Cosmic Rays detection, Statistical Overview and Data analysis with Python

Realized by

Jacopo Martellotto – 5^B
Liceo Scientifico C. De Giorgi – Lecce

- **ABSTRACT**

During the International Cosmic Day 2021, we detect the muons arriving at the Earth's surface through a proper detector. The detector is made up of four different layers and any muon is marked as detected when it crosses at least two layers. In particular, we talk about double, triple or quadruple crossing event when a muon crosses two, three or four adjacent layers, respectively.
In detail, we wanted to verify whether the incidence rate (R) of the muons onto the detector layers was proportional to $\cos^2\theta$, where θ is the zenith angle of the muon flux. We successfully reached our goal, since the dependence of R vs $\cos^2\theta$ is described by a linear fit, having a coefficient of determination equal to:
    - 0.985 (double crossing-event)
    - 0.979 (triple crossing-event)
    - 0.944 (quadruple crossing-event)

- **AIM**
    - Calculate the incidence rate of muons on the detector;
    - check the mathematical law on the muon count;
    - compare the real results with the theoretical ones;

- **THEORETICAL BACKGROUND**

    - <u>**Cosmic rays and muons**</u>

Cosmic rays - high-energy protons and atomic nuclei that move through space at just below the speed of light - once entering the Earth's atmosphere, collide with the molecules of the Earth's atmosphere, giving rise to other particles in cascades (secondary cosmic rays).
Muons - particles resulting from the collisions - are part of cosmic rays: their characteristics are almost the same as those of the electrons, but their mass is 207 times higher.
The average life of a muon at rest is about 2.2 μs.
However, its average life does not appear to be long enough for the particle to touch the Earth's surface. In fact, assuming the speed of light as the reference speed for the muons, just for a rough estimate of covered distance, we would obtain:

$$\boldsymbol{s = ct \cong 3 \cdot \frac{10^5 km}{s} \cdot 2.2 \cdot 10^{-6} s \cong 0.7\ km},$$

which is significantly smaller than the typical height, where the muons form.
Therefore, the Theory of Relativity is essential to justify the detection of muons at Earth's surface. In fact, the time flows slower than in a moving reference frame.

We want to evaluate the probability that the muon survives, after a time lapse of a Δt.
If $\boldsymbol{\tau}$ is the average life of the muon, for a generic instant $\boldsymbol{t = N\Delta t}$ (where N is an integer), we obtain:

$$\boldsymbol{p(t) = \left(1 - \frac{t}{N\tau}\right)}$$

If N is very big, we can rewrite: $\boldsymbol{p(t) = e^{-\frac{t}{\tau}} \cong 10^{-7}}$

This result does not appear to be acceptable. It is necessary to consider the effects of Special Relativity. The muon, in fact, travels at the speed of the light, compared to an observer stationary on the Earth's surface.

We could rewrite the average life as:

$$\boldsymbol{\tau_{Earth} = \gamma\tau}$$

where $\boldsymbol{\gamma} = \left(\mathbf{1} - \frac{v^2}{c^2}\right)^{-1/2}$ is the Lorenz factor.

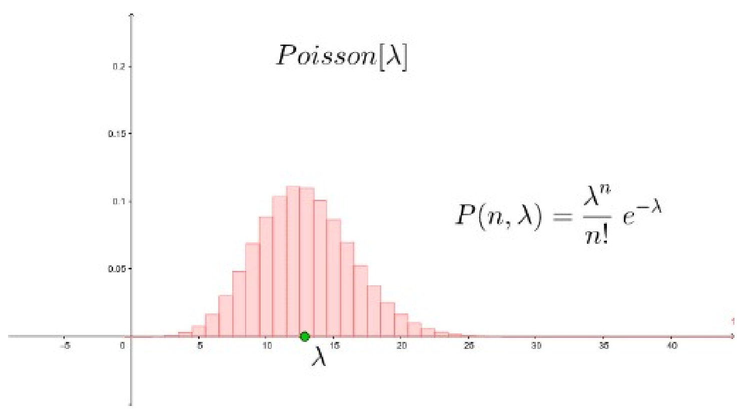
$$\boldsymbol{E = \gamma m_\mu c^2}$$

From the energy measurements, $E = 3\ {GeV}/{c^2}$, we infer the value of $\gamma (\cong 29)$.
By recalculating the probability, we get:

$$\boldsymbol{p(t) \cong 0.6 = 60\%}$$

This time, the result obtained is consistent with the observed phenomenon and is able to explain the presence of muons at sea level.

- **Poisson distribution**



Each measure is affected by systematic and statistical errors. The distribution that best approximates the cosmic ray counting phenomena is the **Poisson distribution**.
Using the definition of the central moment of order 2, it can be shown that the standard deviation is equal to the **square root of the characteristic parameter of the Poisson distribution** (in our case $\boldsymbol{\sqrt{\lambda}}$, where $\boldsymbol{\lambda}$ is the number of events for each measure).

The Poisson distribution is used when an event *E* satisfies the following three assumptions:

1) *The probability of an event occurring in a very short time,* dt, *is proportional to the time interval itself.*

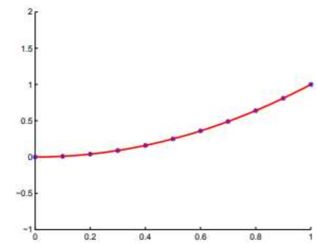
2) *The probability of a second event occurring in the same interval dt is very small.*

3) *Each event is independent of previous or subsequent events.*

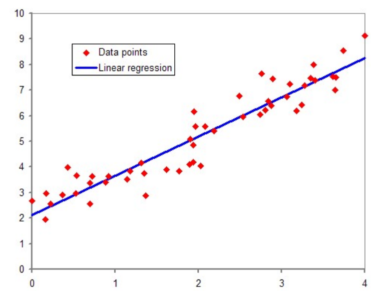
- **<u>The method of least squares</u>**

  We want to determine a function that best approximates our measurement dataset. To do this, it is necessary to talk about interpolation, distinguishing the cases, and the least squares method.

  1) If the function assumes exactly the detected values, and therefore its graph passes through all the points of the scatter plot, we speak of **interpolation for known points**;



  2) If the function assumes values close to the detected values and therefore its graph passes between the points of the scatter plot, we are talking about interpolation between known points or **statistical interpolation.**



There are various ways to determine the interpolating function, however we will use the **least squares method** (operation of "**best fit**").

When in correspondence of x_i we have:

$$\begin{cases} y_i \dashrightarrow \text{ values detected} \\ \hat{y}_i \dashrightarrow \text{theorical values} \end{cases}$$

the condition to obtain the best fit is given by: $\boldsymbol{\varphi = \sum_{i=1}^{n}(y_i - \hat{y}_i)^2 = minimum}$

If we are interested in **<u>a linear fit</u>**, described by**:** $\boldsymbol{\hat{y} = bx + a}$**,** then φ can be rewritten as $\boldsymbol{\varphi(a, b) = \sum_{i=1}^{n}(y_i - a - bx_i)^2}$**.**

For the calculation of the minimum, it is necessary to impose that the first partial derivatives with respect to a and b are 0. The computation yields to the following results:

$$\begin{cases} \boldsymbol{a = \dfrac{\sum y_i \sum x_i{}^2 - \sum x_i \sum x_i y_i}{n \sum x_i{}^2 - (\sum x_i)^2}} \\ \boldsymbol{b = \dfrac{n \sum x_i y_i - \sum x_i \sum y_i}{n \sum x_i{}^2 - (\sum x_i)^2}} \end{cases}$$

It is clear that the slope $\boldsymbol{b} = \frac{\boldsymbol{\sigma_{xy}}}{\boldsymbol{\sigma_x^2}}$, where σ_{xy} is the covariance between x and y, while σ_x^2 is the variance of x. Instead, $\boldsymbol{a}$ is the consequence of the best-fit line, passing through $(\bar{x};\ \bar{y}\ )$, where $\bar{x}$ and $\bar{y}$ are the mean values of x and y, respectively.

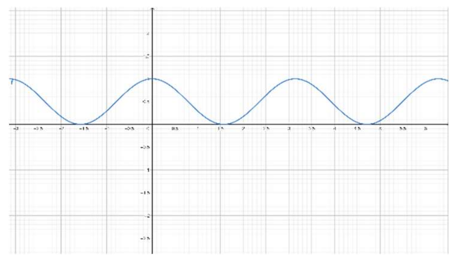
Therefore, the equation of the interpolating line appears also as:

$$\boldsymbol{y - \bar{y} = b(x - \bar{x})}$$

- **Measurement of the cosmic ray flux as a function of the zenith angle**

The flux of muons reaching the ground is not uniformly distributed, but depends on the angle that the particles form with the local zenith.

As you can read in the paper referenced below [1.]: ***The overall angular distribution of muons at the ground is $\propto \cos^2\theta$, which is characteristic of muons with $E_\mu \sim 3$ GeV.*** *At lower energy the angular distribution becomes increasingly steep, while at higher energy it flattens, approaching a sec θ distribution for* $E_\mu \gg$ *115 GeV and θ < 70°.*



Representation of the function cos ^ 2

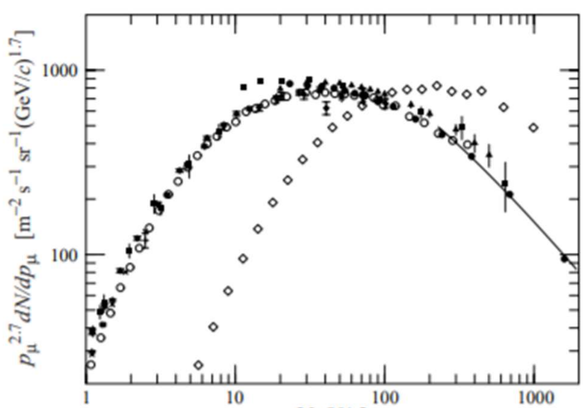


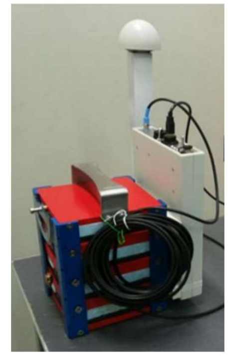
Figure 29.5: Spectrum of muons at $\theta = 0°$ (♦ [50], ■ [56], ▼ [57], ▲ [58], ×, + [52], ○ [53], and ● [54] and $\theta = 75°$ ◊ [59])

It turns out that the maximum flow occurs for θ = 0° and minimum for 90°. Muons arriving perpendicular to the Earth's surface, along the direction of the local zenith, travel less road in the atmosphere.

For higher angles, the distance becomes greater.

In conclusion, the greater the angle from the zenith, the greater the distance travelled, the greater the probability that the muons decay before reaching the earth's surface. Therefore, the flux decreases with the zenith angle.

- **EXPERIMENTAL SETUP**



The CORAM detector is made up of four scintillator layers, named X, Y, Z, W. Each layer consists of a scintillation detector and photodetectors (APD).
A scintillator is a material capable of emitting pulses of light when it is crossed by charged particles, while the photodetectors convert the light into electrical pulses. Between each scintillator layer, a layer of iron is interposed, in order to select the most "penetrating" particles (muons).

For a given interval T, we define the following kind of signals:
• **Single** (signal registered by the single layers)
• **Doubles** (two adjacent layers register a signal within a ‘T’ time)
• **Triple** (three adjacent layers register a signal within a ‘T’ time)
• **Quadruple** (four adjacent layers register a signal within a time ‘T’)

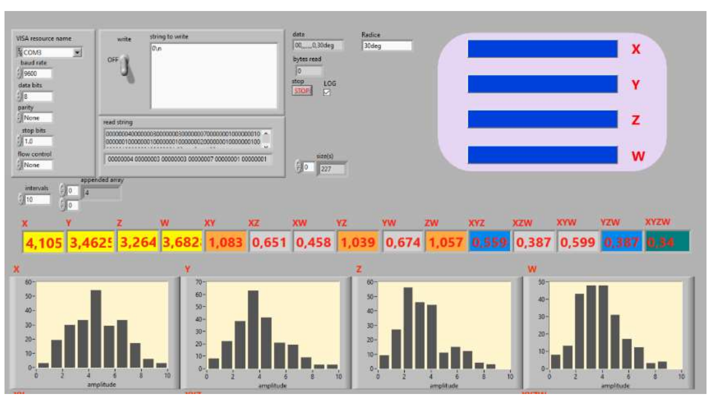
What we are going to calculate is the rate, expressed as:

$$\boldsymbol{R} = \frac{\boldsymbol{C}}{\boldsymbol{\Delta T}} \ \frac{\boldsymbol{part}}{\boldsymbol{s}}$$

Where:

C is the number of cosmic rays recorded by the detector

$\boldsymbol{\Delta T}$ is the time interval taken into consideration (in our case it is $3s$)



Software Interface

- **METHODS**

Data analysis and graph plotting are done with Python. The libraries that are used are *math*, *pandas*, *numpy*, *matplotlib*, *scipy* and *IPython*.
Let's analyse the mathematical concepts that are taken into consideration by libraries to better understand how they work.

- **Variance, covariance and coefficient of determination**

The **variance** is the arithmetic mean of the squares of the differences between each $\boldsymbol{x_i}$ value of the distribution and the average value, $\boldsymbol{\overline{x}}$.
The variance identifies the dispersion of the values of the variable x around the mean value. The smaller the variance, the more the values of the variable are concentrated around the mean value.

$$\boldsymbol{\sigma^2 = \frac{\sum_{i=1}^{n}(x_i - \overline{x})^2}{n}}$$

The **covariance** of two statistical variables or random variables is a numerical value that provides a measure of how much the two vary together, or of their dependence.

The covariance of X and Y is the average of the product of the mean deviations $S(x_i)$ and $S(y_i)$.

$$\boldsymbol{Cov(X,Y) = \frac{1}{n}\sum_{i=1}^{n} S(x_i)S(y_i) = \frac{1}{n}\sum_{i=1}^{n}[x_i - M(x)][y_i - M(y)]}$$

The covariance of X and Y can also be written as the average difference between the products of the X and Y values and the product of the means.

$$\boldsymbol{Cov(X,Y) = \frac{1}{n}\sum_{i=1}^{n}[x_i - M(x)][y_i - M(y)] = \frac{1}{n}\sum_{i=1}^{n} x_i y_i - M(x)M(y)}$$

In statistics, **the coefficient of determination** is an index that quantifies up to each extent the model in use is able to predict the data variability.

$$\boldsymbol{R^2 = 1 - \frac{\sum_{i=1}^{n}(y_i - \widehat{y}_i)^2}{\sum_{i=1}^{n}(y_i - \overline{y})^2}}$$

Where:

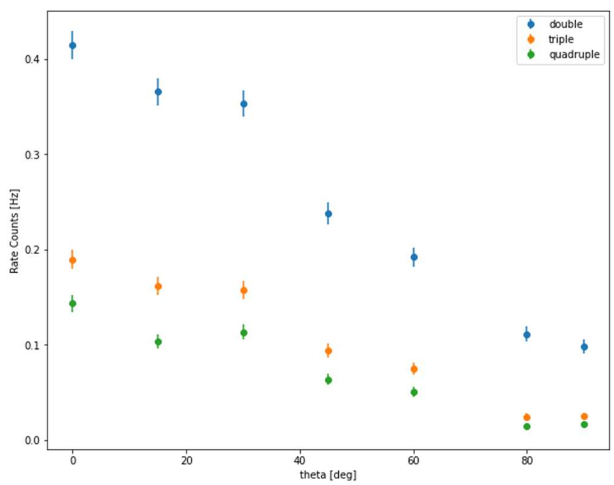
- y_i *are the observed data*, corresponding to x_i;
- $\bar{y}$ *is their average*;
- $\hat{y}_i$ *are the data estimated by the model, corresponding to* x_i.

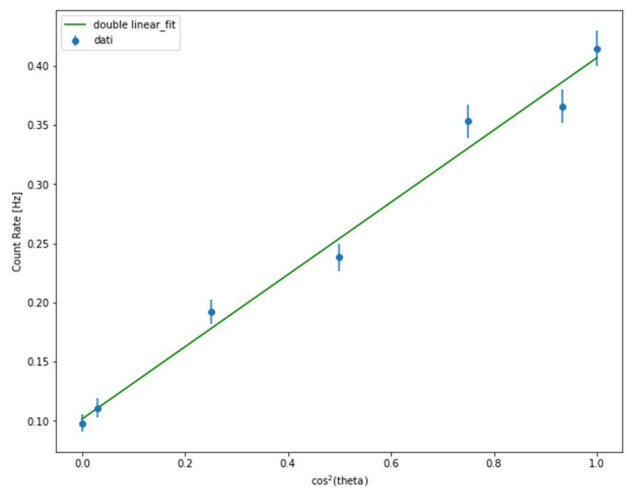
In particular, $\boldsymbol{\sum_{i=1}^{n}(y_i - \overline{y})^2}$ is the total deviance of the data, while $\boldsymbol{\sum_{i=1}^{n}(y_i - \widehat{y}_i)^2}$ is the residual deviance, that is the deviance which is not explained by the model.
The complement to unity of the ratio between the residual deviance and the total deviance gives $\boldsymbol{R^2}$**,** which quantifies the fraction of deviance explained by the model.

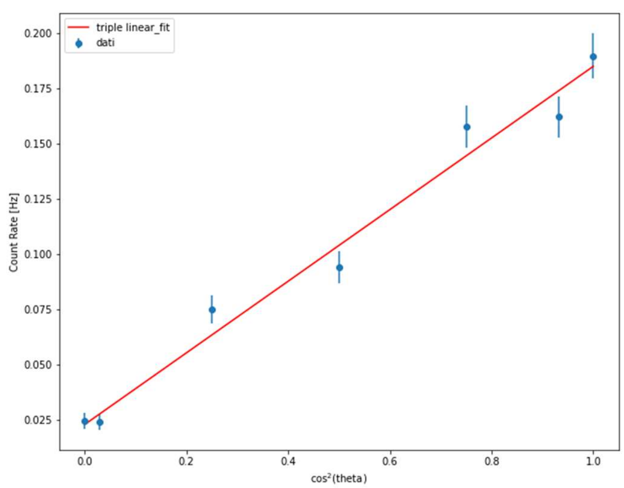
- **RESULTS**

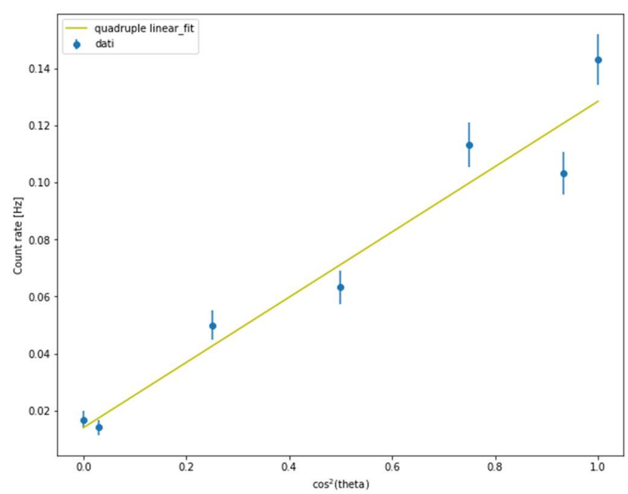
```
#linear fit
fig, ax = plt.subplots(figsize=(10,8))

cos2=np.cos(np.radians(df["angle"]))**2

plt.ylabel("Count Rate [Hz]");
plt.xlabel("cos$^2$(theta)");
plt.errorbar(cos2,rate_double,yerr=rate_error_double,marker='o',linestyle='None',label="dati")

pp,residuals, _, _, _= np.polyfit(cos2,rate_double,1,w=1./rate_error_double,full=True,cov=True)
xval = np.linspace(0, 1)
polynomial_p = np.poly1d(pp)
yval_p = polynomial_p(xval)
plt.plot(xval,yval_p,'-', color='g',label="double linear_fit")
plt.legend(loc='upper left')
```

Below, the plots representing the function:









The following snippets allows you to calculate the coefficient of determination:

```
#determine the coefficient of determination
x_values = [cos2]
y_values = [rate_triple]
correlation_matrix = np.corrcoef(x_values, y_values)
correlation_xy = correlation_matrix[0,1]
print('Coefficient of correlation:', correlation_xy)
r_squared = correlation_xy**2
print('Coeffcient of determination:', r_squared)
```

Or manually:

```
covariance= (len(cos2)-1)/len(cos2)*np.cov([cos2, rate_double])
covar=covariance[1,0]
stdx_0=np.sqrt(covariance[0,0])
stdy_0=np.sqrt(covariance[1,1])
stdx=np.sqrt(np.var([cos2]))
stdy=np.sqrt(np.var([rate_double]))
Coeffcorr=covar/(stdx*stdy)
print('Coefficient of correlation (manual):', Coeffcorr)
r_2 = (Coeffcorr**2)
print('Coefficient of determination:', r_2)
```

These are the results obtained:

- 0.985 (double crossing-event)
- 0.979 (triple crossing-event)
- 0.944 (quadruple crossing-event)

- **DISCUSSION**

1) Let's consider the relation between $\frac{R(\theta)}{R(0)}$ and $\frac{I(\theta)}{I(0)}$.

$R(O)$ and $I(0)$ are respectively the rate and intensity of the muons arriving vertically ($\theta = 0$).

With the same instrument, we can consider:

$$\boldsymbol{I = \frac{R}{G * \varepsilon}}$$

where:

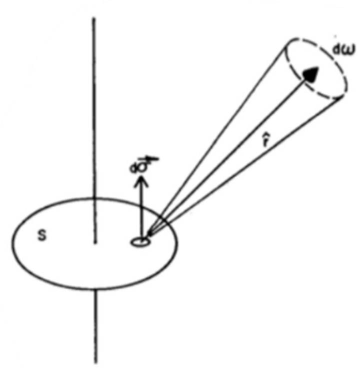
R is the rate;

G is the acceptance (a geometrical factor)

$\boldsymbol{\varepsilon}$ is related to the efficiency of the detector

$\boldsymbol{G * \varepsilon}$ is a constant conversion factor

$$\frac{I}{I(0)} = \frac{R}{G * \varepsilon} * \frac{G * \varepsilon}{R(0)} => \frac{I}{I(0)} = \frac{R}{R(0)}$$



G is described by: $\boldsymbol{G = \int_{\pi} d\omega \int_{S} d\vec{\sigma} \cdot \hat{r} F(\omega)}$

Where:

$d\omega \rightarrow element\ of\ solid\ angle$

$\hat{r}\ d\sigma \rightarrow element\ of\ area\ looking\ into\ \omega$

2) If the intensity is a function of the angle, the acceptance depends on the intrinsic properties of the detector: the greater the solid angle, the greater the acceptance. In fact, a larger solid angle corresponds to a larger collection basin.

3) Before taking measurements, it is necessary to make sure that the detector is perpendicular to the ground, otherwise the measurements may be biased.

- **CONCLUSION**

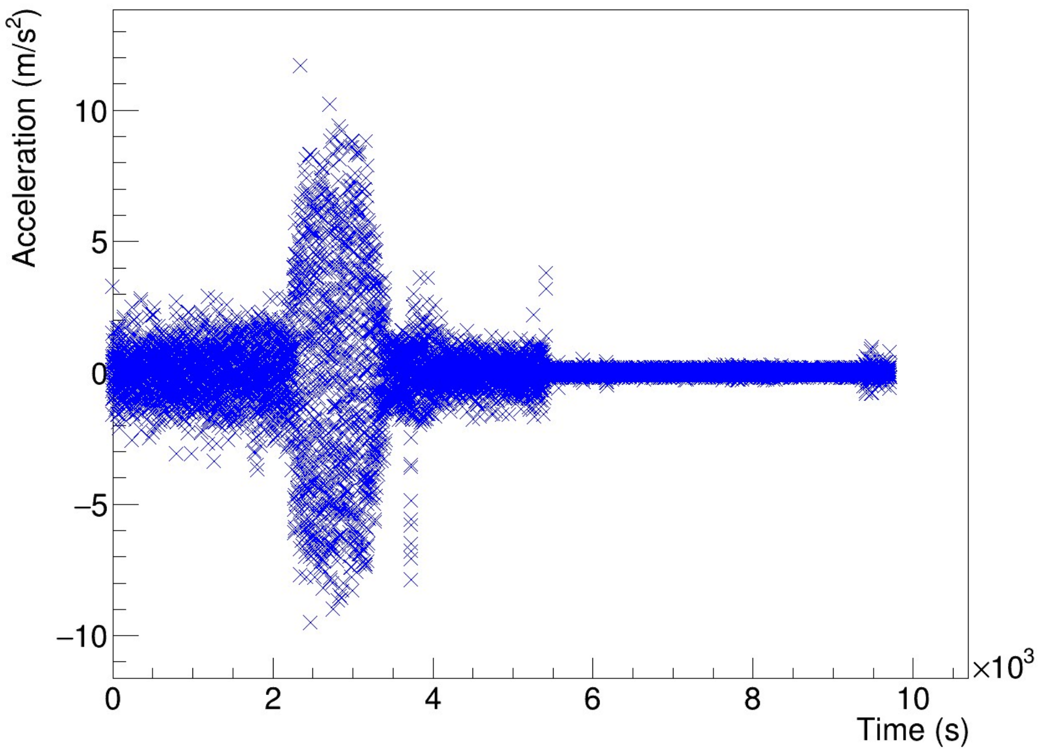
We detected the muons onto the detector layers and we verified that the dependence of the incidence rate R vs $\cos^2\theta$ is linear as expected. In fact, the coefficient of determinations of the best fit models are all between 0.944 and 0.985. This means that the data variability, that is not explained by the linear models is at least below 6%, which appears to be an acceptable threshold, considering that any experiment is affected by errors.

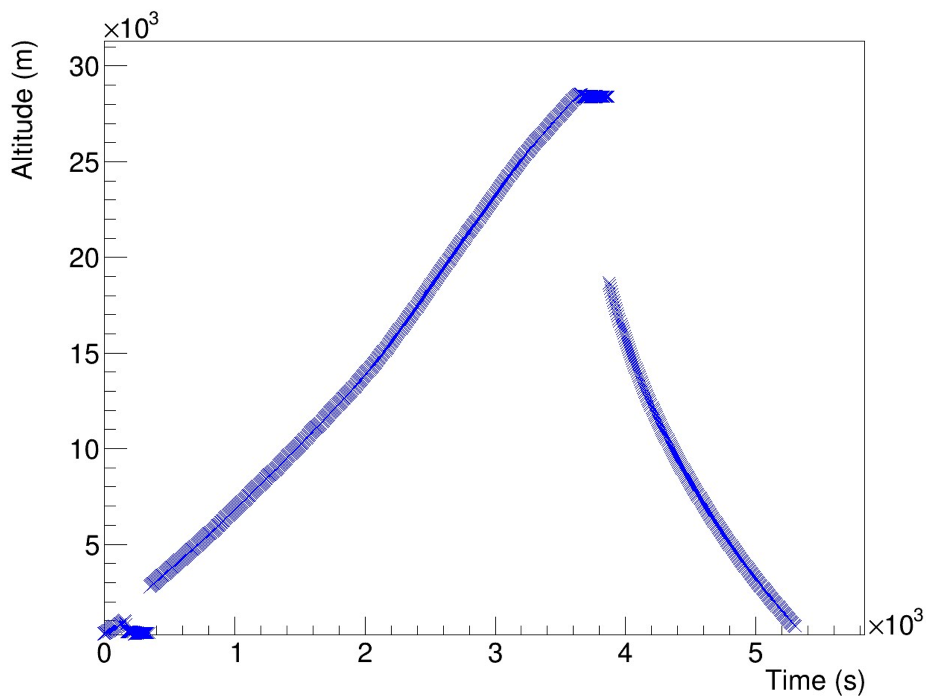
- **REFERENCES**

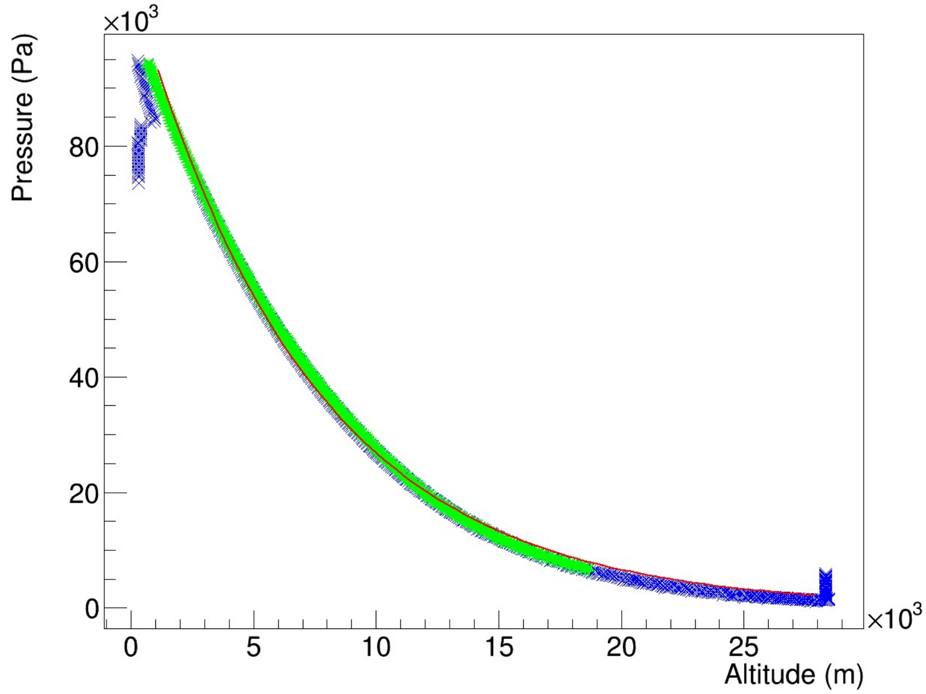
1. https://iopscience.iop.org/1674-1137/40/10/100001/media/rpp2016_0349-0428.pdf
   (C. Patrignani et al. (Particle Data Group), Chinese Physics C, 40, 100001 (2016): Review of Particle Physics -29. Cosmic Rays.)
2. https://home.cern/science/physics/cosmic-rays-particles-outer-space
3. https://web.infn.it/OCRA
4. https://agenda.infn.it/event/24503/timetable/

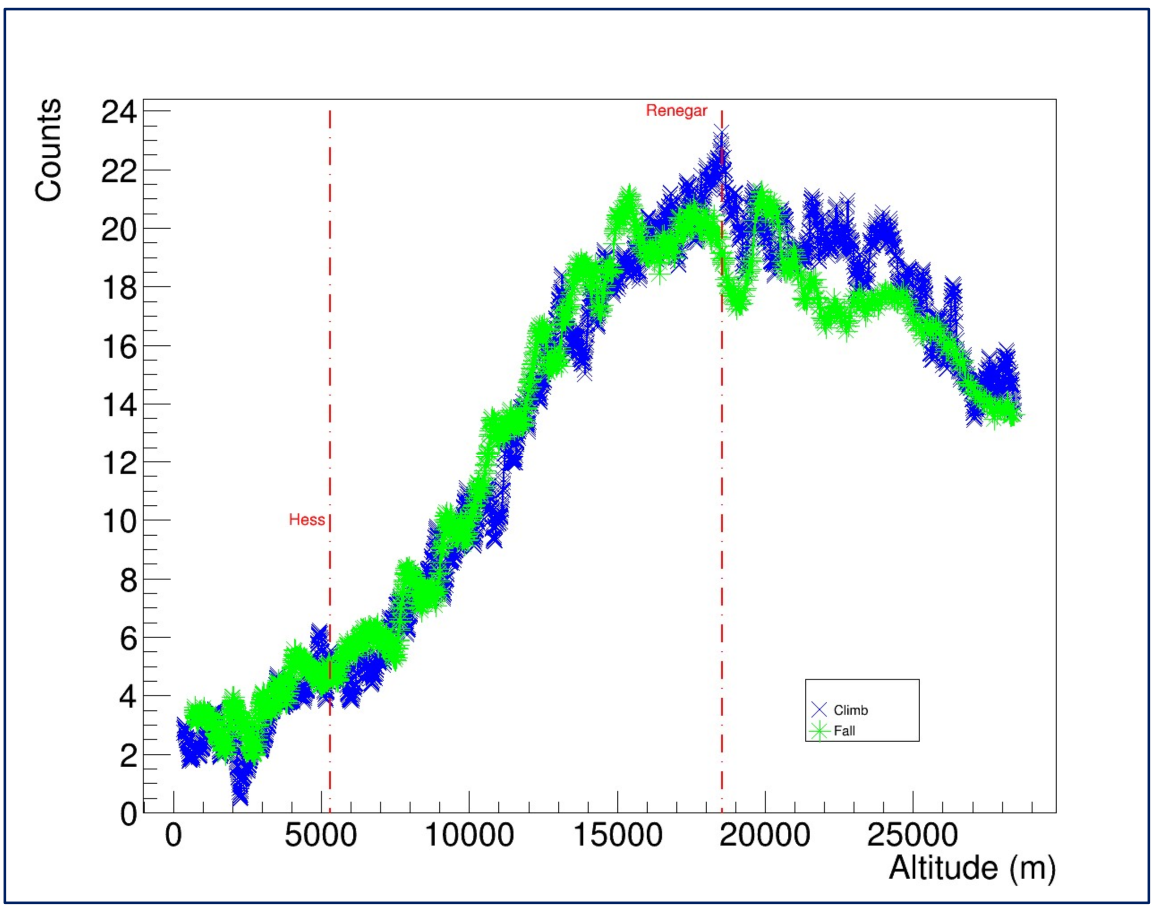
At the **Frascati National Laboratories**, I had the opportunity to measure the trend of cosmic rays at high altitudes, following the launch of a **stratospheric balloon**. The balloon was equipped with numerous sensors: pressure, altitude, temperature, two high precision GPS and two ArduSiPM (cosmic ray detector connected to an Arduino board).

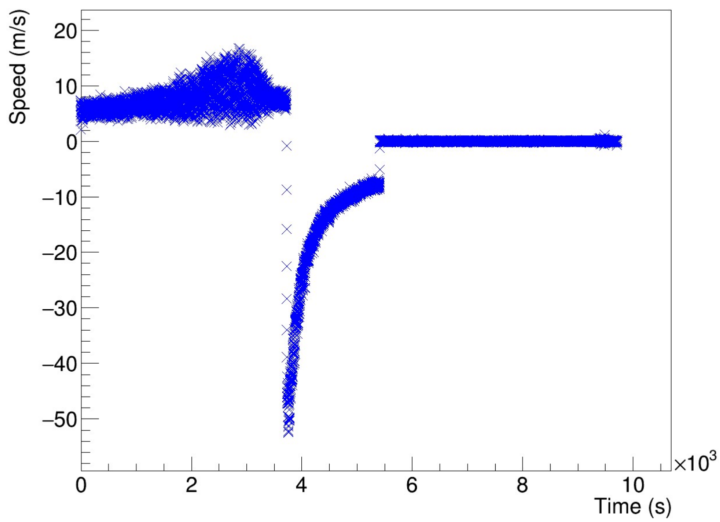
- **RESULTS STRATOSPHERIC BALLOON LAUNCH**

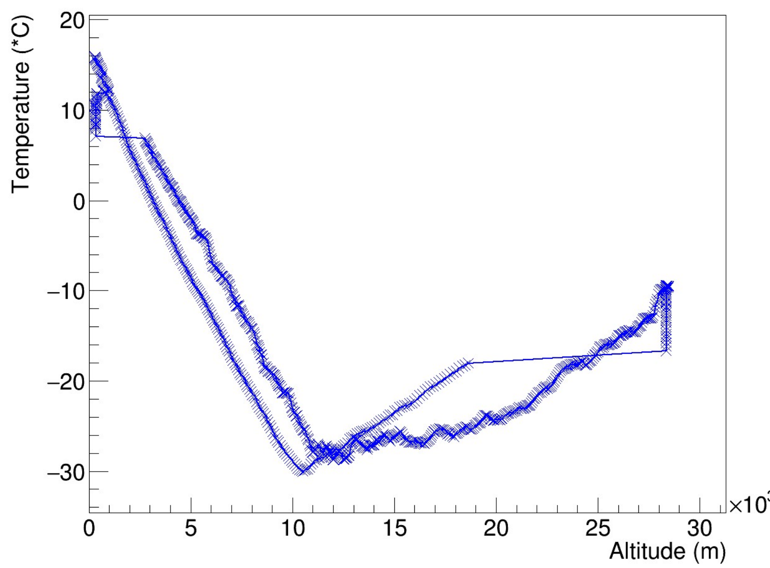


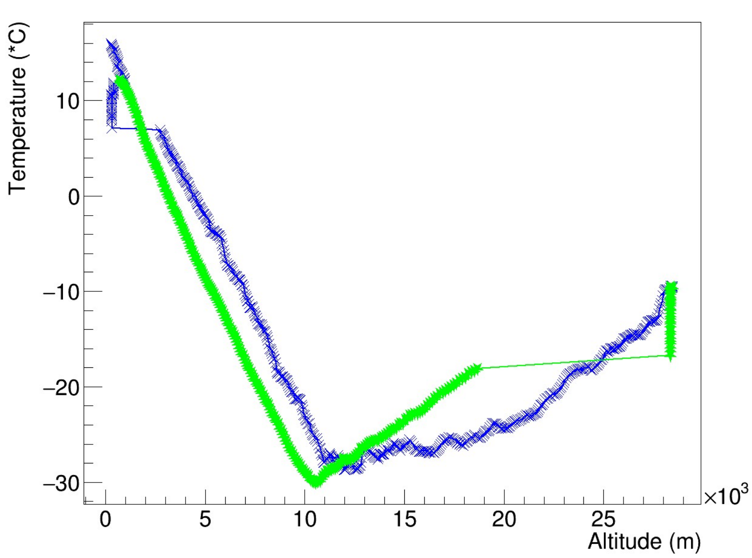



Pressure (Pa)
$\times10^3$
80
60
40
20
0
0
5
10
15
20
25
$\times10^3$
Altitude (m)


Counts
24
22
20
18
16
14
12
10
8
6
4
2
0
0
5000
10000
15000
20000
25000
Altitude (m)
Hess
Renegar
Climb
Fall


Speed (m/s)
20
10
0
−10
−20
−30
−40
−50
0
2
4
6
8
10
×10^3
Time (s)


Temperature (*C)
20
10
0
−10
−20
−30
0
5
10
15
20
25
30
×10^3
Altitude (m)


Temperature (*C)
10
0
−10
−20
−30
0
5
10
15
20
25
×10^3
Altitude (m)