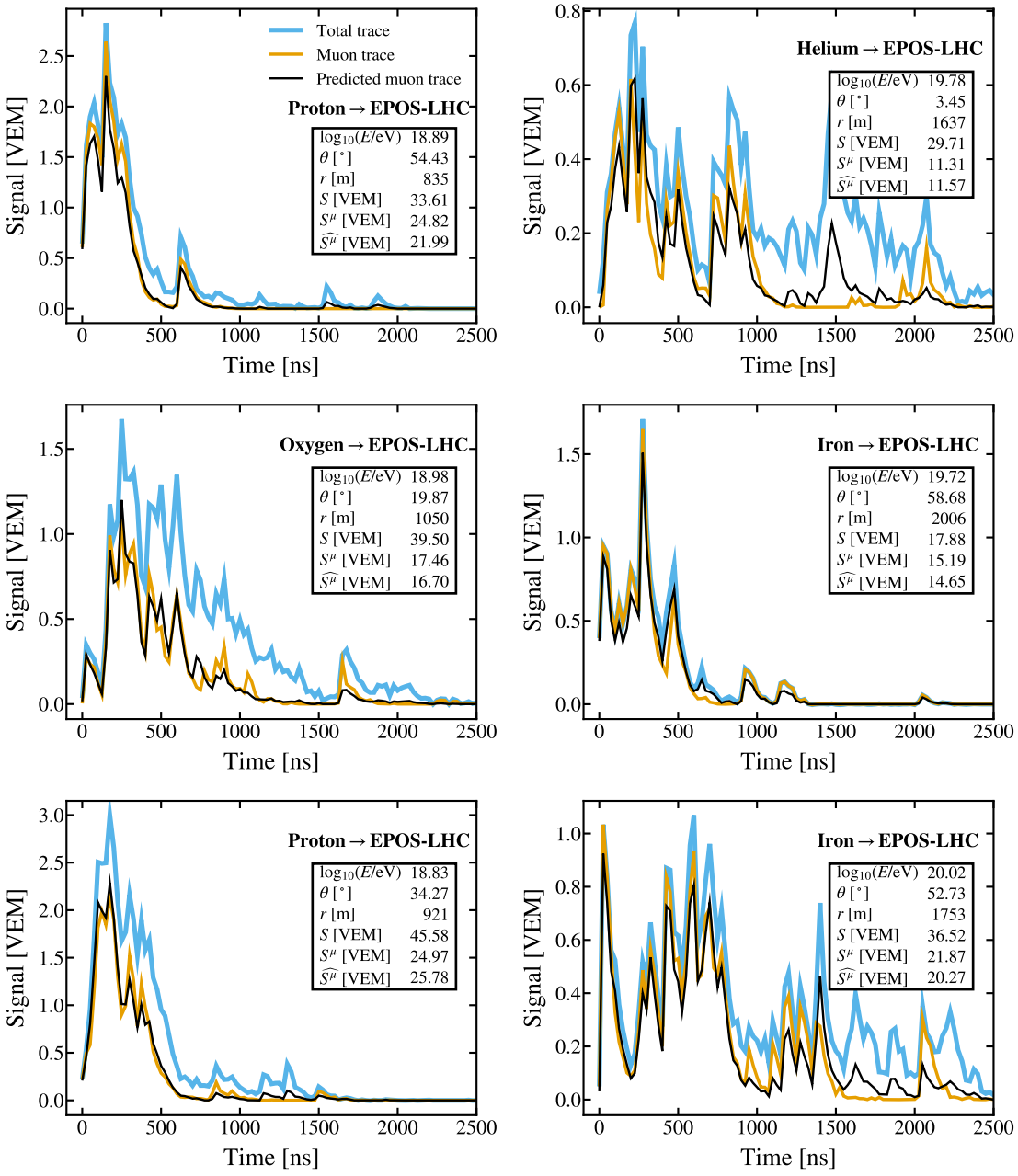

Figure 6: Examples of predicted muon traces for simulated events done with EPOS-LHC with different primaries. The prediction (black line) takes the shape of the true muon trace (orange line) accurately at most times. The blue thicker line corresponds to the total trace, the one that can be measured by the stations of the SD.

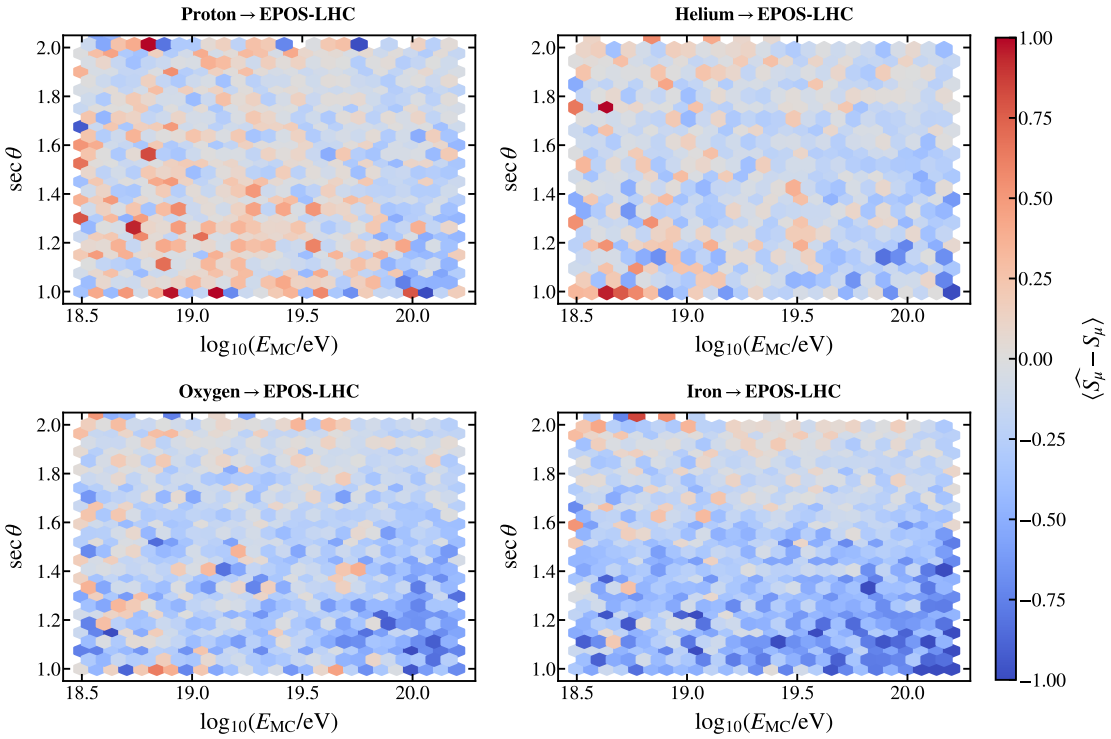

Figure 7: Mean of the difference between the predicted and true value of the muon signal as a function of the energy and zenith angle.

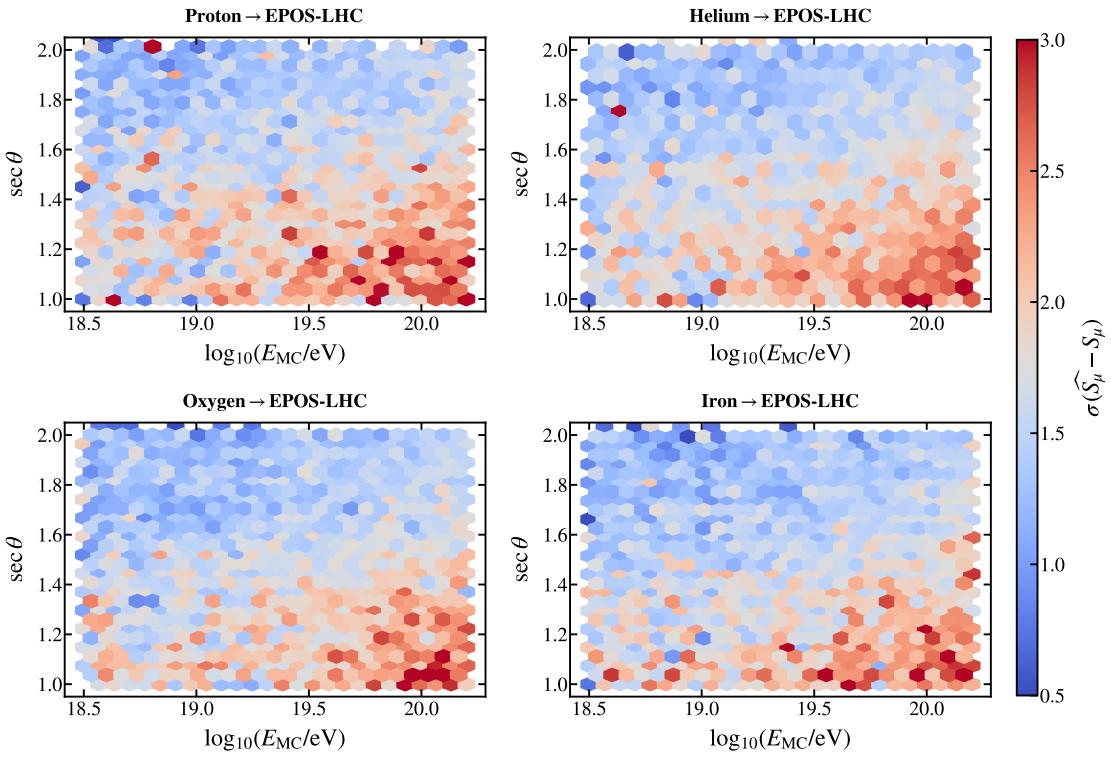

Figure 8: Standard deviation of the difference between the predicted and true value of the muon signal as a function of the energy and zenith angle.

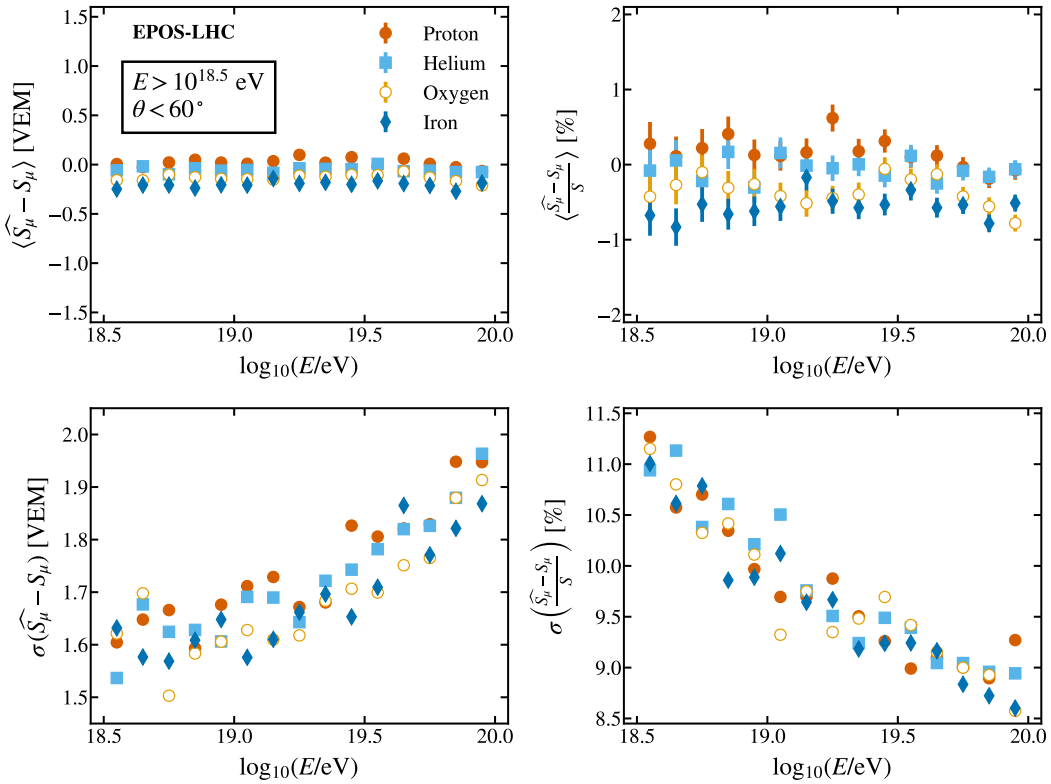

Figure 9: Mean value (first row) and standard deviation (second row) of the difference between the integral of the predicted muon trace and the true muon trace. They are shown as a function of the energy.

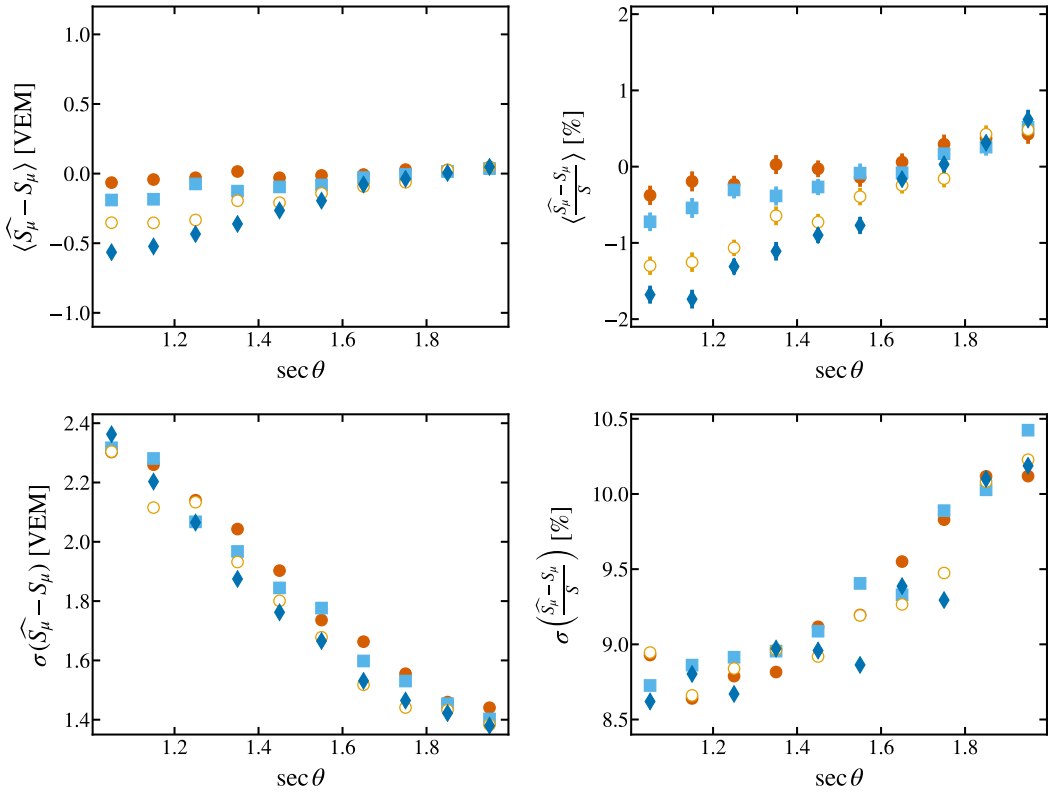

Figure 10: Mean value (first row) and standard deviation (second row) of the difference between the integral of the predicted muon trace and the true muon trace. They are shown as a function of the secant of the zenith angle.

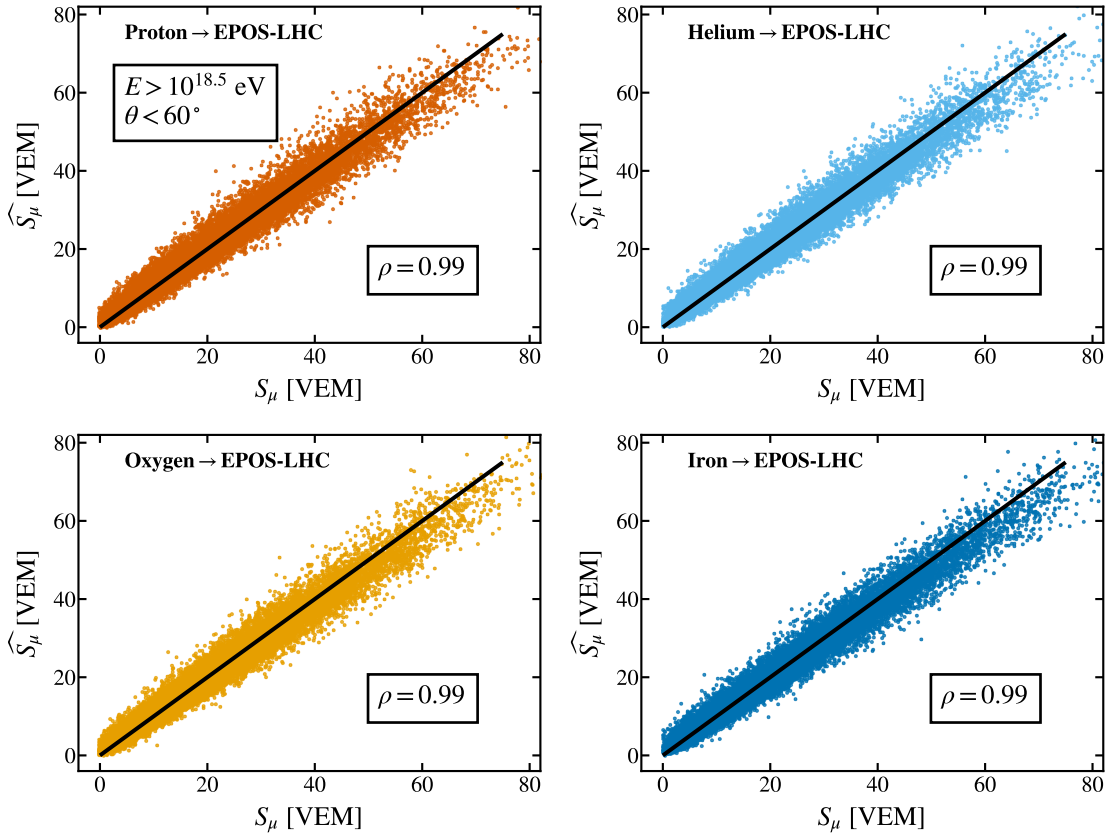

Figure 11: Integral of the predicted muon trace as a function of the integral of the true muon trace. The black line corresponds to a perfect prediction and ρ is the Pearson correlation coefficient.

Appendix for Chapter 8

Plots using EPOS-LHC

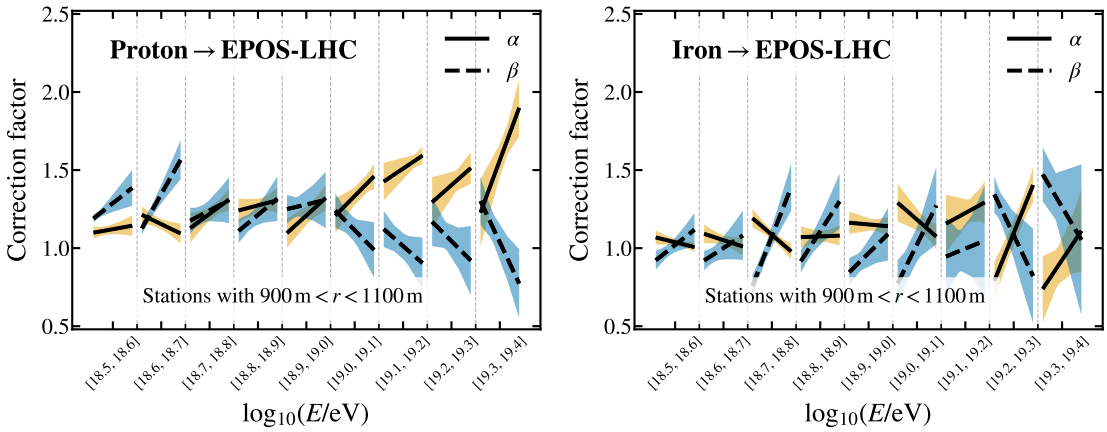

Figure 12: Evolution of α and β with the energy and zenith angle for simulations done with EPOS-LHC. For each energy bin, the zenith angle increases linearly in $\sec\theta$ from left to right from 1 to 2. Linear fits (black lines) are shown instead of the points for each energy bin. The shadowed area corresponds to the uncertainty of the fit.

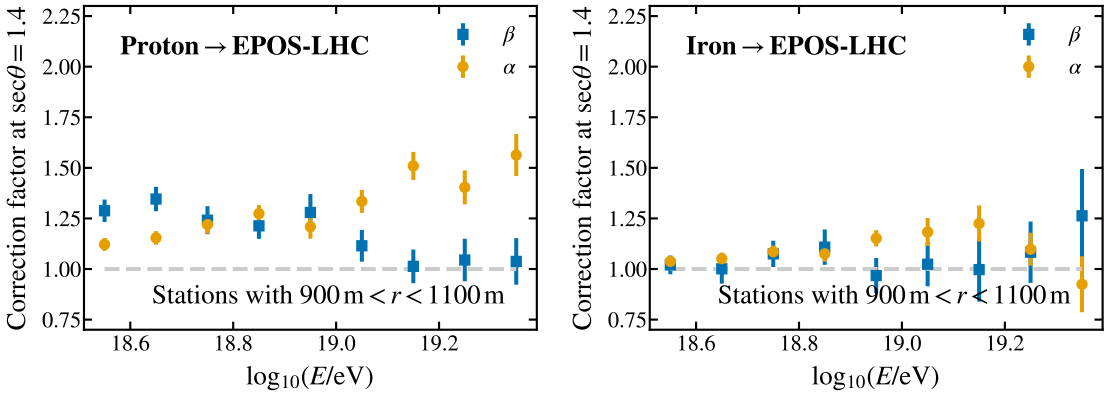

Figure 13: Evolution of α and β when using EPOS-LHC with the energy when picking the values of the fits in Figure 8.4 at $\sec\theta = 1.4$. The error bars correspond to the uncertainty of the fits.

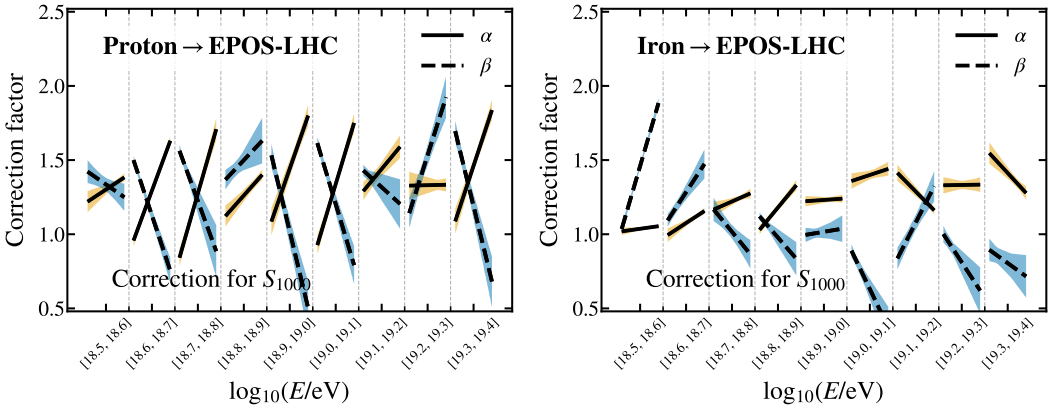

Figure 14: Evolution of α and β with the energy and zenith angle for simulations done with EPOS-LHC. For each energy bin, the zenith angle increases linearly in $\sec\theta$ from left to right from 1 to 2. Linear fits (black lines) are shown instead of the points for each energy bin. The shadowed area corresponds to the uncertainty of the fit.

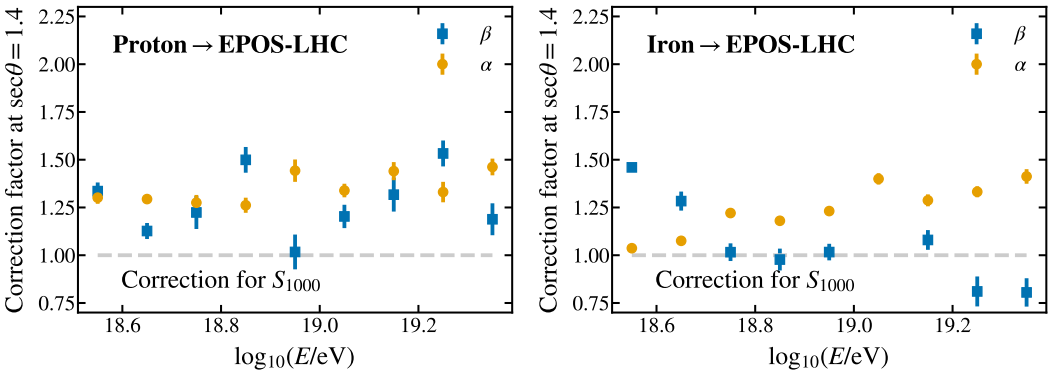

Figure 15: Evolution of α and β when using EPOS-LHC with the energy and zenith angle. For each energy bin, the zenith angle increases linearly in $\sec\theta$ from left to right from 1 to 2. Linear fits (black lines) are shown instead of the points for each energy bin. The shadowed area corresponds to the uncertainty of the fit.

List of Figures

List of Tables

Bibliography

[1] V. V. Ezhela et al. *Particle Physics: One Hundred Years of Discoveries (An Annotated Chronological Bibliography)*. ISBN 978-1-56396-642-2 (1996).

[2] W. Röntgen. *On a New Kind of Rays*. Nature, **53**, 274 (1896). DOI: 10.1038/053274b0.

[3] H. Becquerel. *Sur les radiations émises par phosphorescence*. Comptes Rendus de l'Acad. des Sciences, **122**, 420 (1896).

[4] P. Curie, M. P. Curie, and G. Bémont. *Sur une nouvelle substance fortement radioactive, contenue dans la pechblende*. Comptes Rendus de l'Acad. des Sciences, **127**, 1215 (1898).

[5] C. Coulomb. *Mdm. de l'Acad. des Sciences*. Académie Royale des Sciences de Paris, 612 (1785).

[6] C. T. R. Wilson. *On the ionization of Atmospheric Air*. Proc. Roy. Soc. London, **68**, 151 (1901).

[7] T. Wulf. Phys. Zeit. **1**, 152 (1909).

[8] D. Pacini. *Sulle radiazioni penetranti*. Ann. Uff. Centr. Meteor. **XXXII, parte I**, 123 (1910).

[9] D. Pacini. *La radiazione penetrante sul mare*. Ann. Uff. Centr. Meteor. **XXXII, parte I**, 93 (1912).

[10] D. Pacini. *La radiazione penetrante alla superficie ed in seno alle acque*. Nuovo Cim. **VI/3**, 93 (1912).

[11] A. Gockel. Phys. Zeit. **11**, 280 (1910).

[12] A. Gockel. Phys. Zeit. **12**, 595 (1911).

[13] V. Hess. Phys. Zeit. **13**, 1084 (1912).

[14] V. Hess. *On the Observations of the Penetrating Radiation during Seven Balloon Flights*. (2018). arXiv: 1808.02927.

[15] W. Kolhörster. Phys. Zeit. **14**, 1153 (1913).

[16] W. Kolhörster. Ber. deutsch. Phys. Ques. **16**, 719 (1914).

[17] B. Rossi. *Über die Eigenschaften der durchdringenden Korpuskularstrahlung im Meeresniveau*. Phys. Zeit. **82**, 151 (1933).

[18] P. Auger, R. Maze, and Robley. *Extension et pouvoir pénétrant des grandes gerbes de rayons cosmiques*. Comptes Rendus de l'Acad. des Sciences, **208**, 1641 (1939).

[19] K.-H. Kampert and A. A. Watson. *Extensive Air Showers and Ultra High-Energy Cosmic Rays: A Historical Review*. (2012). DOI: 10.1140/epjh/e2012-30013-x. arXiv: 1207.4827v1.

[20] Particle Data Group. Phys. Rev. D, **98**, 030001 (2018).

[21] W. Heitler. *The Quantum Theory of Radiation*. Dover Books on Physics and Chemistry (1954).

[22] T. K. Gaisser, R. Engel, and E. Resconi. *Cosmic Rays and Particle Physics*. Cambridge University Press (2016).

[23] J. Matthews. *A Heitler model of extensive air showers*. Astroparticle Physics, **22**, 387 (2005). DOI: 10.1016/j.astropartphys.2004.09.003.

[24] A. Hillas. *Shower simulation: lessons from MOCCA*. Nuclear Physics B - Proceedings Supplements, **52**, 29 (1997). DOI: 10.1016/S0920-5632(96)00847-X.

[25] E. S. Seo et al. *Measurement of Cosmic-Ray Proton and Helium Spectra during the 1987 Solar Minimum*. Astrophysical Journal, **378**, 763 (1991). DOI: 10.1086/170477.

[26] Grigorov et al. Proceedings of the 12th ICRC, Tasmania, Australia, 1760 (1971).

[27] A. A. Ivanov, S. P. Knurenko, and I. Y. Sleptsov. *Measuring extensive air showers with Cherenkov light detectors of the Yakutsk array: the energy spectrum of cosmic rays*. New Journal of Physics, **11**, 065008 (2009). DOI: 10.1088/1367-2630/11/6/065008. arXiv: 0902.1016.

[28] M. A. Lawrence, R. J. O. Reid, and A. A. Watson. *The cosmic ray energy spectrum above* $4 \cdot 10^{17}$ *eV as measured by the Haverah Park array*. Journal of Physics G: Nuclear and Particle Physics, **17**, 733 (1991). DOI: 10.1088/0954-3899/17/5/019.

[29] M. Nagano et al. *Energy spectrum of primary cosmic rays above* 10^{17} *eV determined from extensive air shower experiments at Akeno*. Journal of Physics G: Nuclear and Particle Physics, **18**, 423 (1992). DOI: 10.1088/0954-3899/18/2/022.

[30] D. J. Bird et al. *The Cosmic-Ray Energy Spectrum Observed by the Fly's Eye*. Astrophysical Journal, **424**, 491 (1994). DOI: 10.1086/173906.

[31] The HiRes Collaboration. *First Observation of the Greisen-Zatsepin-Kuzmin Suppression*. Phys. Rev. Lett. **100**, 101101 (2008). DOI: 10.1103/PhysRevLett.100.101101. arXiv: astro-ph/0703099.

[32] W. Hanlon and D. Ikeda. *Energy Spectrum and Mass Composition of Ultra-High Energy Cosmic Rays Measured by the hybrid technique in Telescope Array*. PoS, **ICRC2015**, 362 (2016).

[33] Valerio Verzi. *Measurement of the energy spectrum of ultra-high energy cosmic rays using The Pierre Auger Observatory* [137]. 450 (2019).

[34] A. De Angelis and M. Pimenta. *Introduction to Particle and Astroparticle Physics*. Springer, Second Edition (2018).

[35] The Pierre Auger Collaboration. *Observation of the Suppression of the Flux of Cosmic Rays above* $4 \cdot 10^{19}$ *eV*. Phys. Rev. Lett. **101**, 061101 (2008). DOI: 10.1103/PhysRevLett.101.061101. arXiv: 0806.4302.

[36] K. Greisen. *End to the cosmic ray spectrum?* Phys. Rev. Lett. **16**, 748 (1966). DOI: 10.1103/PhysRevLett.16.748.

[37] G. Zatsepin and V. Kuzmin. *Upper limit of the spectrum of cosmic rays*. JETP Lett. **4**, 78 (1966).

[38] K. Kampert. *The chemical composition of cosmic rays*. Workshop: Frontier Objects in Astrophysics and Particle Physics, 485 (2003). arXiv: astro-ph/0212348.

[39] A. Porcelli. *Measurements of* X_{max} *above* 10^{17} *eV with the fluorescence detector of the Pierre Auger Observatory* [138]. Talk at ICRC2015 (2015).

[40] Alexey Yushkov. *Mass composition of cosmic rays with energies above* $10^{17.2}$ *eV from the hybrid data of the Pierre Auger Observatory* [137]. 482 (2019).

[41] The Pierre Auger Collaboration. *Measurement of the Depth of Maximum of Extensive Air Showers above* 10^{18} *eV*. Phys. Rev. Lett. **104**, 091101 (2010). DOI: 10.1103/PhysRevLett.104.091101. arXiv: 1002.0699.

[42] The Pierre Auger Collaboration. *Depth of maximum of air-shower profiles at the Pierre Auger Observatory. I. Measurements at energies above* $10^{17.8}$ *eV*. Phys. Rev. D, **90**, 122005 (2014). DOI: 10.1103/PhysRevD.90.122005. arXiv: 1409.4809.

[43] P. Sanchez Lucas. *The* $\langle \Delta \rangle$ *Method: An estimator for the mass composition of ultra-high-energy cosmic rays*. PhD Thesis. University of Granada (2016).

[44] The Pierre Auger Collaboration. *Inferences on mass composition and tests of hadronic interactions from* 0.3 *to* 100 *EeV using the water-Cherenkov detectors of the Pierre Auger Observatory*. Phys. Rev. D, **96**, 122003 (2017). DOI: 10.1103/PhysRevD.96.122003. arXiv: 1710.07249.

[45] The Pierre Auger Collaboration. *Evidence for a mixed mass composition at the 'ankle' in the cosmic-ray spectrum*. Physics Letters B, **762**, 288 (2016). DOI: 10.1016/j.physletb.2016.09.039. arXiv: 1609.08567.

[46] R. A. Gideon and R. A. Hollister. *A Rank Correlation Coefficient Resistant to Outliers*. Journal of the American Statistical Association, **82**, 656 (1987). DOI: 10.1080/01621459.1987.10478480.

[47] T. Pierog, I. Karpenko, J. M. Katzy, E. Yatsenko, and K. Werner. *EPOS LHC: Test of collective hadronization with data measured at the CERN Large Hadron Collider*. Physical Review C, **92**, 034906 (2015). DOI: 10.1103/PhysRevC.92.034906. arXiv: 1306.0121.

[48] F. Riehn et al. *The hadronic interaction model Sibyll 2.3c and Feynman scaling* [139]. 301 (2017). arXiv: 1709.07227.

[49] A. M. Hillas. *The Origin of Ultra-High-Energy Cosmic Rays*. Annual Review of Astronomy and Astrophysics, **22**, 425 (1984). DOI: 10.1146/annurev.aa.22.090184.002233.

[50] The Pierre Auger Collaboration. *Observation of a large-scale anisotropy in the arrival directions of cosmic rays above* $8{\cdot}10^{18}$ *eV*. Science, **357**, 1266 (2017). DOI: 10.1126/science.aan4338. arXiv: 1709.07321.

[51] S. Ostapchenko. *Monte Carlo treatment of hadronic interactions in enhanced Pomeron scheme: QGSJET-II model*. Phys. Rev. D, **83**, 014018 (2011). DOI: 10.1103/PhysRevD.83.014018. arXiv: 1010.1869.

[52] The Pierre Auger Collaboration. *Testing Hadronic Interactions at Ultrahigh Energies with Air Showers Measured by the Pierre Auger Observatory*. Phys. Rev. Lett. **117**, 192001 (2016). DOI: 10.1103/PhysRevLett.117.192001. arXiv: 1610.08509.

[53] The Pierre Auger Collaboration. *Muons in air showers at the Pierre Auger Observatory: Mean number in highly inclined events*. Phys. Rev. D, **91**, 032003 (2015). DOI: 10.1103/PhysRevD.91.032003. arXiv: 1408.1421.

[54] The Telescope Array Collaboration. *Study of muons from ultra-high energy cosmic ray air showers measured with the Telescope Array experiment*. Phys. Rev. D, **98**, 022002 (2018). DOI: 10.1103/PhysRevD.98.022002. arXiv: 1804.03877.

[55] J. Linsley. *Evidence for a Primary Cosmic-Ray Particle with Energy* 10^{20} *eV*. Phys. Rev. Lett. **10**, 146 (1963). DOI: 10.1103/PhysRevLett.10.146.

[56] P. Sommers. *Capabilities of a giant hybrid air shower detector*. Astroparticle Physics, **3**, 349 (1995). DOI: 10.1016/0927-6505(95)00013-7.

[57] B. Dawson, H. Dai, P. Sommers, and S. Yoshida. *Simulations of a giant hybrid air shower detector*. Astroparticle Physics, **5**, 239 (1996). DOI: 10.1016/0927-6505(96)00024-2.

[58] The Pierre Auger Collaboration. *The Pierre Auger Cosmic Ray Observatory*. Nucl. Instrum. Meth. **A798**, 172 (2015). DOI: https://doi.org/10.1016/j.nima.2015.06.058. arXiv: 1502.01323.

[59] I. Allekotte et al., for the Pierre Auger Collaboration. *The surface detector system of the Pierre Auger Observatory*. Nucl. Instrum. Meth. **A586**, 409 (2008). DOI: 10.1016/j.nima.2007.12.016. arXiv: 0712.2832.

[60] D. Veberič. *Estimation of the Total Signals in Saturated Stations of the Pierre Auger Observatory* [140]. 0633 (2013).

[61] X. Bertou et al., for the Pierre Auger Collaboration. *Calibration of the surface array of the Pierre Auger Observatory*. Nucl. Instrum. Meth. **A568**, 839 (2006). DOI: 10.1016/j.nima.2006.07.066.

[62] The Pierre Auger Collaboration. *Trigger and aperture of the surface detector array of the Pierre Auger Observatory*. Nucl. Instrum. Meth. **A613**, 29 (2010). DOI: 10.1016/j.nima.2009.11.018. arXiv: 1111.6764.

[63] The Pierre Auger Collaboration. *The Fluorescence Detector of the Pierre Auger Observatory*. Nucl. Instrum. Meth. **A620**, 227 (2010). DOI: 10.1016/j.nima.2010.04.023. arXiv: 0907.4282.

[64] The Pierre Auger Collaboration. *Properties and performance of the prototype instrument for the Pierre Auger Observatory*. Nucl. Instrum. Meth. **A523**, 50 (2004). DOI: 10.1016/j.nima.2003.12.012.

[65] J. Bauml. *Measurement of the Optical Properties of the Auger Fluorescence Telescopes* [140]. 0806 (2013).

[66] J. Rautenberg. *Remote operation of the Pierre Auger Observatory* [141]. 1203 (2011).

[67] K. Kamata and J. Nishimura. *The Lateral and the Angular Structure Functions of Electron Showers*. Proc. Theor. Phys. Supplement, **6**, 93 (1958). DOI: 10.1143/PTPS.6.93.

[68] K. Greisen. *The extensive air showers*. Progress in Cosmic Ray Physics, **3**, 1 (1956).

[69] D. Newton, J. Knapp, and A. A. Watson. *The optimum distance at which to determine the size of a giant air shower*. Astroparticle Physics, **26**, 414 (2007). DOI: 10.1016/j.astropartphys.2006.08.003. arXiv: astro-ph/0608118.

[70] C. Bonifazi, A. Letessier-Selvon, and E. Santos, for the Pierre Auger Collaboration. *A model for the time uncertainty measurements in the Auger surface detector array*. Astroparticle Physics, **28**, 523 (2008). DOI: 10.1016/j.astropartphys.2007.09.007. arXiv: 0705.1856.

[71] C. Bonifazi, for the Pierre Auger Collaboration. *The angular resolution of the Pierre Auger Observatory*. Nuclear Physics B Proceedings Supplements, **190**, 20 (2009). DOI: 10.1016/j.nuclphysbps.2009.03.063. arXiv: 0901.3138.

[72] J. Hersil, I. Escobar, D. Scott, G. Clark, and S. Olbert. *Observations of Extensive Air Showers near the Maximum of Their Longitudinal Development*. Phys. Rev. Lett. **6**, 22 (1961). DOI: 10.1103/PhysRevLett.6.22. URL: https://link.aps.org/doi/10.1103/PhysRevLett.6.22.

[73] A. Schulz. *Measurement of the Energy Spectrum of Cosmic Rays above* $3{\cdot}10^{17}$ *eV with the Pierre Auger Observatory* [140]. 0769 (2013).

[74] R. Pesce. *Energy calibration of data recorded with the surface detectors of the Pierre Auger Observatory: an update* [141]. 1160 (2011).

[75] M. Tueros. *Estimate of the non-calorimetric energy of showers observed with the fluorescence and surface detectors of the Pierre Auger Observatory* [140]. 0705 (2013).

[76] H. Dembinski. *The Cosmic Ray Spectrum above* $4{\cdot}10^{18}$ *eV as measured with inclined showers recorded at the Pierre Auger Observatory* [141]. 0724 (2011).

[77] V. Verzi. *The Energy Scale of the Pierre Auger Observatory* [140]. 0928 (2013).

[78] I. Valiño. *A measurement of the muon number in showers using inclined events recorded at the Pierre Auger Observatory* [140]. 0635 (2013).

[79] The Pierre Auger Collaboration. *Reconstruction of inclined air showers detected with the Pierre Auger Observatory*. Journal of Cosmology and Astroparticle Physics, **2014**, 019 (2014). DOI: 10.1088/1475-7516/2014/08/019. arXiv: 1407.3214.

[80] F. Suarez. *The AMIGA muon detectors of the Pierre Auger Observatory: overview and status* [140]. 0712 (2013).

[81] F. Sánchez. *The AMIGA detector of the Pierre Auger Observatory: an overview* [141]. 0742 (2011).

[82] M. Platino et al. *AMIGA at the Auger Observatory: the scintillator module testing system*. Journal of Instrumentation, **6**, P06006 (2011). DOI: 10.1088/1748-0221/6/06/p06006.

[83] I. Mariş. *The AMIGA infill detector of the Pierre Auger Observatory: performance and first data* [141]. 0711 (2011).

[84] M. Nagano et al. *Energy spectrum of primary cosmic rays above* 10^{17} *eV determined from extensive air shower experiments at Akeno*. Journal of Physics G: Nuclear and Particle Physics, **18**, 423 (1992). DOI: 10.1088/0954-3899/18/2/022.

[85] S. Maldera. *Measuring the accuracy of the AMIGA muon counters at the Pierre Auger Observatory* [140]. 0748 (2013).

[86] D. Ravignani. *Measurement of the energy spectrum of cosmic rays above* $3{\cdot}10^{17}$ *eV using the AMIGA 750 m surface detector array of the Pierre Auger Observatory* [140]. 0693 (2013).

[87] A. D. Supanitsky et al. *Underground muon counters as a tool for composition analyses*. Astroparticle Physics, **29**, 461 (2008). DOI: 10.1016/j.astropartphys.2008.05.003. arXiv: 0804.1068.

[88] F. Sánchez. *The muon component of extensive air showers above* $10^{17.5}$ *eV measured with The Pierre Auger Observatory* [137]. 411 (2019).

[89] A. Watson and J. Wilson. *Fluctuation studies of large air showers: the composition of primary cosmic ray particles of energy* E_p *approximately* 10^{18} *eV*. J. Phys. B, **7**, 1199 (1974).

[90] The Pierre Auger Collaboration. *Azimuthal asymmetry in the risetime of the surface detector signals of the Pierre Auger Observatory*. Phys. Rev. D, **93**, 072006 (2016). DOI: 10.1103/PhysRevD.93.072006. arXiv: 1604.00978.

[91] H. Cook. *GAP Report 107*. Pierre Auger Collaboration internal publication, unpublished (2012).

[92] I. Lhenry-Yvon. *GAP Report 033*. Pierre Auger Collaboration internal publication, unpublished (2016).

[93] The Pierre Auger Collaboration. *Interpretation of the depths of maximum of extensive air showers measured by the Pierre Auger Observatory*. JCAP, **2013**, 026 (2013). DOI: 10.1088/1475-7516/2013/02/026. arXiv: 1301.6637.

[94] A. Bueno, J. M. Carceller, P. Sanchez-Lucas, and A. A. Watson. *GAP Report 068*. Pierre Auger Collaboration internal publication, unpublished (2017).

[95] M. D. Castro and E. M. Santos. *GAP Report 031*. Pierre Auger Collaboration internal publication, unpublished (2016).

[96] A. Bueno, D. García-Gámez, L. Molina, and B. Zamorano. *GAP Report 007*. Pierre Auger Collaboration internal publication, unpublished (2014).

[97] A. Watson and J. Wilson. *Fluctuation studies of large air showers: the composition of primary cosmic ray particles of energy* $E_p \sim 10^{18}$ *eV*. J. Phys., B (London), **7**, 1199 (1974). DOI: 10.1088/0305-4470/7/10/013.

[98] A. Bueno, J. M. Carceller, and A. A. Watson. *GAP Report 042*. Pierre Auger Collaboration internal publication, unpublished (2018).

[99] G. Carleo et al. *Machine learning and the physical sciences*. Rev. Mod. Phys. **91**, 045002 (2019). DOI: 10.1103/RevModPhys.91.045002. arXiv: 1903.10563.

[100] I. Goodfellow, Y. Bengio, and A. Courville. *Deep Learning*. MIT Press (2016). URL: http://www.deeplearningbook.org.

[101] A. Ng. *Linear Regression and Gradient Descent*. Stanford CS229: Machine Learning (Autumn 2018). Accesed 31-05-2020. URL: https://www.youtube.com/watch?v=4b4MUYve_U8.

[102] T. Hastie, R. Tibshirani, and J. Friedman. *The Elements of Statistical Learning. Data Mining, Inference and Prediction*. Springer, Second Edition (2017).

[103] M. Nielsen. *Using neural nets to recognize handwritten digits*. Accesed 31-05-2020. URL: http://neuralnetworksanddeeplearning.com/chap1.html.

[104] J. Schmidhuber. *Deep Learning in Neural Networks: An Overview*. Neural Networks, **61**, 85 (2015). DOI: 10.1016/j.neunet.2014.09.003. arXiv: 1404.7828.

[105] Y. LeCun, Y. Bengio, and G. Hinton. *Deep Learning*. Nature, **521**, 436 (2015). DOI: 10.1038/nature14539.

[106] *Heaviside step function*. Accesed 27-04-2020. URL: https://en.wikipedia.org/wiki/Heaviside_step_function.

[107] M. Nielsen. *How the backpropagation algorithm works*. Accesed 31-05-2020. URL: http://neuralnetworksanddeeplearning.com/chap2.html.

[108] *Kronecker delta*. Accesed 27-04-2020. URL: https://en.wikipedia.org/wiki/Kronecker_delta.

[109] *Pytorch*. Accesed 21-04-2020. URL: https://pytorch.org/.

[110] *TensorFlow*. Accesed 21-04-2020. URL: https://www.tensorflow.org/.

[111] *Keras*. Accesed 21-04-2020. URL: https://keras.io/.

[112] The Pierre Auger Collaboration. *The Pierre Auger Observatory Upgrade - Preliminary Design Report*. (2016). arXiv: 1604.03637.

[113] P. Virtanen et al. *SciPy 1.0: Fundamental Algorithms for Scientific Computing in Python*. Nature Methods, **17**, 261 (2020). DOI: 10.1038/s41592-019-0686-2.

[114] F. Pedregosa et al. *Scikit-learn: Machine Learning in Python*. Journal of Machine Learning Research, **12**, 2825 (2011).

[115] F. Chollet et al. *Keras*. 2015. URL: https://keras.io.

[116] *Python*. Accesed 02-04-2020. URL: https://www.python.org/.

[117] V. Nair and G. E. Hinton. *Rectified linear units improve restricted boltzmann machines*. Proceedings of the 27th International Conference on International Conference on Machine Learning, 807 (2010).

[118] X. Glorot, A. Bordes, and Y. Bengio. *Deep Sparse Rectifier Neural Networks*. Proceedings of the Fourteenth International Conference on Artificial Intelligence and Statistics, **15**, 315 (2011).

[119] *Softmax function*. Accesed 27-04-2020. URL: https://en.wikipedia.org/wiki/Softmax_function.

[120] G. Klambauer, T. Unterthiner, A. Mayr, and S. Hochreiter. *Self-Normalizing Neural Networks*. Proceedings of Advances in Neural Information Processing Systems, **30** (2017). arXiv: 1706.02515.

[121] D. P. Kingma and J. Ba. *Adam: A Method for Stochastic Optimization*. Conference paper in ICLR 2015 (2015). arXiv: 1412.6980.

[122] M. Mitchell. *An Introduction to Genetic Algorithms*. MIT Press (1996).

[123] J. González et al. *Improving the accuracy while preserving the interpretability of fuzzy function approximators by means of multi-objective evolutionary algorithms*. International Journal of Approximate Reasoning, **44**, 32 (2007). DOI: 10.1016/j.ijar.2006.02.006.

[124] A. Guillén et al. *Parallel multiobjective memetic RBFNNs design and feature selection for function approximation problems*. Neurocomputing, **72**, 3541 (2009). DOI: 10.1016/j.neucom.2008.12.037.

[125] A. Guillén et al. *Evolutive Approaches for Variable Selection Using a Non-parametric Noise Estimator*. Studies in Computational Intelligence, **415**, 243 (2012). DOI: 10.1007/978-3-642-28789-3-11.

[126] A. Guillén, D. Sovilj, A. Lendasse, F. Mateo, and I. Rojas. *Minimising the delta test for variable selection in regression problems*. International Journal of High Performance Systems Architecture, **1**, 269 (2008). DOI: 10.1504/IJHPSA.2008.024211.

[127] S. Argirò et al. *The offline software framework of the Pierre Auger Observatory*. Nucl. Instrum. Meth. **A580**, 1485 (2007). DOI: 10.1016/j.nima.2007.07.010. arXiv: 0707.1652.

[128] D. Heck, G. Schatz, T. Thouw, J. Knapp and J. Capdevielle. *CORSIKA: A Monte Carlo code to simulate extensive air showers*. Accesed 02-04-2020. URL: https://www.ikp.kit.edu/corsika/.

[129] A. Guillén et al. *Deep learning techniques applied to the physics of extensive air showers*. Astroparticle Physics, **111**, 12–22 (2019). DOI: 10.1016/j.astropartphys.2019.03.001. arXiv: 1807.09024.

[130] A. Karpathy. *The Unreasonable Effectiveness of Recurrent Neural Networks*. Accesed 18-05-2020. URL: https://karpathy.github.io/2015/05/21/rnn-effectiveness.

[131] S. Hochreiter and J. Schmidhuber. *Long Short-term Memory*. Neural computation, **9**, 1735 (1997). DOI: 10.1162/neco.1997.9.8.1735.

[132] C. Olah. *Understanding LSTM Networks*. Accesed 18-05-2020. URL: https://colah.github.io/posts/2015-08-Understanding-LSTMs.

[133] E. W. Kellermann and L. Towers. *The electromagnetic component of large air showers*. Journal of Physics A: General Physics, **3**, 284 (1970). DOI: 10.1088/0305-4470/3/3/015.

[134] N. Hayashida et al. *Muons ($\geq$ 1 GeV) in large extensive air showers of energies between $10^{16.5}$ eV and $10^{19.5}$ eV observed at Akeno*. Journal of Physics G: Nuclear and Particle Physics, **21**, 1101 (1995). DOI: 10.1088/0954-3899/21/8/008.

[135] K. Greisen. *Cosmic Ray Showers*. Annual Review of Nuclear Science, **10**, 63 (1960). DOI: 10.1146/annurev.ns.10.120160.000431.

[136] J. Vicha, A. Yushkov, D. Nosek, and P. Travnicek. *GAP Report 058*. Pierre Auger Collaboration internal publication, unpublished (2018).

[137] The Pierre Auger Collaboration. *The Pierre Auger Observatory: Contributions to the 36th International Cosmic Ray Conference (ICRC 2019)*. PoS, **ICRC2019** (2019). arXiv: 1909.09073v1.

[138] The Pierre Auger Collaboration. *The Pierre Auger Observatory: Contributions to the 34th International Cosmic Ray Conference (ICRC 2015)*. PoS, **ICRC2015** (2015). arXiv: 1509.03732.

[139] The Pierre Auger Collaboration. *The Pierre Auger Observatory: Contributions to the 35th International Cosmic Ray Conference (ICRC 2017)*. PoS, **ICRC2017** (2017). arXiv: 1708.06592.

[140] The Pierre Auger Collaboration. *The Pierre Auger Observatory: Contributions to the 33rd International Cosmic Ray Conference (ICRC 2013)*. PoS, **ICRC2013** (2013). arXiv: 1307.5059.

[141] The Pierre Auger Collaboration. *The Pierre Auger Observatory: Contributions to the 32nd International Cosmic Ray Conference (ICRC 2011)*. PoS, **ICRC2011** (2011). arXiv: 1107.4806.