Facultad de Ciencias
Universidad de Cantabria

Statistical approach to muography as a non-destructive testing technique for industry problem solving

(Título en español)

Trabajo de fin de Máster
para acceder al

MÁSTER UNIVERSITARIO EN CIENCIA DE DATOS

Autor : Cédric PRIEËLS

Director : Pablo MARTÍNEZ RUIZ DEL ÁRBOL
Co-director : CARLOS DÍEZ

Julio 2020

Abstract

In this work, cosmic muons have been used to perform a muography experiment, a non-destructive method allowing us to study the internal properties of several physical objects with various geometries. Such study is extremely useful in the industry, in order for example to study the degradation of pipes, and finds many application in our daily lives.

In particular, a new C++ framework has been developped in order to study the results of such muography experiments and study their results from a statistical point of view, to find the optimal geometrical parameters of different objects placed between two muon detectors.

Key words : cosmic muons, muography, geometry, likelihood

Resumen

Palabras claves :

Acknowledgments

I would like to mainly thank Pablo for his outstanding work as the director of this work over the past few months. He has been incredibly helpful and present all along the way, teaching me the basic of muography and helping and guiding me when needed. I know it has not always been easy but I am very grateful for the time he spent with me, and I am more than happy with the final result and glad I decided to choose this particular subject.

Thank you to Muon Systems and its CEO Carlos Díez in particular, the co-director of this work, for allowing me be a part of their experiment during a few months and allowing me to get a privileged access to actual data collected. Having the opportunity to work in collaboration with a real industry has been an extremely interesting experience that will for sure help me a lot in the future. I obviously wish to the company all the best since the technology developped definitely show a lot of potential.

I would also like to thank all the other students involved in this Master degree, in particular to Pedro and Nicolò which have been extremly helpful as well. Many thanks to Lara for organizing this degree in the best way possible given the special circonstances which happened this year and for setting up everything allowing me to follow the classes remotely during the first semester.

Finally, many thanks to my parents and family in general for being there whenever I need it and for supporting my decision to come and study in Spain 6 years ago. It has been a great experience that tought me a lot and that I will for sure never forget.

Contents

Chapter 1

Introduction

Nowadays, muon tomography is an active field of research since it is a non destructive method allowing to map the inside of large objects difficult and/or dangerous to access, without any contact or damage, and without even having physical access to them. To do so, this technique uses muons, elementary particles similar to electrons but with a much greater mass, which allows them to penetrate much deeper and probe matter in a more efficient way, since they suffer less from the bremsstrahlung radiation affecting all leptons. Such technology is relatively well-know and presents several advantages over other techniques such as X-ray imaging since it is globally safe and clean and uses natural radiation, cosmic muons, while providing an excellent penetration in matter in order to study it.

In this particular work, muon tomography will be applied to industry and used in order to probe several kinds of physical objects, such as pipes, to study their degradation without damaging them. This is the main objective of the company called Muon systems, founded in 2017 and based in Bilbao.

This company already built powerful muon detectors allowing to perform such experiment and developped over the years a powerful framework allowing to study the degradation of pipes using convolutional neural networks, but this method is limited in the sense that training such networks is computing intensive and time consuming, and therefore does not allow to study several properties of such pipes at once. In this work, an alternative method based on a statistical analysis of the problem is presented, allowing a greater generalization of the previous method to other object geometries and to a higher dimension parameters phase space.

This project is highly relevant in today's society because it allows to combine data science algorithms and high-performing statistical methods that have been developped in this context to solver this particular muography problem, putting them in practice in an actual industry.

Chapter 2

Muons and muography

The main particles used throughout this work, muons, will be briefly introduced in Section 2.1 of this Chapter. Muon are particles produced naturally by cosmic rays, introduced in Section 2.2 and, once produced, they obviously interact when traversing the atmosphere and matter in general in ways described in Section 2.3 that actually need to be understood extremely well for the muon tomography process introduced in Section 2.4 to be useful and appliable to industry. Finally, the actual experimental setup used for this work will be explained in Section 2.5.

2.1 Particle physics and muons

Particle physics is the field which studies the matter surrounding us, along with the fundamental interactions between the particles. In this context, the Standard Model of particle physics [1] is nowadays the most accepted mathematical model used to describe the elementary particles and three of the 4 fundamental forces of nature (electromagnetic, weak and strong interactions, while the gravitational interaction is out of reach of this model). Even though quite simple in concept, it has been able to describe most of the phenomena observed in nature so far with an incredible level of precision, and has made a lot of predictions that have now been proven to be true, such as the discovery of the top quark [2] in 1995, the tau neutrino [3] in 2001 and the Higgs boson itself [4, 5], the last missing piece of the Standard Model, annouced at CERN in July 2012.

According to this model, 12 different fermions (along with their 12 corresponding anti-particles) exist in nature, as shown in Figure 2.1, most of them being unstable. These fermions can be divided into two fundamentally different categories, the quarks and the leptons, containing each 6 particles and sensitive to different forces. Even though quite interesting, the quarks do not play a fundamantal role in the muon tomography detailed in this work, so only leptons will be considered from now on. In particular, leptons can be divided even more into three different generations of particles, and the muon, one particular lepton belonging to the second generation, will be the main focus of this work.

Muons μ^- [9] are therefore fundamental particles having a negative charge and quite similar in nature to electrons, even though they have a quite high mass (200 times larger than the electron), which implies that their are not stable particles: they have a lifetime of approximately 2.2μs, and typically decay into an electron and a pair of neutrinos. However, this lifetime is actually quite long with respect to other fundamental particles and muons are on average able to travel more than 700 meters, allowing us to consider them to be stable particles in many processes, such as the one presented in this work. Muons also have a relatively small interaction cross-section with ordinary matter, even though they do interact with baryonic matter by several processes described in Section 2.3.

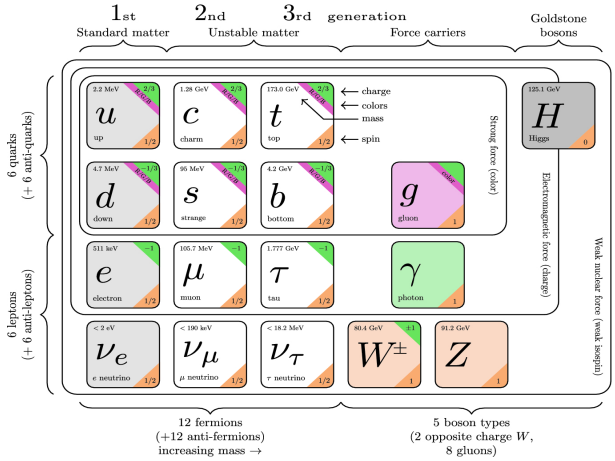


Figure 2.1: Representation of the 12 fermions of the Standard Model [8] along with the main force carriers and the Higgs boson, discovered in 2012 and completing this model.

2.2 Cosmic rays

Being unstable by nature, once produced, muons decay almost instantly by a weak process into an electron and a pair of neutrinos. However, it is possible to observe them in nature, since they are continually produced, mainly thanks to cosmic rays [10], constant flux of high energy particles (mostly protons and atomic nuclei) coming mostly from supernovae explosions and AGN emissions and reaching the Earth every day. Indeed, as they impact our atmosphere, these particles start a chain reaction, as shown in Figure 2.2: first of all, several neutral and charged pions are produced, decaying themselves into a pair of photons (and, later on, electron and positron pairs) and muons, respectively.

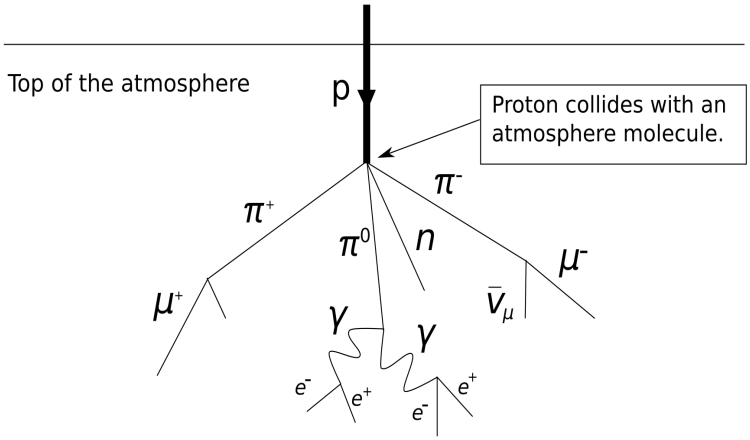


Figure 2.2: Typical chain of decays induced by highly energetic cosmic rays when reaching the Earth's atmosphere.

Muons are the most abundant charged particles produced by these processes actually reaching the sea level, as shown in Figure 2.3. Even though they are unstable and have a limited lifetime, around 0.06% of muons produced by such processes indeed do manage to reach the sea level thanks to the temporal distorsion induced by their high energy and relativistic speed. As a rule of thumb, one can expect to observe 10.000 muons per square meter and per minute at the sea level. Muons were actually discovered thanks to cosmic rays in 1936 [11].

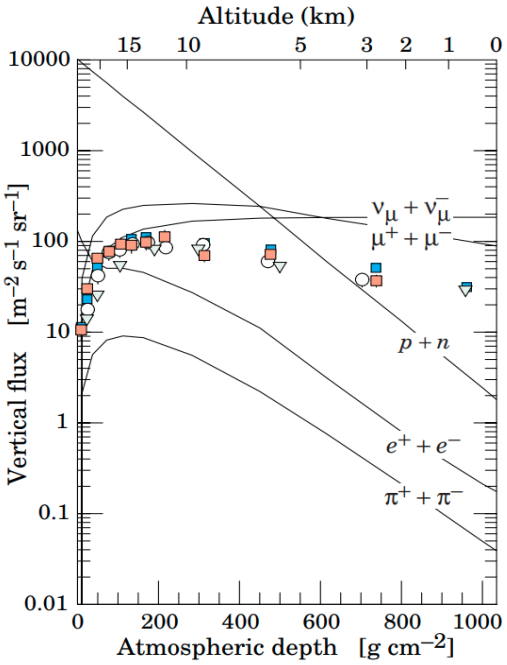


Figure 2.3: Abundance of particles observed at the sea level, due to cosmic rays [10].

With this work, millions of cosmic muons will be generated in order to simulate an actual experiment. This will allow us to check several of the well-known properties of such cosmic muons in Section 5.2, such as their distribution angle, following a $\cos^2$ law, and their exponentially falling energy spectrum.

2.3 Muons interaction with matter

Muons typically interact with ordinary matter through two main processes: ionization and multiple scattering, both resulting in different effects on the incident particle.

2.3.1 Ionization process

First of all, cosmic muons can interact with matter through **ionization**, when the incident muon gives some of its energy to the electrons of the absorber. This process is well described by the famous Bethe-Bloch formula shown in Equation 2.1 [9], relating the average loss of energy over a distance $\frac{dE}{dx}$ (typically referred to as the **mass stopping power**) of material with several parameters, such as the charge number of incident particle z (for cosmic muons, $z = 1$), the atomic mass and charge of absorber A and Z, the relativistic factors β and γ, the maximum possible energy transfer to an electron in a single collision $W_{\max}$ and the mean excitation energy I.

$$-\left\langle \frac{dE}{dx} \right\rangle = Kz^2 \frac{Z}{A} \frac{1}{\beta^2} \left[\frac{1}{2} \ln \left(\frac{2m_e c^2 \beta^2 \gamma^2 W_{\max}}{I^2} - \beta^2 - \frac{\delta(\beta\gamma)}{2} \right) \right] \tag{2.1}$$

This previous equation gives an accuracy of a few percent in the range $0.1 < \beta < 1000$ and we can easily see that the quantity of energy lost by in a muon when traversing any given medium actually depends on the energy of the incident muon, as shown in Figure 2.4. In practice, this means that

cosmic muons, having a mean energy of a few GeV, have mean energy loss rates actually close to the minimum: they are usually called for this reason *minimum ionizing particles* or MIPs. Energy losses of muons due to this ionization process are for this reason quite small and difficult to detect, so this interaction process will not be considered in this work.

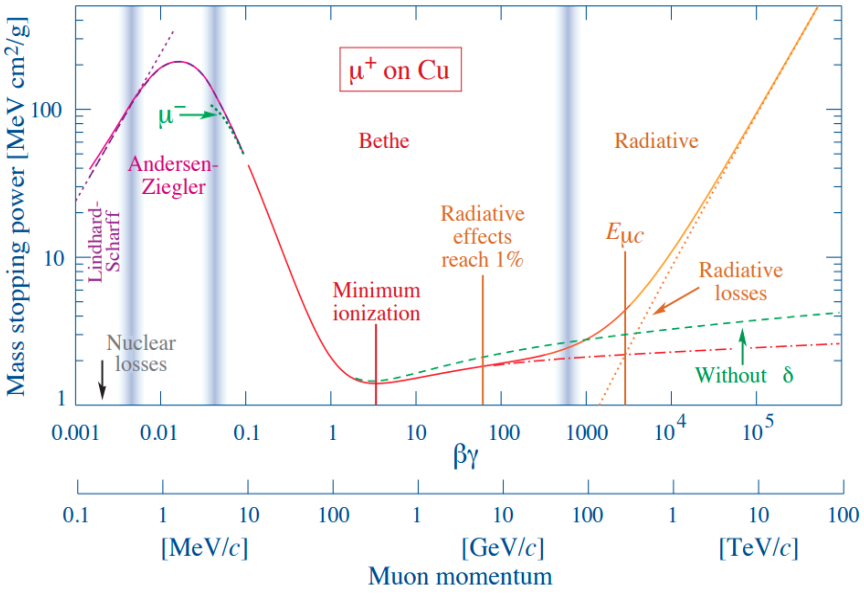


Figure 2.4: Mass stopping power of copper energy lost by a muon in copper due to the ionization process with respect to its momentum [9].

2.3.2 Multiple scattering process

Muons also interact with matter through another process, called **multiple scattering**. Since muons have a negative electric charge, by getting close to the nuclei of the absorber, they are suffering from Coulomb scattering. Given the high number of nuclei in matter, this process is repeated many times, deflecting each time the muon by a small angle in a stochastic way, meaning that there is no way of calculating this deviation exactly, but only using probabilities and the so-called theory of Molière [13].

According to this theory, the central 98% of the projected angular distribution due to Coulomb scattering can de described with a Gaussian function, whose width is given by the θ_0 parameter shown in Equation 2.2 [12], where p is the momentum of the incident particle and X_0 is the radiation length, defined as the characteristic amount of matter traversed by the incident particle for a particular interaction. The second term of the relation is actually typically neglected, being much smaller and therefore having a small impact on the final results obtained.

$$\theta_0 = \frac{13.6 \text{ MeV}}{\beta c p}\sqrt{\frac{x}{X_0}}\left[1 + 0.038 \ln\left(\frac{x}{X_0 \beta^2}\right)\right] \tag{2.2}$$

This formula is expected to be valid for distances up to $\sim 100 X_0$, giving an error smaller than 11%. Corrections do exist though in order to get slightly better results, especially in the tails of the distribution, but this theory is precise enough for our needs, given the small deviation angles we expect to observe given the experimental conditions considered.

Even though this deviation angle is the most important parameter when considering the multiple

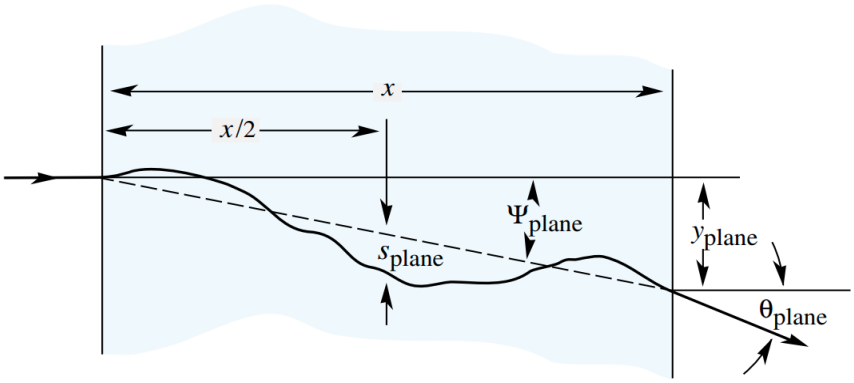


Figure 2.5: Schematic representation of the deviation induced by an absorber to an incident muon because of the multiple scattering effect [12].

scattering, other parameters can be important as well, as shown in Figure 2.5, such as the value of y_{plane}, corresponding to the deviation observed with respect to the initial expected position. All these additional parameters are typically highly correlated to the deviation angle previously defined but cannot be easily described with this simple theory and not relevant to this work anyway, since the deviation angle will be the main parameters studied with the muon tomography method used.

2.4 Muon tomography

Given Molière's theory, it is obvious to see that, instead of *calculating* the expected deviation of a muon crossing a given material, we can instead try and *measure* it, by inversing the two previous relations. Since this deviation depends on several parameters related to the absorber itself, such as its width and radiation length X_0, we can then infer such parameters experimentally and determine the properties of the medium crossed by cosmic muons. This is the so-called muon tomography, or **muography**, taking advantage of the interaction of muons with matter to use it in practice.

Muography is therefore a non-destructive imaging technique producing a density map of the inside of an object by measuring a flux of muons. Such method presents many different advantages over other imaging techniques such as X-rays, since it usually uses natural cosmic rays to make the measurements, being therefore completely safe. Additionally, muons interact lightly with matter, meaning that they typically have high penetrating capabilities and can therefore probe even large and/or dense objects. Muography can in this sense be applied to many different fields: it has for example even been used in 1970 in order to try and find hidden cavities of pyramids in Egypt [14] and can also be used in vulcanology, to determine whether a pocket inside of a volcano is empty or full of lava, among many other practical applications.

Such imaging techniques can be divided into two categories:

- **Absorption muography**. In this case, the observed muon flux in a given direction is compared to what is expected from cosmic rays, trying to determine the inner structure of the absorber as discrepancies between these two values. Only one detector is needed in this case, making this technique useful mostly to study large objects, even though the time to make a single measurement can last up to a few months.
- **Scattering muography**. On the other hand, the multiple scattering of muons can be used,

by placing one detector on each side of the object being studied to determine the deviation of the flux of incoming muons. The denser the material put in between, the larger the observed deviation will be, as shown in Equation 2.2. This technique is mostly used to study smaller objects and is able to make quick measurements.

In this particular work, scattering muography is being applied to industry in order to try and determine the degradation of the interior of industrial equipment, as we will now see.

2.5 Experimental setup

2.5.1 Muon detectors

If we want to work with cosmic muons, we need to be able to build some detectors able to spot them and give us back useful data, such as their energy and/or direction of propagation. Many different technologies exist nowadays in order to detect muons but in this particular case, the typical **multi-wire proportional chambers** have been used.

Invented at CERN in 1968, these detectors use an array of high-voltage wires (playing the role of the anode), running through a chamber filled with gas and whose walls are typically grounded (the cathode), as shown in Figure 2.6. Such an experimental setup therefore creates an electric field inside of the chamber, that needs to be made as large and uniform as possible.

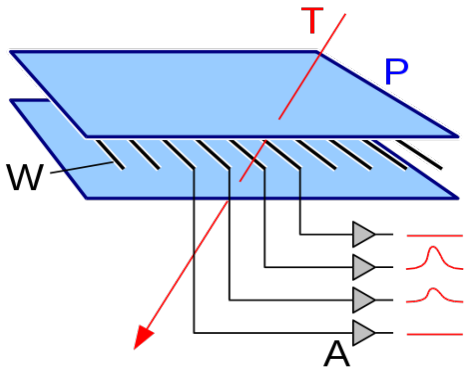


Figure 2.6: Schematic representation of a wire chamber muon detector.

When a charged particle such as a muon crosses this chamber, it will ionize the gas of the chamber and leave small electric charges along its path which will start to drift until reaching one of the wire. This drift induces an electric signal proportional to the ionisation effect in the different wires surrounding the particle path, and the combination of all the signals collected is able to give information regarding the actual path followed by the incoming particle.

Several properties are extremely important when designing a detector:

- First of all, the **spatial resolution**, the precision by which we can tell the position of the muon, should be ideally as small as possible, depending on the actual problem faced.
- The **efficiency** is also an important parameter, since we want to be able to detect as many muons as possible, to make the measurement faster and more precise.

- Finally, we typically want the detector to be large enough to avoid any **acceptance** issues.

2.5.2 Working setup

For this particular work, four muon chambers of 1 meter by 1 meter have been build using these principles, as shown in Figure 2.7. As we can see, more than 200 wires connected to the high voltage and made out of gold and tungsten have been placed every 4 mm in two different planes rotated by 90 degrees, to measure the x and y position of cosmic muons.

These chambers, filled while a mixture of Argon and CO_2, are then setup in pairs above and below the object being studied, in order to determine the position and the direction of the incoming and outgoing cosmic muon. This setup allows to measure with a good spatial resolution the deviation of muons, to determine as precisely as possible the properties of the object put in between both detectors.

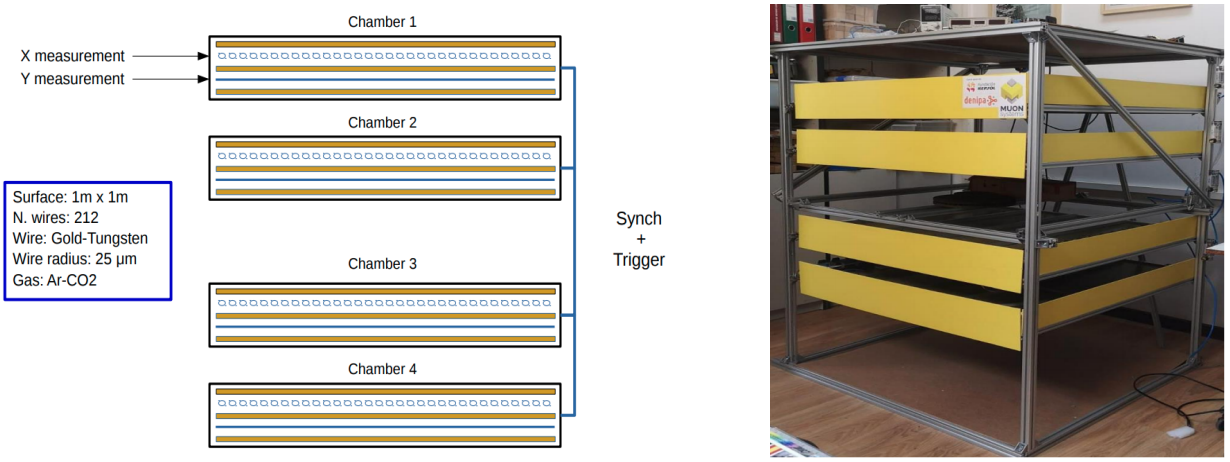


Figure 2.7: Actual working setup used for this work, with two multi-wires muon chambers of $1m^2$ placed on each side of the object being studied.

2.5.3 Data flow

Once collected, the data needs to be pass through several differents in order to convert electric signals into files that can be processed, as shown on Figure 2.8. As we can see, the data follows a different path depending on its nature, if it has been collected by the detector or if we are considering Monte-Carlo simulations that will be described in details in the next Chapter. The stream of data is collected by a simple USB, and then sent to a DAQ/DQM system and to an unpacker, which takes into account the detector gemetry and translates into a physical position the information received from the DAQ, typically only telling that a certain wire has been activated.

Both streams then join at the 1D histogram format level, gathering a collection of all the hits measured for every event. All these hits are then processed and reconstructed into two trajectories, one above and one below the object, using advanced techniques that will also be described in next Chapter. Finally, 4D segments are made available from this reconstruction process, having the position and direction information needed, from which the actual analysis can take place.

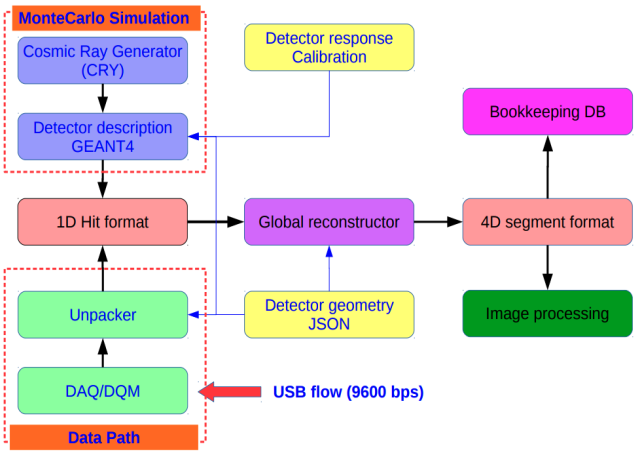


Figure 2.8: Schematic representation of the detector data flow.

Chapter 3

Statistical parameters

The general objective of this work is to find a way to determine the geometry of an eventual object placed between the two muon detectors by measuring the deviation affecting incident cosmic muons, a typical non-deterministic process which therefore needs to be adressed using advanced statistical parameters ans methods, such as probability and kernal density functions (Sections 3.1 and 3.4), Monte-Carlo simulations (Section 3.2), p-values (Section 3.3) and likelihood function (Section 3.5).

3.1 Probability density functions

The probability density functions, or PDFs, are statistical expressions defining probability distributions representing the likelihood of any given outcome. They are typically represented as curves in the two dimentional space (x, y), as shown in Figure 3.1, in which the total area below the curve in an interval is equal to the probability of a discrete random variable occurring.

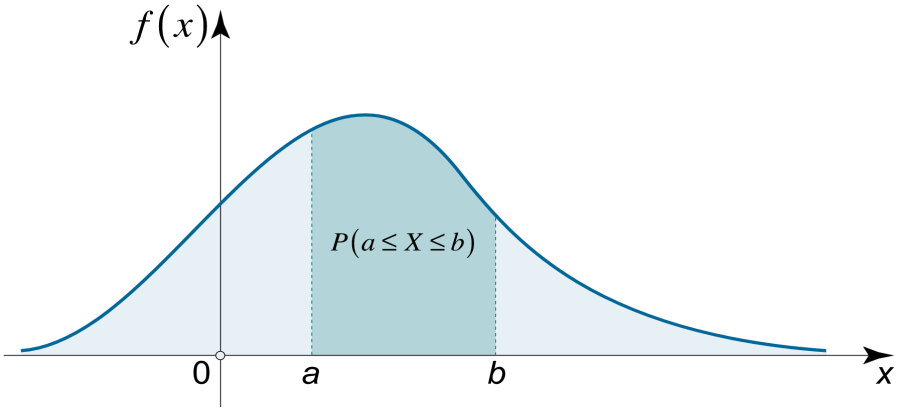


Figure 3.1: Schematic representation of a random PDF [15].

Gaussian distributions are in this sense probability density functions having a particular expression defined in Equation 3.1, where σ is the so-called **standard deviation** of the distribution, representing its width: a low σ indicates that the values of the distribution tend to be close to its mean value μ, while a high σ indicates that the values are spread out over a wider range.

$$f(x) = \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2} \tag{3.1}$$

These definitions are important because we already know that the multiple scattering process which affects the incident cosmic muons is a stochastic process: this means that two muons having similar kinematics can leave the detector with very different output directions and positions, whose actual distribution can be approximated by a gaussian function for small deviation angles (at larger angles, the distribution is behaving like Rutherford scattering, with slightly larger tails).

The standard deviation of the distribution of such output paraemeters is actually highly dependant on the geometrical parameters θ_i of the eventual object put between the two detectors, as we saw in Equation 2.2. Indeed, the mean deviation observed is expected to depend on the number of radiation lengths crossed, typically much higher when a muon crosses denser objects, such as a steel pipe, as shown in Figure 3.2. This means that in first approximation, the tichker the object investigated is, the higher the expected deviation will be and this simple observation will actually be used as the driving process of the statistical study performed in this work.

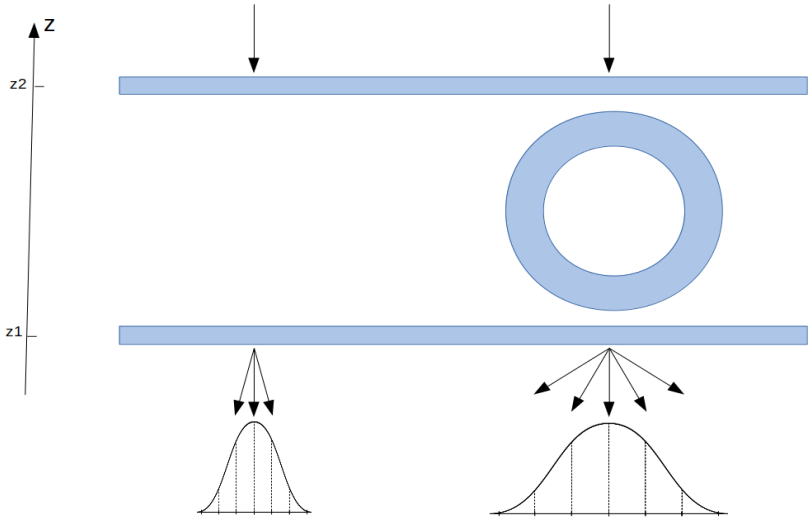


Figure 3.2: Schematic representation of the deviation expected for an incoming muon and PDFs associated without (on the left) and with (on the right) an object placed between the detectors.

3.2 Monte-Carlo simulations

Monte-Carlo simulations are algorithms developped in order to compute approximate numerical values to stochastical problems using random processes and advanced probabilistic techniques. In this particular case, such computation methods apply extremely well since, by being a stochastical process, the actual probability density function associated to a given geometry cannot be computed analytically (and if we were able to do it, a simple gradiant descent method would be enough to solve this problem and find the optimal parameters θ_i of the object).

The only way available to actually estimate these parameters is by using thousands of Monte-Carlo simulated toys, simulating thousands of different incident muons for different geometries. The principle is simple: for each incident muon measured, a large number N of Monte-Carlo simulations will be performed, propagating the muon using an algorithm described in Chapter 4 until reaching the lower detector. Repeating this experiment over and over again always gives different results, giving us the opportunity to build the expected PDF for a given object geometry and for the incident muon that has been measured and that will be used later on.

At the end of the day, the main objective of this process is to be able to estimate the probability

of observing a certain deviation given an object geometry and the input trajectory of the muon, $P(\text{deviation}|\text{input})$ in simulation. Then, this process will be reversed using actual data in order to try and obtain the geometry of the object from the actual measurement of the input and output muon trajectories and positions, using a likelihood minimization method.

3.3 p-values

So far, we developed a technique allowing us to reconstruct the expected probability density function for a given incident muon using Monte-Carlo techniques. However, we still need a way to estimate the goodness of the actual measurement with this simulated PDF: this is where the so-called **p-values** enter, a key ingredient for this work since they allow to relate the two main parts of the problem: the simulations performed and the actual data collected.

Probability values, or p-values, are based on the concept of **null hypothesis** H_0, a general statement or default position telling that there is no actual relationship between two measured phenomena and assumed to be true until an evidence indicates otherwise. This hypothesis is then typically opposed to the **alternative hypothesis** H, usually more interesting by being the interesting hypothesis of the experiment being performed. The main goal is to compare the data with both hypotheses: if the data is consistent with H_0, then the null hypothesis can simply not be rejected. However, we can reject H_0 (and therefore accept H, its exact opposite) if the data collected is significantly unlikely to have occurred if the null hypothesis were to be true, according to a **confidence level** previously defined.

From these definitions, a p-value is then defined as the probability for a variable to be observed equal or more extreme (lower or higher, depending on the cases) than the actual value x observed under the null hypothesis. This definition, can be represented in a schematic way in Figure 3.3.

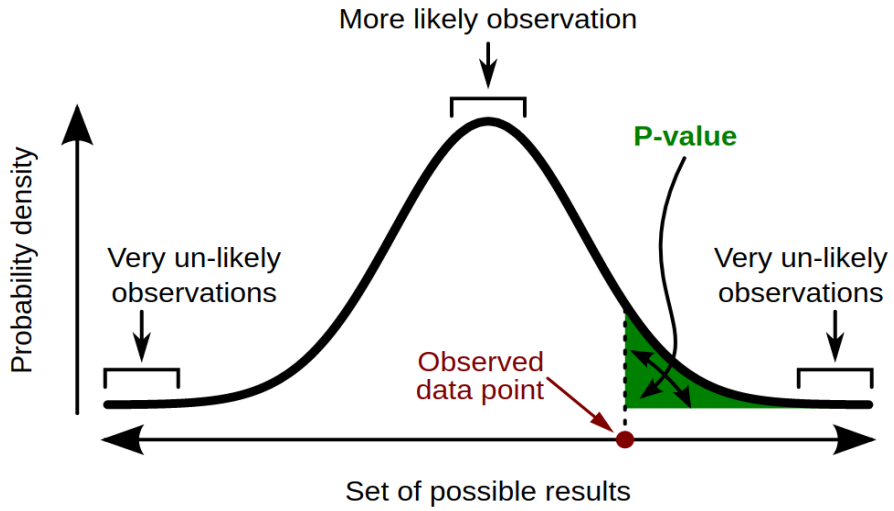


Figure 3.3: Schematic representation of the statistical concept of p-value.

In general, a smaller the p-value tells us that the hypothesis under consideration may not adequately explain the observation, leading to a higher statistical significance. At the end of the day, the null hypothesis is rejected if the p-value obtained is smaller than a previously defined, arbitrary and fixed threshold α, the **level of significance** of the experiment. This parameter can take a large range of values, typically ranging from 0.05 to 0.001.

In this work, the p-value will then allow us to compare a single data measurement with the probability

density functions obtained from Monte-Carlo and quantify the probability that such muon was actually observed for a given object geometry. This value will then be feeded to a likelihood function, as explained in Section 3.5.

3.4 Kernel density estimations

In statistics, the kernel density estimation is a method allowing to estimate the shape of the density probability f, the PDF defined in Section 3.1, of a random variable $\boldsymbol{x}$, from N observations drawn from this unknown function PDF. The usual way to proceed is to place each observation made in an histogram, where the density in each point x can be therefore estimated as the proportion of observations close to x. However, this method depends strongly on the binning used and the function obtained by this process is non-continuous by definition. The **kernel method** has been developped in this context, to try and solve the non-continuity of the PDF obtained, by simply using various continuous functions instead of bins.

Mathematically, we can define a function called the **density estimator** of f, $\hat{f}_h(x)$, defined as a sum of such general functions K over the number of observations drawn, as shown in Equation 3.2.

$$\hat{f}_h(x) = \frac{1}{Nh} \sum_{i=1}^{N} K\left(\frac{x - x_i}{h}\right) \tag{3.2}$$

In this last equation, two parameters that need to be defined by hand before using this method are actually extremely important:

- First, the **kernel** K, is a non-negative window function defined depending on the problem, as the "building block" of the final function that needs to be estimated: indeed, as we can see, the function $\hat{f}_h(x)$ will be given as a simple sum of these kernels. In most of the problems, including this work, a standard gaussian kernel is used, taking the particular shape shown in Equation 3.3.

$$K(x) = \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}x^2} \tag{3.3}$$

- The **bandwidth** h is another interesting parameter, affecting the final smoothness of the resulting curve. The optimal selection of such parameter is not obvious at all [16] but, as a rule of thumb, it can be shown that for gaussian kernels of standard deviation σ, it should be approximately equal to $1.06\ \sigma N^{-1/5}$ in order to minimize the mean integrated square error obtained.

A comparison of the different methods so far and different possible is shown in Figure 3.4.

In summary, this method is used in order to model the distribution of an arbitrary input dataset as a superposition of Gaussian kernels, one for each data point, each contributing 1/N to the total integral of the PDF. This method allows us to construct continous functions and recover for the eventual low statistics and possibles holes in the bi-dimentional histograms for the position and direction of the muons obtained in x and y, following the method that will be described in Chapter 4.

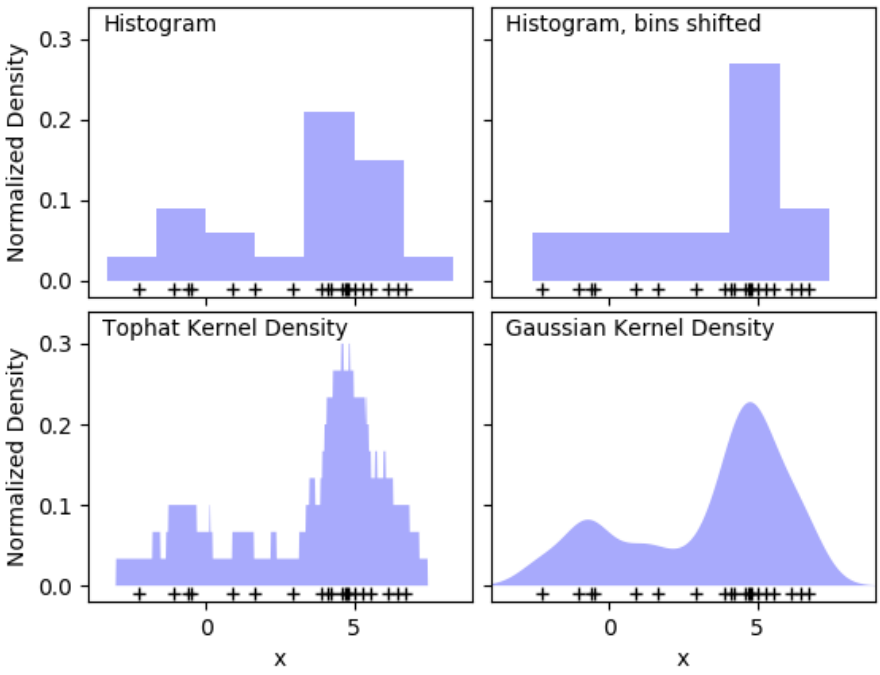


Figure 3.4: Comparison of the histogram and the kernel density estimations methods for different kernels in order to estimate a PDF [17].

3.5 Likelihood

So far, we have been able to generate thousands of different Monte-Carlo experiments for a given muon and for a given object geometry, but we still need to find a way to reverse this process to estimate the geometry of the object from a given measurement.

This is where the likelihood becomes useful, defined as a function that measures the goodness of a fit with respect to a sample of data, for several unknown parameters of a mathematical model. In general, this function can be defined in Equation 3.4, where $\boldsymbol{\theta} = \{\theta_1, ..., \theta_i\}$ are the parameters of the model and x is the actual measurement of a random variable X following a density function f.

$$\mathcal{L}(\boldsymbol{\theta}|x) = f_{\boldsymbol{\theta}}(x) = P(X = x|\boldsymbol{\theta}) \tag{3.4}$$

It is important to note at this point that the likelihood is a function of the parameters $\boldsymbol{\theta}$ but not a probability density function itself. Additionally, it should in general not be confused with the probability $p(\boldsymbol{\theta}|x)$ since it is equal to the probability of observing a given outcome x when the true values of the parameters are $\boldsymbol{\theta}$: this means that the likelihood is equal to a probability density over the outcome x, not over the parameters $\boldsymbol{\theta}$.

To understand this concept better, the simple example of an unfair coin toss can be considered, where the fairness of the coin is the parameter of the model, represented by the probability of obtaining head p_H and taking values between 0 and 1 (for a regular coin, $p_H = 0.5$). The likelihood is then defined

from a given observation, such as obtaining two heads in a row HHT in the following way:

$$\mathcal{L}(p_H|HHT) = P(HHT|p_H) = P(H|p_H) \cdot P(H|p_H) \cdot P(T|(1-p_H)) \tag{3.5}$$

For each value of p_H, the value of the likelihood can then be computed and ultimately plotted, as shown in Figure 3.5, showing a minimum value for a particular value of p_H.

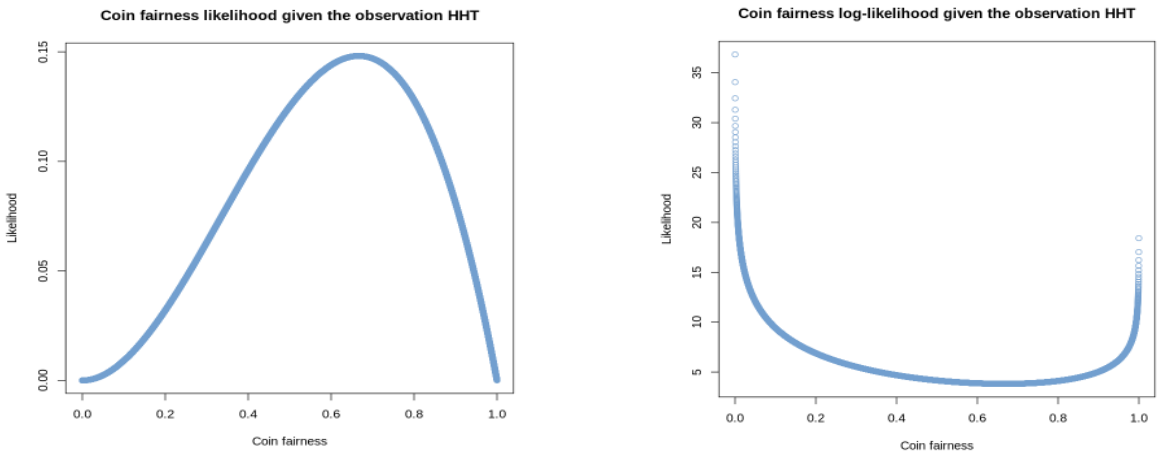


Figure 3.5: Likelihood (on the left) and log-likelihood (on the right) obtained for our particular example of the fairness of a coin given the observation HHT.

The likelihood is in this sense interesting because it can be described as a hypersurface whose peak gives the optimal set of parameters maximizing the probability of drawing the actual sample measured. Often, the **log-likelihood** $l(\boldsymbol{\theta}|x) = -2\log(\mathcal{L}(\boldsymbol{\theta}|x))$ is actually used instead, being a bit more convenient to deal with the special part concavity plays in the maximization process. Given the properties of the logarithm, maximizing the likelihood is equivalent to minimizing this particular log-likelihood.

Another important property of the likelihood that will be used extensively and already assumed in the previous example is the fact that, for two independant measurements, the likelihood of both measurements is equal to the product of both likelihoods independantly computed for each event.

In this work, a likelihood function will be obtained as the product of the thousands of muons collected, but for a fixed geometry of the object. A minimization method will then be performed in order to try and minimize the value obtained by trying out different parameters $\boldsymbol{\theta}$, until finding the optimal set of parameters able to describe best the object put in between the detectors.

Chapter 4

The algorithm

The algorithm we developped to solve this particular problem has been written in C++ and can be divided in two categories. First of all, we have the so-called **PipeReconstructor** (Section 4.1), a set of classes able to perform several tasks:

- Define the geometry of the problem (the detector, and the volume under investigation are therefore defined using the smallest possible set of parameters at this stage);
- Compute the different intersection points of an incident cosmic muons with our geometry;
- Propagate the muons through the different medium of interaction encountered along their path;
- Calculate the likelihood of a given measurement for a given volume;
- Finally, probe different volumes and perform a descent method in order to minimize the likelihood encountered and find the optimal volume to solve our problem given the measurements collected.

This class is complemented with the **Generator** (Section 4.2), a small class allowing us to generate muons simulating our actual experiment, using some of the functions of the PipeReconstructor.

4.1 PipeReconstructor

The PipeReconstructor is a set of classes, allowing us to define different kind of objects:

- First of all, a muon is defined from the **MuonState** class and from 7 parameters: 3 parameters (x, y, z) representing the position of the muon, 3 additional parameters (v_x, v_y, v_z) representing its three-dimensional direction and the actual value of its momentum. The measurement performed at the top detector will be labelled with the index 1 (x_1, y_1, z_1, v_{x1}, v_{y1}, v_{z1}, p_1) while the measurement performed by the lower detector will take the index 2.
- **Surfaces** and **Volumes**, virtual classes allowing us to define general surfaces (their spatial position and their geometrical center) and volumes, defined as vectors of several surfaces: a general volume is in this sense defined by its surfaces, central position and global density. Defining such virtual classes is extremely important since it would allow us to consider different geometries, not just a pipe put in between the two detectors. The constant size ($1m^2$) and z position of both the lower and upper detectors (-37 and 37 cm, respectively) is also defined at this stage.

- **Cylinders** and **Pipes**, particular subclasses of the virtual classes previously defined in order to define the geometry of this particular problem. A pipe is in this sense defined as a set of two cylinders and from 7 different parameters: its central position (x, y, z), along with its density, the radii r and R of both cylinders and its length along the y-axis L, as shown in Figure 4.1. For simplicity, the pipes considered in this work are therefore considered to have a constant density.

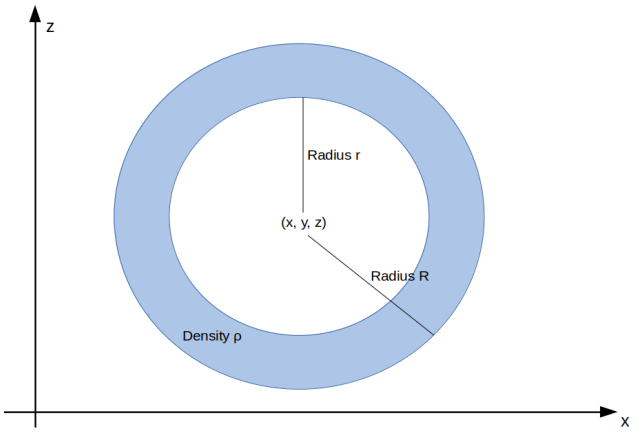


Figure 4.1: Schematic representation of the pipe considered in this work and its respective parameters.

One method computing the exact cut point between a muonState and a Cylinder has been defined at this point as well, from the general equation of a cylinder of radius R centered in (X, $Y = 0$, Z) (given that the y-axis is defined along the cylinder, we can assume that the cylinder is actually centered in $Y = 0$ without any loss of generalization).

$$(x - X)^2 + (z - Z)^2 = R^2 \tag{4.1}$$

From this, computing the intersection between such cylinder and a general MuonState (x_0, y_0, z_0, v_x, v_y, v_z, p) is trivial, since a MuonState can be represented by a straight line:

$$\begin{cases} x = x_0 + \lambda v_x \\ y = y_0 + \lambda v_y \\ z = z_0 + \lambda v_z \end{cases} \tag{4.2}$$

By putting this into the equation of the cylinder, we get the following expressions:

$$(x - X)^2 + \left(z_0 + \frac{v_z}{v_x}(x - x_0) - Z \right)^2 = R^2 \tag{4.3}$$

Once both the squared values applied, we get an equation having the shape $Ax^2 + Bx + C = 0$, where both cutting points will be given therefore by $x^{+/-} = \frac{-B \pm \sqrt{B^2 - 4AC}}{2A}$:

$$\begin{cases} A = \left(1 + \left(\frac{v_z}{v_x} \right)^2 \right) \\ B = -2X - 2x_0 \left(\frac{v_z}{v_x} \right)^2 + (2z_0 - 2) \left(\frac{v_z}{v_x} \right) \\ C = X^2 + Z^2 + z_0^2 - 2z_0 Z + x_0^2 \left(\frac{v_z}{v_x} \right)^2 + x_0(2Z - 2z_0) \left(\frac{v_z}{v_x} \right) - R^2 \end{cases} \tag{4.4}$$

Finally, methods allowing us to determine whether a muon state is inside or outside of the pipe have also been written, taking into account the particular geometry of such objects in order to determine the medium in which the muon is propagating and therefore use the correct density value in all the calculations performed.

- A **Propagator** is then defined as the object allowing us to propagate a MuonState through a general Volume. The way it works is quite simple:
  - First of all, the distance between the initial MuonState and all the surfaces of the object considered is computed and the first intersection point is kept.
  - The muon is then propagated in a distance corresponding to 90% of the total distance to this first cut point by taking into account the multiple scattering effect and the Molière theory, and a new MuonState is therefore obtained. We cannot propagate the muon 100% of the distance immediately to avoid effects related to the actual shape of the Cylinder and due to the fact that the Equation 2.2 is only valid in case of normal incidence between the MuonState and the Cylinder, which is typically not verified, as shown in Figure 4.2.

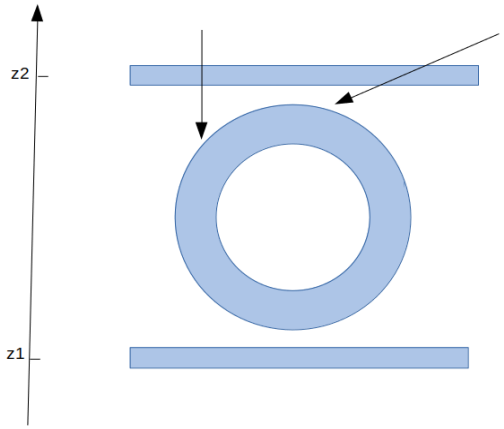


Figure 4.2: Typical example of two incident muons not respecting the normal incidence required by the Molière equation for the multiple scattering of muons.

  - This propagation process is then several times, until reaching a position extremely close to the volume that is being studied (our tolerance parameter has been fixed to 0.1 mm). The muon is then manually propagated along its direction of propagation along a distance of 0.1 mm in order to force its crossing with respect to the first surface, so that the propagation can be repeated for the next surface.
  - This whole proces is repeated for all the different surfaces encountered by the muon (up to four times in total for our particular geometry), and one final time until reaching the bottom detector. At this point, the MuonState is kept as the second measurement (x_2, y_2, z_2, v_{x2}, v_{y2}, v_{z2}, p_2) that will be compared with the simulation using statistical methods.
  - Eventual muons out of the acceptance of any of the detectors are of course rejected at this stage and simply not considered when computing the value of the likelihood.

- The **Likelihood** object itself can then be calculated. This class takes as input a pipe geometry and a file containing a set of measurements (either obtained from Monte-Carlo or actually measured by the detectors) and its main objective is to estimate how likely it is that we obtained exactly these measurements for the geometry given. In order to reach this goal, several calculations are performed for each event found in the input file ($n_{\text{gen}} \sim 100.000$):

– First of all, three MuonStates are computed: the so-called *IncomingMuon* and *Outgoing-Muon*, directly read as the measurement at the upper and lower detectors from the input file, and the *IncomingMuonLinearDown*, defined as the prolongation along a straight line of the *IncomingMuon*. Each *IncomingMuon* is then propagated along the geometry using our Propagator, defining a fourth *IncomingMuonDown* MuonState, as shown in Figure 4.3.

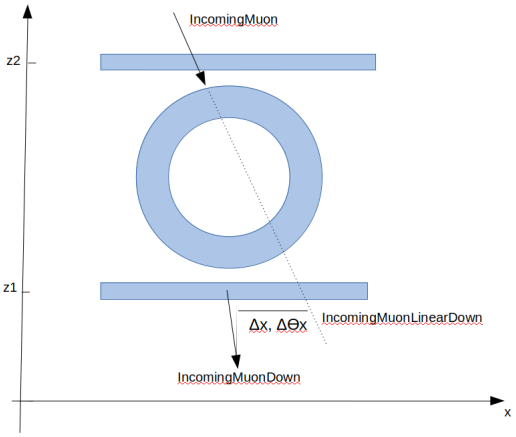


Figure 4.3: Schematical representation of the 4 four muonStates and main parameters used for the computation of the likelihood.

– From these muonStates, the Δx, $\Delta\theta_x$, Δy and $\Delta\theta_y$ parameters can easily be computed as the difference between the position and angle measured between the *IncomingMuonDown* and *IncomingMuonLinearDown* muonStates. Such parameters are important and will be used a lot in this work since they represent the actual deviation suffered by the muon.

– This propagation process is repeated over and over again ($n_{\text{prop}} \sim 100$ times in total) since we known that the multiple scattering is a stochastic process. Each time the propagation method is called for the same incident muon, different deviations parameters are therefore obtained, each kept in a vector, which can be represented in an histogram and smoothen using the RooFit package and kernel density functions, as explained in Section 3.4.

– At this stage, all the ingredients needed to compute a p-value are available. A p-value for each *IncomingMuon* is then computed by using the vectors previously defined and representing an histogram of all the possible deviation values, and by comparing the actual measurement *OutgoingMuon* for such distributions.

– Two p-values, one along each of the x and y axes, are then obtained for each of the N events found in the input file. The total value returned by this function can then simply be computed from Equation 4.5.

$$\mathcal{L} = \sum_{i=1}^{N} -2.0 \cdot \log(\text{pvalue}_{x,i}) \cdot \log(\text{pvalue}_{y,i}) \tag{4.5}$$

- So far, we described a method allowing to compute the likelihood of observing a given measuremen by using Monte-Carlo simulations for a given volume, but one last important step is still missing: we need to find a way to find the optimal volume for a given mesurement, the main objective of this process. Talk about the minimization process once implemented.

4.2 Generator

As we have just seen, the computation of the p-values required for the likelihood takes as input a file of Monte-Carlo generated events, which has been previously produced thanks to this **Generator**. This class takes as input one rootfile containing the Monte-Carlo generation of incoming cosmic muons previously produced by the company using CRY [18], a dedicated generator for cosmic muons. To simplify the current problem, the goal is then to simply read the input parameters of such muons, such as the position of their impact with the upper detector, and to propagate them using the previously defined functions, to generate a distribution of the expected output.

The way this algorithm works is quite simple: a general volume matching the one that needs to be studied (in this case, a general Pipe located in the center (0, 0, 0) of the detector, having an inner radius variable, an outter radius of 20 cm and a length of 50 cm) is first of all defined. Then, for each event we want to generate ($n_{\text{gen}} \sim 100.000$), a loop is performed: in each iteration of this loop, the incident muon is propagated throughout the geometry, histograms for the parameters Δx, $\Delta\theta_x$, Δy and $\Delta\theta_y$ are obtained and, more importantly, measurements in the lower detector are artificially generated to test our framework without having to consider detector effects for now.

One last important notion to introduce at this point is the difference between the actual measurement and the Monte-Carlo truth value, that can both be computed by the generator. The existence of both variables is related to the fact that the detector is made out of 200 discrete wires, each separated by a distance of 4 mm, meaning that the actual position measured is typically a discrete number. However, when considering simulation files, both the real and the measured position can be obtained and kept in different MuonStates labelled as (px_i, py_i, pz_i, pv_{xi}, pv_{yi}, pv_{zi}, p_i) and (mx_i, my_i, mz_i, mv_{xi}, mv_{yi}, mv_{zi}, m_i) respectively, even though the second MuonState will be used by default for the rest of this work, allowing for a direct comparison between simulation and data.

Add plot to show the difference?

Chapter 5

Results obtained

Now that the experimental setup and main statistical parameters used have been fully described and that the algorithm developped for this particular exercise has been introduced, we can present the results obtained. This chapter is divided into several sections: first of all, we are going to mention the different validation steps performed to make sure that this algorithm's results can be trusted in Section 5.1, and several properties of cosmic muons will be checked in Section 5.2. Then, more general results will be presented in Section 5.3.

5.1 Generator validation

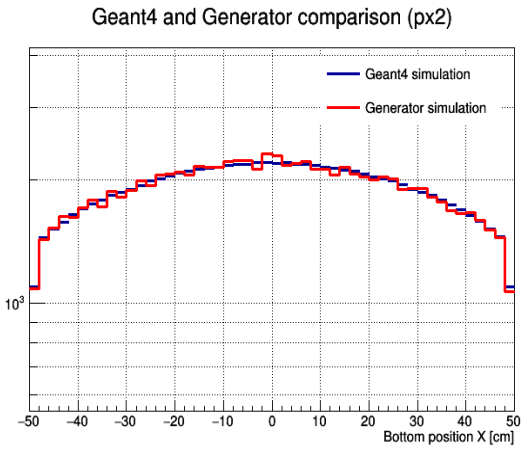
The validation step consists in comparing the results previously obtained with Geant 4, a a toolkit for the simulation of the passage of particles through matter [19], with a complete and realistic description of the interaction between the cosmic muons and the detectors, and the ones obtained with our Generator, assuming a perfect detector.

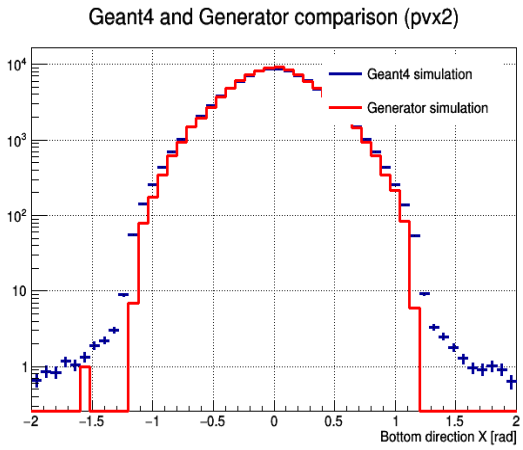
This step is actually extremely important, as it allowed us to find several bugs in the code. In this context, several files have been generated using both generators and different pipe geometries, allowing us to perform several different checks. In each case, the objective was to obtain the output MuonState (x_2, y_2, z_2, v_{x2}, v_{y2}, v_{z2}, p_2) after propagating the initial MuonState throughout a given geometry.

For a given pipe having a geometry (r = 17.2 cm, R = 20 cm, L = 50 cm) and located in (0, 0, 0), thousands of muons have then been generated using Geant 4 and our Generator. The output distributions (the positions x and y and directions v_x and v_y after propagation, along with the deviations in position ΔX and ΔY, defined in Equation**??** where d is the vertical distance between the two detectors, and the deviation angles $\Delta\theta_x$ and $\Delta\theta_y$, defined in Equation **??**) have then been obtained in this case in Figure 5.1. All these plots have been normalized to unity in order to account for the eventual different number of events generated in both cases.

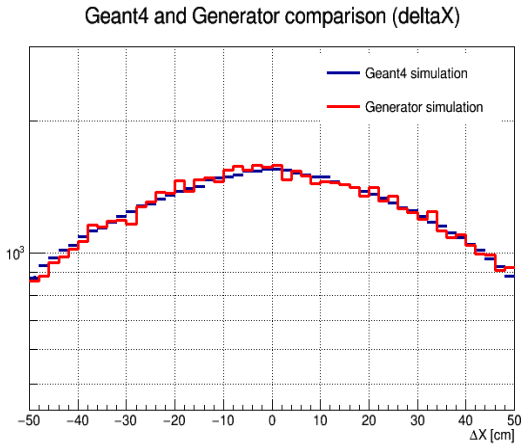
5.2 Cosmis muons properties

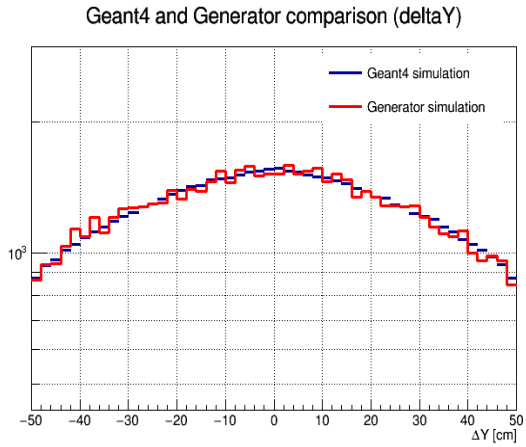
One of the most famous properties of cosmic muons is their regular angular distribution T_θ, given by Equation 5.1, where I_0 and n are parameters depeding on several factors, such as the altitude and



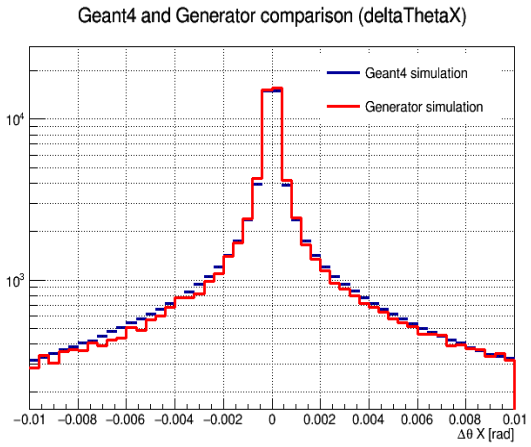


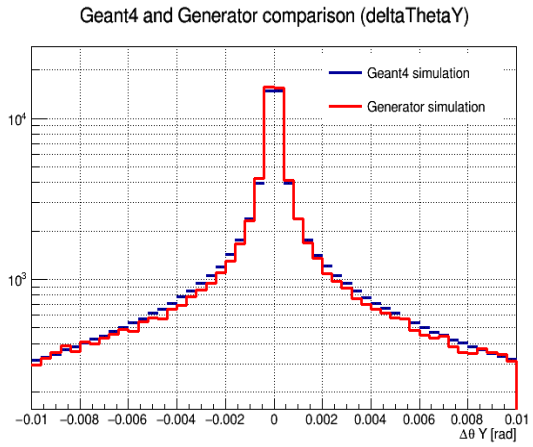
(a) Muon position (on the left) and direction (on the right) along the x-axis at the bottom detector after propagation





(b) Muon ΔX (on the left) and ΔY (on the right) as measured after propagation





(c) Muon $\Delta\theta_x$ (on the left) and $\Delta\theta_y$ (on the right) as measured after propagation

Figure 5.1: Normalized variables measured on the bottom detector using the Geant4 (in blue) simulation process and our custom Generator (in red).

latitude of the place where the measurement is performed.

$$I_\theta = I_0 \cos^n \theta \propto cos^n \theta \tag{5.1}$$

At our latitude, we can approximately estimate that $n \simeq 2$ and this is something we can easily check with our experimental setup, given the fact that the upper detector is able to give us such angular distribution. Another important property of cosmic muons is their relatively well known energy spectrum, that we can also represent here, in order to check for the validity of the cosmic muons generated with CRY. The results obtained in both cases are shown in Figure 5.2.

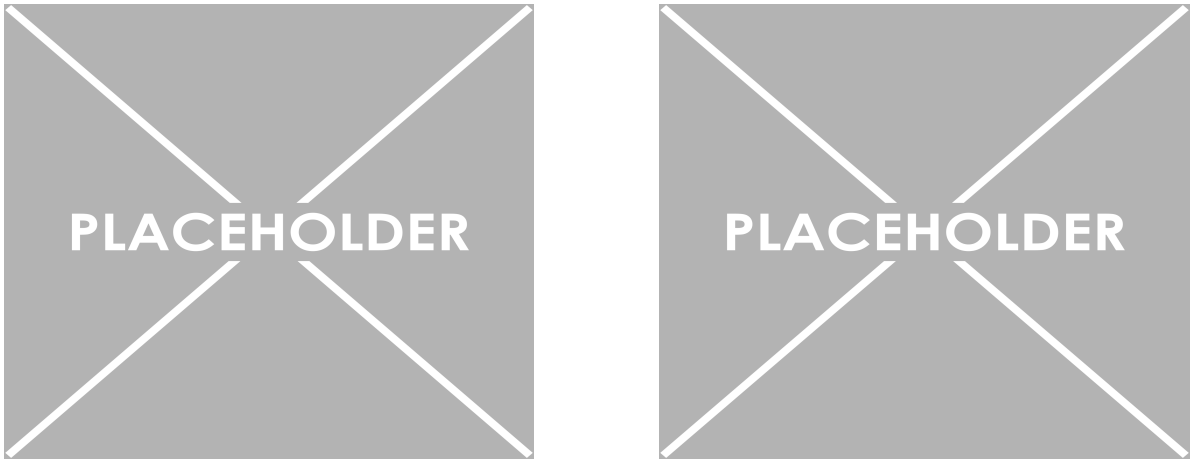


Figure 5.2: Study of the angular (on the left) and energy (on the right) distributions of cosmic muons generated by CRY and used throughout this work.

5.3 General results

5.3.1 Pipe geometries impact

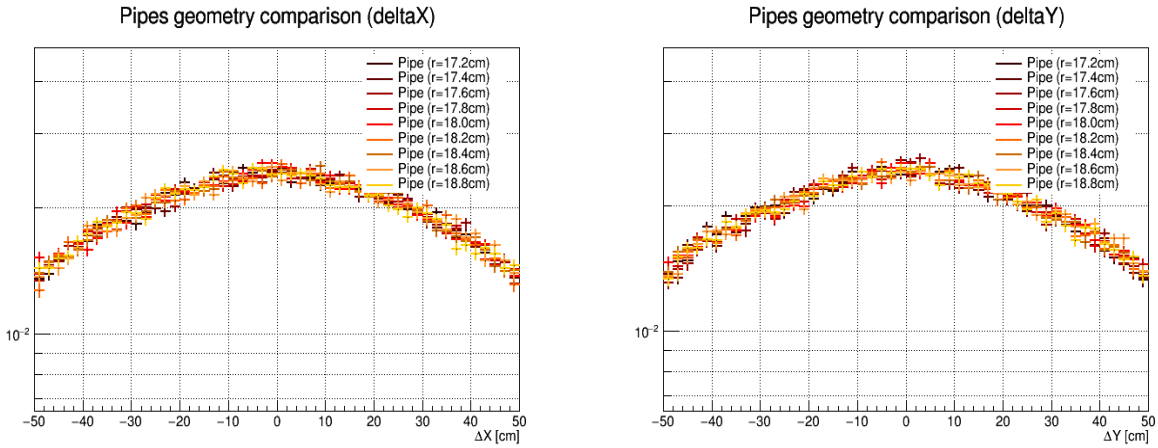
Now that we know we can trust the results obtained from the Generator and that general properties of cosmic muons have been derived, we can start comparing the ΔX, ΔY, $\Delta\theta_x$ and $\Delta\theta_y$ distributions obtained for different geometries. This is done in Figure 5.3, considering pipes with a constant length $L = 50$ cm and outer radius of $R = 20$ cm, while the inner radius changes from $r = 17.2$ to 18.8 cm.

As we can see, in this case, and as expected, the geometry has a little impact on the deviation observed in the position and direction, along both the x and y axes. We can however observe slightly larger tails for the pipes having a larger width, as expected. The study of this small differences between the geometries will be the starting point of the determination of the properties of an unknown pipe placed between the detectors.

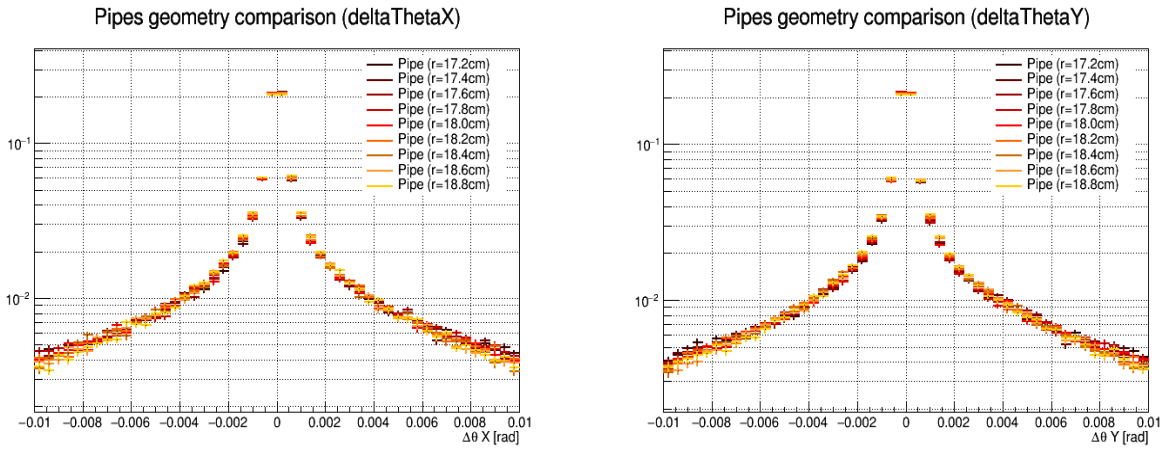
5.3.2 Kernel density functions

5.3.3 Likelihoods

The goal is now to estimate the value of the total likelihood obtained for different pipe geometries, characterized by different inner radii, ranging from 17.2 to 18.8cm, by steps of 0.2cm. The first



(a) Muon ΔX (on the left) and ΔY (on the right) as measured after propagation



(b) Muon $\Delta\theta_x$ (on the left) and $\Delta\theta_y$ (on the right) as measured after propagation

Figure 5.3: Normalized deviation variables generated using different pipe geometries.

step has then been to generate 9 Monte-Carlo files with 100.000 events each, using our Generator, corresponding to these 9 possible geometries.

The idea is then to compute the value of the likelihood obtained for each geometry by comparing an actual measurement obtained by our propagator with these generated files, in order to estimate the p-value in each case and, at the end of the day, try to figure out which pipe geometry is more likely to give rise to such measurements. The results obtained are shown in Figure 5.4, where we expect to see a minimum for the actual geometry used, represented by the vertical red line in each case.

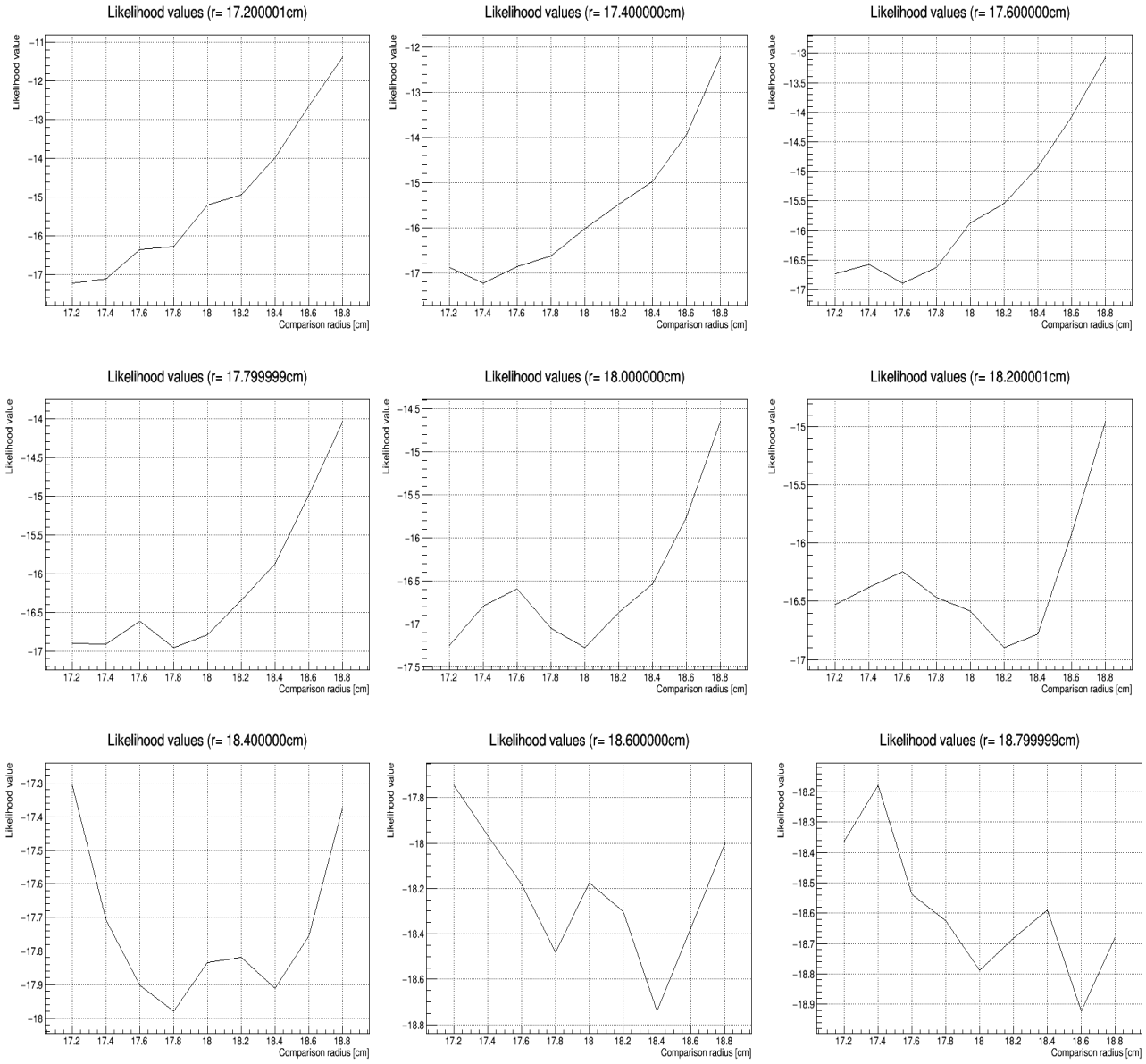


Figure 5.4: Likelihoods values obtained for different pipe geometries, ranging from 17.2 (on the top left) to 18.8cm of inner radius (on the bottom right).

5.3.4 Minimization process

Chapter 6

Conclusions

Improvements: consider ionization as well, detectors effect, additional parameters

Bibliography

[1] G. Altarelli, "The Standard Model of Particle Physics", CERN-PH-TH/2005-206, 2005

[2] D0 Collaboration, "Observation of the Top Quark", Phys.Rev.Lett.74:2632-2637, 1995

[3] DONUT Collaboration, "Observation of Tau Neutrino Interactions", Phys.Lett.B504:218-224, 2001

[4] S. Chatrchyan et al., "Observation of a new boson at a mass of 125 GeV with the CMS experiment at the LHC", Phys.Lett.B716:30-61, 2012 [arXiv: 1207.7235]

[5] G. Aad et al., "Observation of a new particle in the search for the Standard Model Higgs boson with the ATLAS detector at the LHC", Phys.Lett.B716:1-29, 2012 [arXiv: 1207.7214]

[6] CMS Collaboration, "The CMS Experiment at the CERN LHC", JINST 3 S08004, 2008

[7] ATLAS Collaboration, "The ATLAS Experiment at the CERN Large Hadron Collider", JINST 3 S08003, 2008

[8] S. Manzoni, "The Standard Model and the Higgs Boson", Physics with Photons Using the ATLAS Run 2 Data, Springer Theses, 2019

[9] "Muon", Particle Data Group, 2018

[10] "Cosmic rays", Particle Data Group, 2017

[11] S.H. Neddermeyer and C.D. Anderson, "Note on the Nature of Cosmic-Ray Particles", Physical Review Vol. 51, 1936

[12] "Passage of Particles Through Matter", Particle Data Group, 2019

[13] H.A. Bethe, "Molière's Theory of Multiple Scattering", Physical Review Vol. 89, 1953

[14] L. Alvarez et all., "Search for Hidden Chambers in the Pyramid", Science, Volume 167, Issue 3919, 1970

[15] Math24, "Probability Density Function", as seen in May 2020

[16] S. Chen, "Optimal Bandwidth Selection for Kernel Density Functionals Estimation", Journal of Probability and Statitics vol. 2015 art. 242683, as seen in June 2020

[17] Scikit-learn: Machine Learning in Python, "Simple 1D Kernel Density Estimation", as seen in June 2020

[18] M. Holmann et all, "GEANT4 Simulation of a Cosmic Ray Muon Tomography System with Micro-Pattern Gas Detectors for the Detection of High-Z Materials", as seen in June 2020

[19] S. Agostinelli et all, "Geant4?a simulation toolkit", Nuclear Instruments and Methods in Physics Research Section A, vol. 506, issue 3, pages 250-303