Development of a statistical analysis framework in the context of Muography applied to the industry

Cédric Prieëls

Director - Pablo Martínez Ruíz del Árbol
Co-director - Carlos Díez

Universidad de Cantabria
Muon systems

July 17th 2020

Table of contents

Section I
General introduction

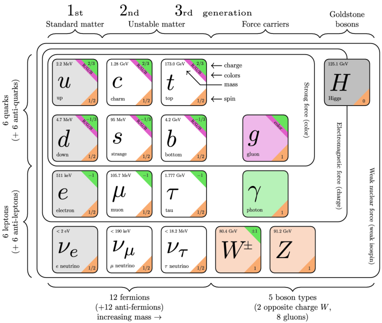
This work is framed in the context of Muography, an emerging non-destructive testing (NDT), being used in the industry to perform preventive maintenance of industrial equipment.

The project aims at developing a framework to perform inspection of the inner properties of objects using advanced statistical concepts such as the maximum likelihood estimation method.

Particle physics and muons

The Standard Model **describes the fundamental particles** existing and their interactions:

- Introduced in the 1970s and still considered to be valid, but probably incomplete
- Simple in concept but extremely precise
- Lots of successful predictions made over the years, such as the existence of the top quark and the Higgs boson



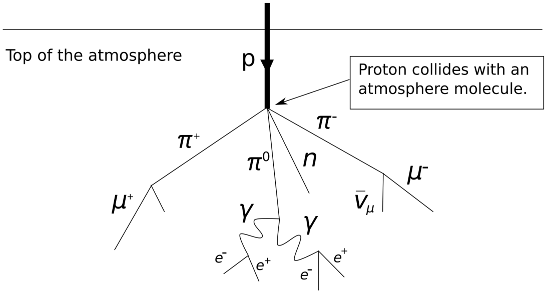
Muons

- Muons μ^- are one of the 12 fundamental particles existing
- They have a relatively small interaction cross-section with ordinary matter, allowing them to cross material without being stopped, making them interesting for applications.

Cosmic rays

Comic rays are **high energy particles** reaching the Earth from the outer space:

- About 98% are protons and the rest mostly helium nuclei
- Trigger a decay chain by interacting with the atmosphere, producing muons



- Muons are not stable ($\tau \simeq 2.2\mu$s) and quickly decay into one electron and two neutrinos
- Because of relativistic time dilation, they live long enough to reach the Earth surface
- Rule of thumb: 10.000 cosmic muons per m^2 and per minute are observed at sea level.

Muons interact with matter through two main processes:

Ionization

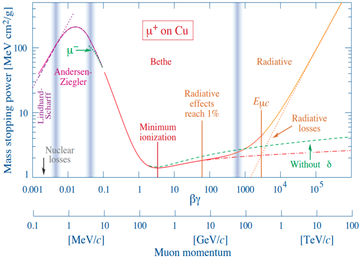
Ionization happens when the incident muon gives some of its energy to the electrons of the absorber, as described by the Bethe-Bloch formula.

→ Basis for the **absorption muography**.

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

The **mass stopping power** of material depends on:

- The charge number of incident particle z
- The atomic mass/charge of absorber A/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.

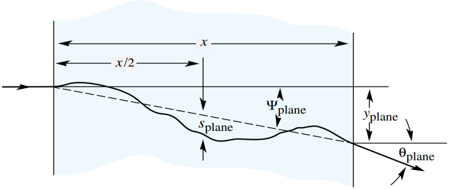


Multiple Coulomb scattering

Coulomb interactions induce a **stochastic deviation** whose central angular deviation can be described by two Gaussians of width θ_0 along the x and y-axes independently:

$$\theta_0 \simeq \frac{13.6 \text{ MeV}}{\beta c p} \sqrt{\frac{x}{X_0}}$$

θ_0 depends on the number of radiation lengths X_0 and on the medium crossed.
$\rightarrow$ Basis for the **scattering muography**.



The pairs $(\theta_{\text{plane}}, y_{\text{plane}})$ are approximately distributed by a bi-dimensional Gaussian distribution with a given covariance:

$$\text{Cov}(\theta_{\text{plane}}, y_{\text{plane}}) = \begin{bmatrix} \theta_0^2 & \frac{\theta_0^2 x}{2} \\ \frac{\theta_0^2 x}{2} & \frac{\theta_0^2 x^2}{3} \end{bmatrix}$$

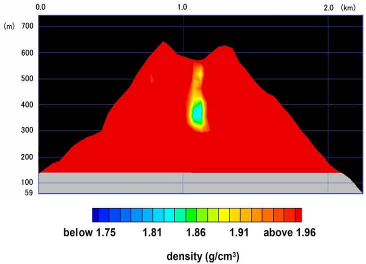
Usage of the attenuation or scattering of cosmic muons to estimate the properties (geometry and density) of the medium they cross.

Several advantages over other imaging techniques:

- Non-destructive technique
- High penetrating capabilities allowing to probe large and dense objects
- Completely safe as it uses natural cosmic rays for the measurement.

Muography can be used in many different fields:

- Archeology applications such as hidden chambers finding
- Nuclear waste/facilities inspection
- Volcanology, to know whether a pocket is empty or full of lava (Sakurajima volcano)



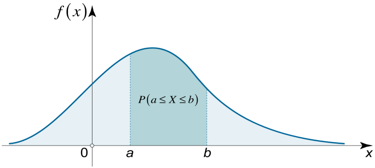
Section II
Statistical basis of the algorithm

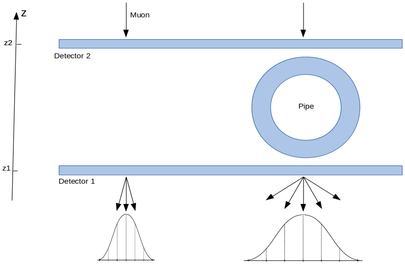
The algorithm developed heavily relies on several important statistical concepts that we can now define.

Probability density functions

PDFs are mathematical expressions defining probability distributions **which represent the likelihood of any given outcome**.

The area below the PDF in an interval can be interpreted as the value of the probability of a random variable X occurring.



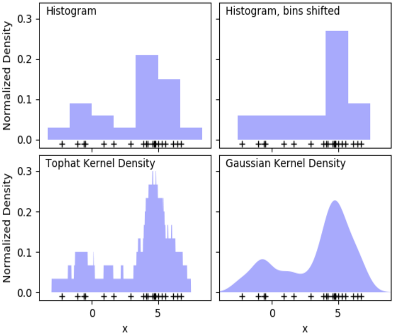


The multiple scattering is a stochastic process, making PDFs extremely important.

A thicker and denser object results in a statistically higher expected deviation and a larger standard deviation σ of the Gaussian PDF.

Kernel density estimation

This method allows us to estimate the shape of an unknown PDF f of a random variable X from a set of N observations.



The usual way to proceed is to simply put the observations in an histogram, but this results in a non continuous function with possible gaps.

We then define $\hat{f}_h(x)$, an estimator of f defined as the sum of continuous functions instead.

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

Two important parameters

- The **kernel** K, chosen as Gaussian functions in this case
- And the **bandwidth** h, a smoothing parameter.

Maximum likelihood estimation

The likelihood measures the goodness of a fit with respect to a sample of data for one or several unknown parameters. If $\boldsymbol{\theta}$ are the parameters of the model and x is the measurement of a random variable X defined from a PDF f, then:

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

The likelihood can be described as an hypersurface whose peak gives the optimal set of parameters maximizing the probability of drawing the actual sample measured $\rightarrow$ the objective is to **find the set of parameters minimizing the log-likelihood** $l(\boldsymbol{\theta}|x) = -2\log(\mathcal{L}(\boldsymbol{\theta}|x))$.

In this work, the parameter to be optimized will be the thickness of a steel pipe.

Section III
Algorithm implementation

A C++ framework has been developed in order to solve this problem, relying on two main parts: a PipeReconstructor and a Generator, both described in the next slides.

The idea behind the algorithm is to simulate the propagation of each muon in a dataset several times to estimate its output PDF for a given geometry and evaluate a given measurement to compute a likelihood.

General idea

We developed a framework to study the results from a muography experiment to characterize the inner properties of physical objects such as the thickness of a steel pipe.

PipeReconstructor

General set of classes defined in the next slides allowing us to:

- Define the geometry of the problem (mainly, the detector and object position and size)
- Compute the intersection points between cosmic muons and this geometry
- Propagate these muons through the geometry using the multiple scattering
- Calculate the likelihood of a given measurement/simulation for a given geometry
- Finally, find the optimal object parameters from this likelihood.

Generator

Another class allowing us to generate muons by performing Monte-Carlo simulations using some of the functions of the PipeReconstructor and without relying on Geant4, a toolkit developed by CERN in order to simulate the passage of particles through matter.

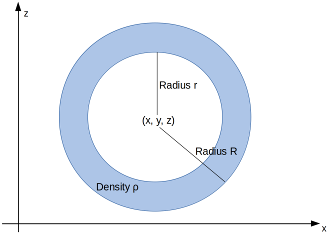
Surfaces, volumes and MuonStates

Surfaces and Volumes

General virtual classes defining the geometry of the problem ; a Volume is defined as a vector of a fixed number of Surfaces.

In particular, the subclasses Cylinder and Pipe are mostly used in this work. A Pipe is defined from 7 parameters:

- Its central position (x, y, z)
- Its inner r and outer radii R
- Its constant density ρ
- And its length L along the axis of the cylinder.



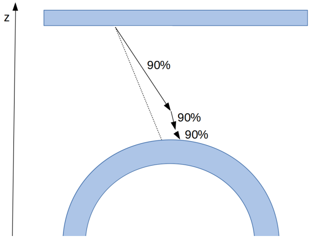
MuonStates

A MuonState is a class defining a muon from two vectors representing its position (x, y, z), direction (v_x, v_y, v_z) and momentum value p.

Propagator

The propagator is the object allowing to **propagate a MuonState through a Volume**:

1. First, the distances between a MuonState and all the Surfaces of the Volume are computed and the first intersection point is kept
2. The muon is propagated 90% of the distance to this cut point using the multiple scattering, and this process is repeated several times until being closer to 0.1mm to the first surface
3. The Surface is manually crossed by slightly moving the MuonState along its direction
4. This process is repeated for all the Surfaces of the Volume and then one last time until reaching the bottom detector, where the MuonState is returned.

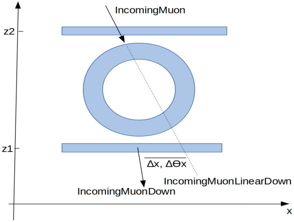


Eventual muons which do not actually cross the pipe or out of the acceptance of any of the detectors are rejected at this stage.

Likelihood

The Likelihood class takes as input a Volume and a data/MC file, to tell us **how likely it is that we obtained exactly these measurements/simulations for the geometry given**:

1. First, four MuonStates are computed:
   - The IncomingMuon and OutgoingMuon, as the actual simulations made at the top/bottom detectors
   - The IncomingMuonLinearDown, the linear propagation of the IncomingMuon, used as reference
   - And the IncomingMuonDown, by propagating the initial state with our Propagator through the Volume



2. This process is repeated n_{iter} times for each input event, obtaining each time values for the Δx, $\Delta\theta_x$, Δy and $\Delta\theta_y$ parameters, filling 2 bi-dimensional PDFs histograms
3. The histograms obtained are smoothen using the kernel density estimation method
4. The probability to observe the actual value of OutgoingMuon is computed, and summed over all the N_{MC} events in the file, returning the value:

$$\mathcal{L} = \sum_{i=1}^{N_{\text{MC}}} -2.0\left(\log(\text{value}_{x,i}) + \log(\text{value}_{y,i})\right)$$

Section IV
Results obtained

First of all, the results obtained by our Generator whill be shown and compared to Geant4, before moving on to the actual results obtained with our algorithm.

All the results shown here are based on Monte-Carlo simulations, relying on CRY, a cosmic ray generator.

Generator and Monte-Carlo simulations

Monte-Carlo simulations are obtained from algorithms developed to **compute approximate numerical values to stochastic problems** using random processes and probabilistic techniques.

Instead of relying on actual data collected from the experiment, we can:

1. Use CRY, a cosmic ray generator to simulate thousands of incident muons
2. Make these muons go through a complete simulation of the detector done with Geant4
3. Go through the same post-processing process as actual data
4. Build the PDF for a given experiment from these simulations
5. Repeat this experiment N_{MC} times, **simulating thousands of experiments** without the need to perform them.

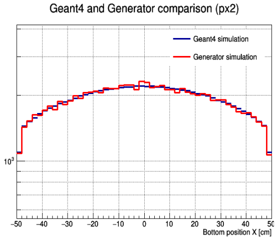
Simulating an experiment is cheaper and faster than running it and allows to compare results obtained from both channels. In this work, the dependence on Geant4 will be removed as well, to make this process even faster.

Generator validation

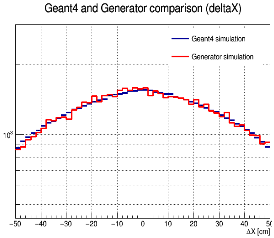
The Generator allows us to generate Monte-Carlo experiments in a faster way than using Geant4, which is more complex, relying on a complete description of the detector.

The first step was then to validate our Monte-Carlo simulations by comparing measurements simulated by both Geant4 and our own Generator for a given geometry.

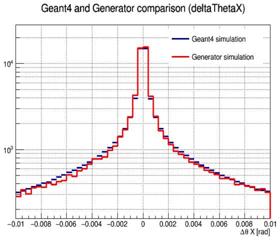
Bottom X position



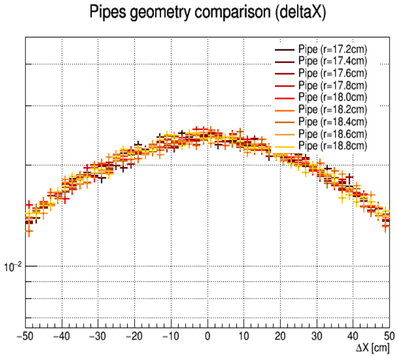
Δ_x parameter

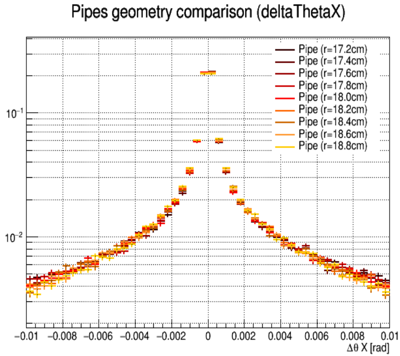


$\Delta\theta_x$ parameter



Once validated, 12 different Monte-Carlo files have been generated with our Generator with 10 to 50.000 events each, for different pipe geometries, having an outer radius of 20cm and inner radii ranging from 16.8 to 19.0cm.

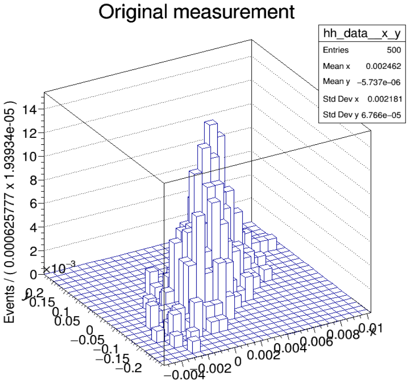


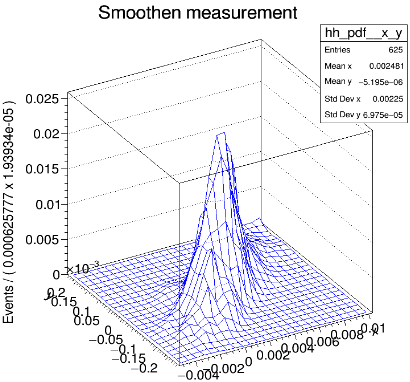


A sophisticated method is needed to analyze such small deviations between the geometries.

It took 8 seconds to produce 50.000 events with our Generator, more than **two orders of magnitude faster** than a complete Geant4 simulation.

Since the likelihood needs to be computed for every single event of the input simulation file by performing another computing extensive loop, the n_{iter} parameter is kept as small as possible using the kernel density estimation method.



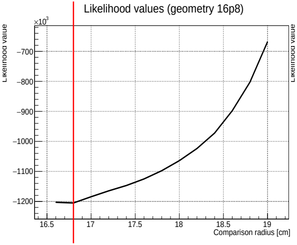


Likelihood curves

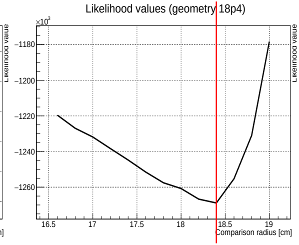
Finally, we estimated the value of the total likelihood obtained for different pipe geometries, characterized by different inner radii, ranging from 16.6 to 19.0cm, by steps of 0.2cm.

The idea was to plot the likelihood obtained comparing one by one the generated file with the different geometries available and try to figure out which pipe geometry, in the x-axis, is more likely to give rise to the file considered.

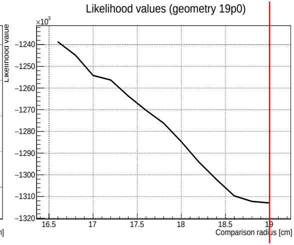
$r = 16.8\text{cm}$



$r = 18.4\text{cm}$



$r = 19.0\text{cm}$

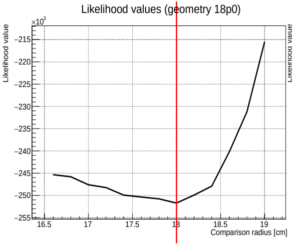


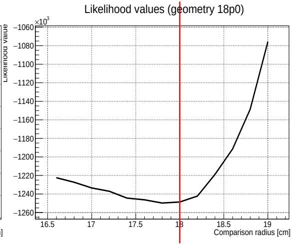
The minimum observed matches the correct radius with a precision of a few millimeters for all the geometries, as seen in the backup.

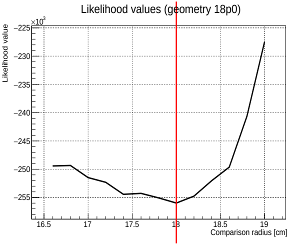
Likelihood curves

We also estimated the impact of the number of simulated events n_{MC} and the number of likelihood computation iterations n_{iter} on the likelihood curves, for the $r = 18.0$cm geometry.

$n_{\text{iter}} = 100$ $n_{\text{MC}} = 10.000$	$n_{\text{iter}} = 100$ $n_{\text{MC}} = 50.000$	$n_{\text{iter}} = 250$ $n_{\text{MC}} = 10.000$







In this case, it does seem that increasing the number of simulated events improve quite a lot the results (note that 10.000 events correspond to 1 minute of data taking only, **15 times faster** than the current algorithm implementation).

In conclusion, we developed a new framework allowing us to:

- Quickly generate thousands of Monte-Carlo experiments for different pipe geometries without relying on Geant4
- Compare the output distributions expected at the bottom detector after propagating a muon through our Volume, using our own Propagator
- Compute the likelihood of a given measurement with respect to different pipe geometries, in order to find a way to estimate the thickness of a steel pipe by studying the deviation of incident cosmic muons
- And plot all the results obtained.

With this framework, we were able to determine the thickness of such pipes with **a precision of the order of the millimiter** by considering 10.000 events, which is equivalent to **1 minute of data taking only**, therefore solving the initial problem solved.

This exercise is only a first approach to the problem, and different improvements can be considered to **improve and/or generalize the results obtained**.

Backup slides

Ionization

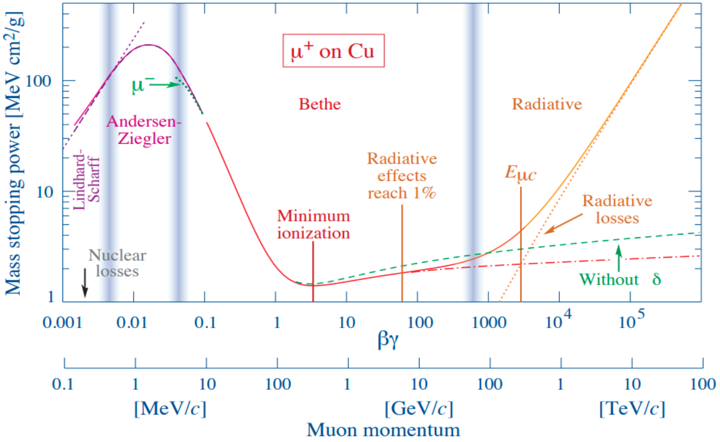
Ionization happens when the incident muon gives some of its energy to the electrons of the absorber, as described by the Bethe-bloch formula.

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

The **mass stopping power** of material depends on:

- The charge number of incident particle z
- 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.

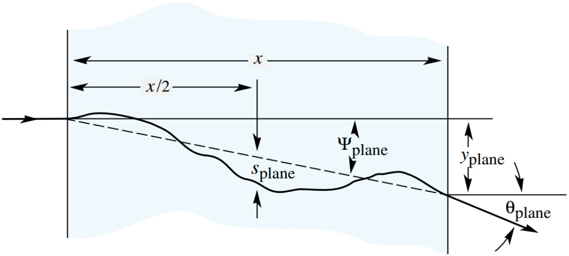
Cosmic muons have an energy of the order of the GeV and are therefore referred to as minimum ionizing particles, so ionization is not considered in this work.



Multiple Coulomb scattering

Inducing a **stochastic deviation** whose central angular deviation can be described by two Gaussians of width θ_0 along the x and y-axes under our experimental conditions.

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



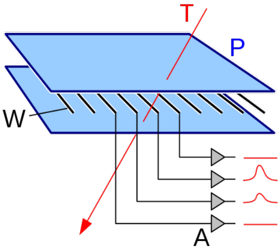
Highly correlated deviation parameters ($\rho_{\theta_{\text{plane}}, y_{\text{plane}}} = \sqrt{3}/2$). The pairs ($\theta_{\text{plane}}$, y_{plane}) are approximately distributed by a bi-dimensional Gaussian distribution with a given covariance:

$$\text{Cov}(\theta_{\text{plane}}, y_{\text{plane}}) = \begin{bmatrix} \theta_0^2 & \frac{\theta_0^2 x}{2} \\ \frac{\theta_0^2 x}{2} & \frac{\theta_0^2 x^2}{3} \end{bmatrix} \quad (1)$$

Muon detectors

Multiwire proportional chambers use an array of high-voltage wires, placed within a chamber filled with a gas, in which an electric field is created.

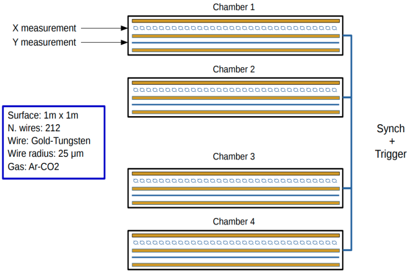
A muon crosses the detector leaves small electric charges behind, collected by the wires while leaving a signal. The combination of the signals on the different wires give us information regarding the muon.



Most important parameters of a muon detector:

- The **spatial resolution**, ideally as small as possible
- The **acceptance**, related to the size of the detector
- The **efficiency**, which should be high to make the measurement reliable and fast.

Experimental setup

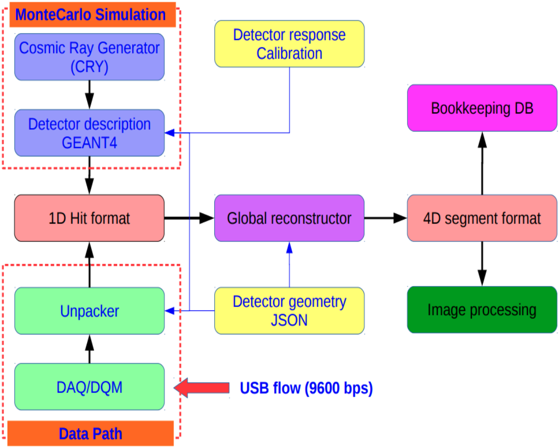


Quite simple experimental setup:

- Two $1m^2$ detectors placed below and above the object under investigation
- Two chambers in each detector, to measure the position and direction of muons along the x and y axes
- Each chamber is filled with a mixture of Argon and CO_2
- More than 200 wires separated by 4mm make up each chamber

The data is collected from a USB stick and goes through a complete **reconstruction process** before being available in a rootfile, as detailed in the backup.

Data flow


MonteCarlo Simulation
Cosmic Ray Generator (CRY)
Detector description GEANT4
Detector response Calibration
Bookkeeping DB
1D Hit format
Global reconstructor
4D segment format
Unpacker
Detector geometry JSON
Image processing
DAQ/DQM
USB flow (9600 bps)
Data Path

Positional and angular deviation parameters

Main parameters used throughout this work:

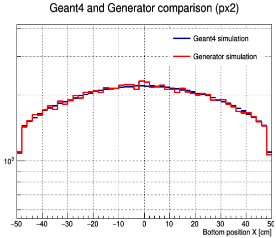
$$\begin{cases} \Delta \mathrm{x} = x_2 + d(v_{x2} - x_1) \\ \Delta \mathrm{y} = y_2 + d(v_{y2} - y_1) \end{cases}$$

$$\begin{cases} \Delta\theta_x = \arctan\left(\dfrac{v_{x2}}{v_{z2}}\right) - \arctan\left(\dfrac{v_{x1}}{v_{z1}}\right) \\ \Delta\theta_y = \arctan\left(\dfrac{v_{y2}}{v_{z2}}\right) - \arctan\left(\dfrac{v_{y1}}{v_{z1}}\right) \end{cases}$$

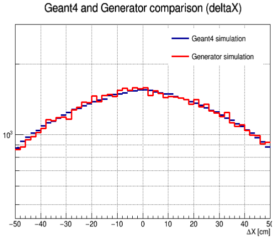
Bi-dimensional histograms (Δx vs $\Delta\theta_x$ and Δy vs $\Delta\theta_y$) are filled with these values, smoothened and used for the computation of the likelihood.

Generator validation

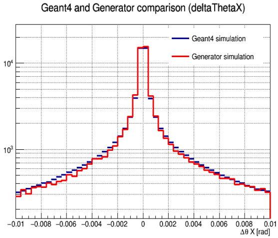
Bottom X position



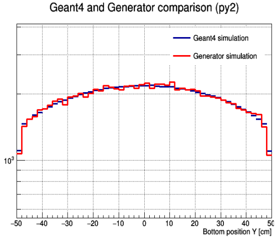
Δ_x parameter



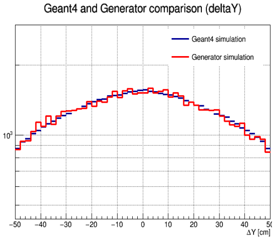
$\Delta\theta_x$ parameter



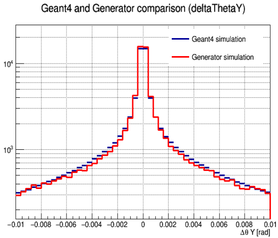
Bottom Y position



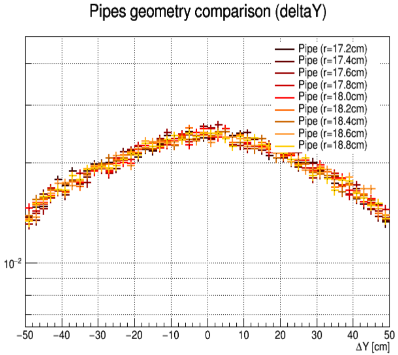
Δ_y parameter

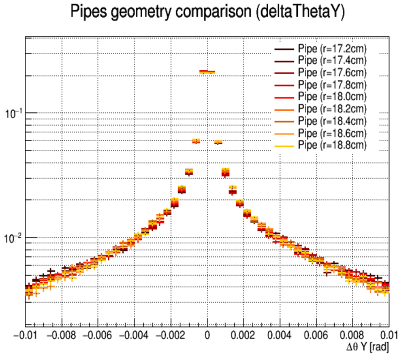


$\Delta\theta_y$ parameter



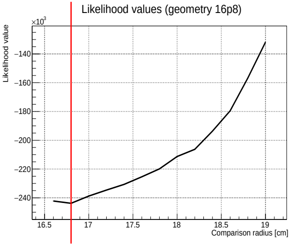
The same comparison has been performed along the Y-axis.



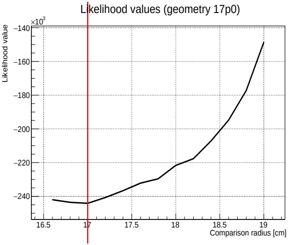


Likelihood curves (10.000 events, 100 iterations)

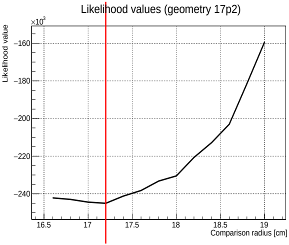
$r = 16.8$cm



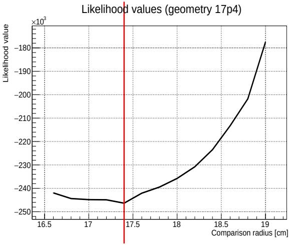
$r = 17.0$cm



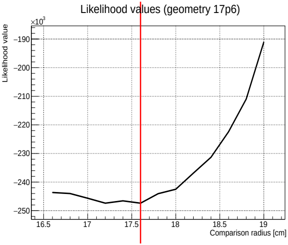
$r = 17.2$cm



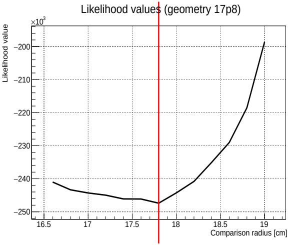
$r = 17.4$cm



$r = 17.6$cm

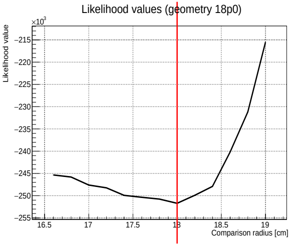


$r = 17.8$cm

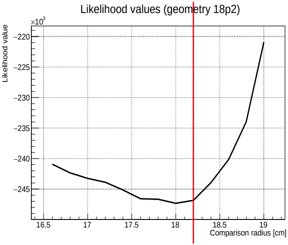


Likelihood curves (10.000 events, 100 iterations)

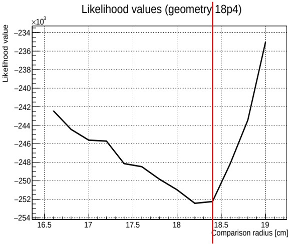
$r = 18.0$cm



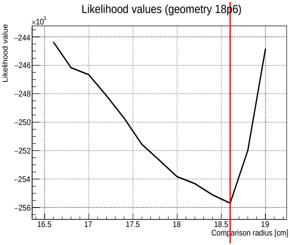
$r = 18.2$cm



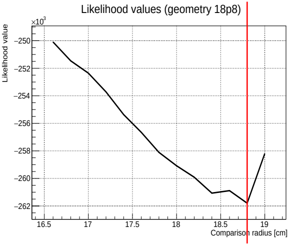
$r = 18.4$cm



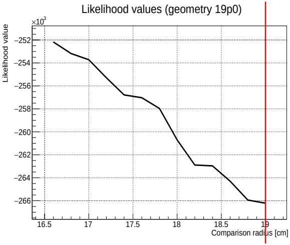
$r = 18.6$cm



$r = 18.8$cm

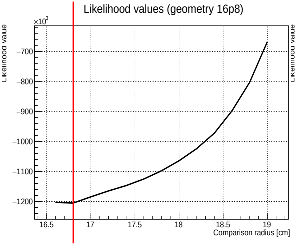


$r = 19.0$cm

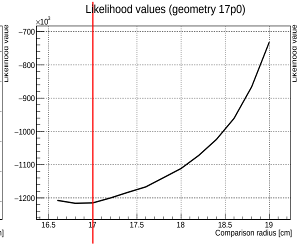


Likelihood curves (50.000 events, 100 iterations)

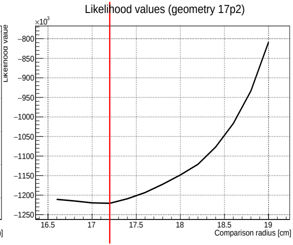
$r = 16.8$cm



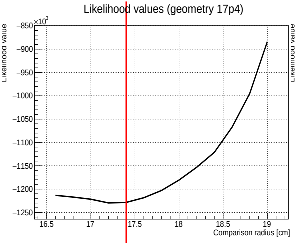
$r = 17.0$cm



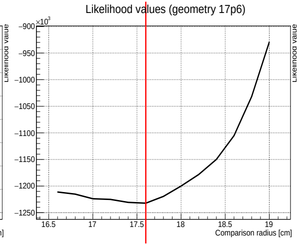
$r = 17.2$cm



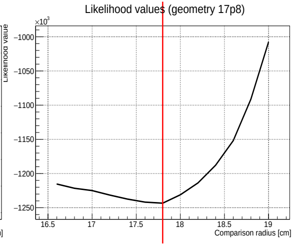
$r = 17.4$cm



$r = 17.6$cm

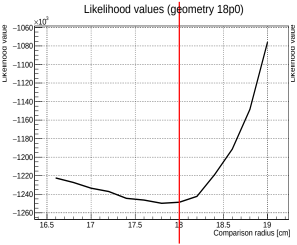


$r = 17.8$cm

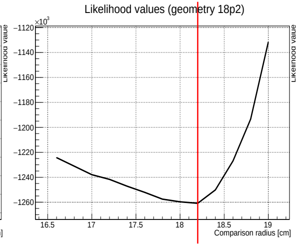


Likelihood curves (50.000 events, 100 iterations)

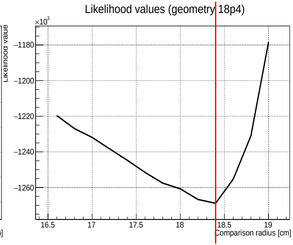
$r = 18.0$cm



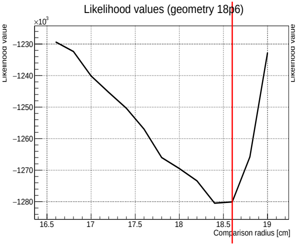
$r = 18.2$cm



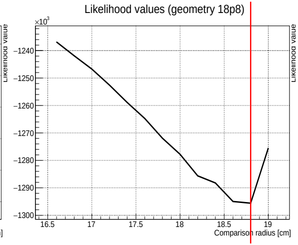
$r = 18.4$cm



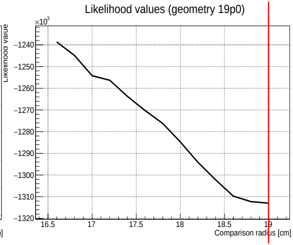
$r = 18.6$cm



$r = 18.8$cm

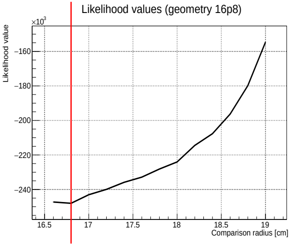


$r = 19.0$cm

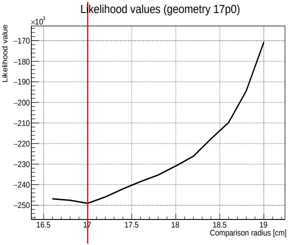


Likelihood curves (10.000 events, 250 iterations)

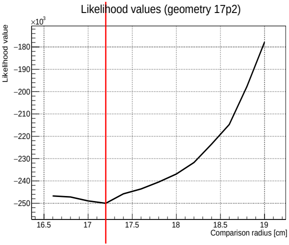
$r = 16.8$cm



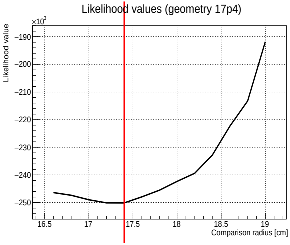
$r = 17.0$cm



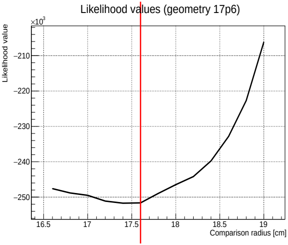
$r = 17.2$cm



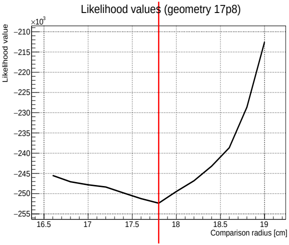
$r = 17.4$cm



$r = 17.6$cm

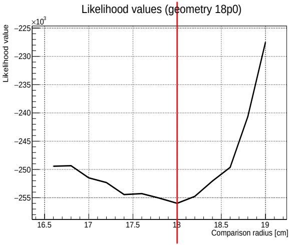


$r = 17.8$cm

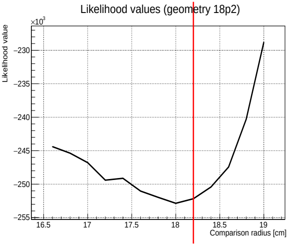


Likelihood curves (10.000 events, 250 iterations)

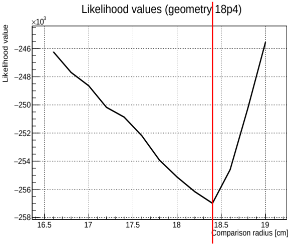
$r = 18.0$cm



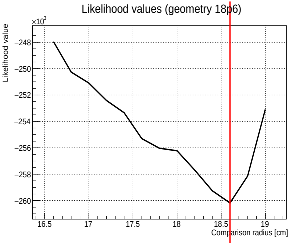
$r = 18.2$cm



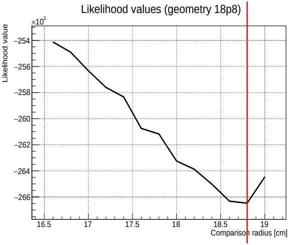
$r = 18.4$cm



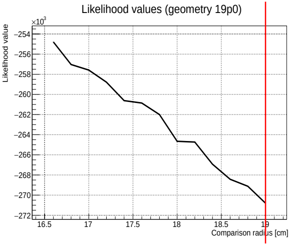
$r = 18.6$cm



$r = 18.8$cm



$r = 19.0$cm



This exercise is only a first approach to the problem, and different improvements can be considered to **improve and/or generalize the results obtained**:

- Consider more general geometries that a pipe, or perform the likelihood minimization with respect to additional parameters, not only the thickness of the pipe
- The analysis can be repeated using actual data collected by the detector
- We could also consider the ionization process to improve these results
- The interaction between the detector and the cosmic muons could be considered in our Generator to make it more reliable and precise
- A further analysis estimating the Hessian of the likelihoods and the impact of systematic uncertainties would definitely improve the algorithm
- Finally, we have been limited computionally in this case, taking a few hours to produce a single plot with few statistics. Accessing to computers with higher capacities will be extremely interesting to improve the results obtained.