

ÉCOLE POLYTECHNIQUE

PHY 584

Experimental aspects of High Energy Physics

M1 High Energy Physics

Thomas LEPLUMEY, Mathieu FEREY

Neutrinos detection at Super-Kamiokande

January-March 2023



Supervisors :

Thomas Mueller, LLR
Antoine Beauchêne, LLR

Contents

Introduction	3
1 Theoretical context	4
1.1 Neutrino oscillations	4
1.1.1 Two flavours oscillations	4
1.2 Detecting neutrinos at Super-Kamiokande	7
1.2.1 Neutrino reactions	7
1.2.2 Cherenkov radiation	8
1.2.3 Dispersion of the Cherenkov light	9
2 Data handling	10
2.1 Experimental setup	10
2.2 The Event Display	11
2.3 Discrimination criteria	12
2.4 Spherical projection	12
2.5 Direction of propagation	12
2.6 Extracting distribution	13
2.6.1 Angular distribution	13
2.6.2 Distances distribution	14
3 Particle identification	19
3.1 The RISPID score	19
3.2 Single-moment scores	20
3.2.1 Mean	20
3.2.2 Standard deviation	21
3.2.3 Skewness	22
3.2.4 Kurtosis	22
3.2.5 Summary	23
3.3 Multi-moment scores	23
3.3.1 Model	23
3.3.2 Training	24
3.3.3 Results	24
3.4 Maximum likelihood method	24
3.4.1 Principle	24
3.4.2 Probability distribution	25
3.4.3 Results	27
4 Vertex reconstruction	28
4.1 Likelihood method	28
4.2 Initialisation with heuristic	28
4.3 Results	28

Conclusion	30
APPENDIX	32
A Prediction analysis	32
B Bloopers	34

Introduction

In order to explain conservation of energy-momentum in beta decay reactions, Wolfgang Pauli introduced in 1930 a new neutral particle which he called "neutron". Of course in 1932, Chadwick discovered what we actually call "neutron", and Pauli's particle had to be renamed "neutrino" -little neutral particle- to be distinguished from the neutral much heavier nucleon. In 1956, Clyde Cowan and Frederick Reines finally managed to detect a neutrino produced in a nuclear reactor. Neutrinos therefore earned their place among the other particles of the Standard Model. There are three kinds of neutrinos, one for each flavour of charged leptons. Neutrinos are neutral leptons, and therefore only interact through weak interaction. It is quite remarkable that, even though neutrinos are produced massively in stars and our atmosphere by cosmic radiations, they remain extremely elusive particles due to how feebly they interact with matter. It is however of the utmost importance to study neutrinos. They only exist in left-handed states of helicity, thus illustrating the tendency of weak interactions to break parity. Oscillations of flavours of neutrinos indicate that they are massive particles, thus raising the question of their coupling to the Higgs field. Some candidates for dark matter are even kinds of neutrinos, or their supersymmetric counterparts. To answer these burning questions, experimentalists needed to find indirect ways to detect neutrinos despite their elusiveness. Huge detectors have been built, and some larger are under construction. Kamiokande, Super-Kamiokande and soon to be Hyper-Kamiokande are of this calibre. Takaaki Kajita earned the 2015 Nobel Prize for his work on neutrino flavour oscillation thanks to the Kamiokande detector, and Super-Kamiokande is still involved in the T2K experiment, dedicated to the measurement of parameters of the oscillation. In short, detecting neutrino flavour is at the heart of these massive international experiments, and such is the purpose of the following report.

Chapter 1

Theoretical context

To begin, before getting into the heart of the matter, some context is necessary. We will first present and explain the recent discoveries concerning the neutrino oscillation, and the associated major experiment T2K, which justify more than anything the interest of the flavor reconstruction in Super-Kamiokande. An explanation of the operation of the Super-Kamiokande detector and of the physical processes involved is also necessary, especially since a good understanding of these processes will be very useful in the implementation of data analysis techniques. Finally, we will discuss the phenomenon at the heart of our study for flavor reconstruction, the dispersion within Super-Kamiokande, which will lead us to the construction of criteria for this task.

1.1 Neutrino oscillations

We are constantly bombarded by a humongous flux of neutrinos coming from the Sun (about 6.5×10^{10} neutrinos per second per centimetre squared !). Since they barely interact inside the star they are coming straight out of its core, giving solar astrophysicists precious information about the structure of the Sun. Yet we detect far less electronic neutrinos coming from the Sun than what is predicted by our models of fusion in stars. This strange fact gave physicists the hint that neutrinos could oscillate in flavour, much like quarks or kaons.

We know neutrinos to come in three different flavour states $|\nu_\alpha\rangle$, $\alpha = e, \mu, \tau$. These states are eigenstates of the weak interaction Hamiltonian, through which they are produced or interact with other particles. We suppose that they are not eigenstates of the propagation Hamiltonian, whose eigenstates we write $|\nu_i\rangle$, $i = 1, 2, 3$, associated with eigenvalues m_i . We write U the unitary matrix mixing these different states (the Pontecorvo-Maki-Nakagawa-Sakata matrix, PMNS for short) [8]:

$$|\nu_\alpha\rangle = \sum_{i=1}^3 U_{\alpha i}^* |\nu_i\rangle \quad (1.1)$$

If the flavour eigenstates differ from the mass eigenstates, and the mass are non-zero, we expect flavour oscillations along the propagation of neutrinos.

1.1.1 Two flavours oscillations

Let us consider only two neutrino flavours for simplicity, $|\nu_e\rangle$, $|\nu_\mu\rangle$, eigenstates of the weak interaction Hamiltonian. We must also consider the propagation Hamiltonian, with eigenstates $|\nu_1\rangle$, $|\nu_2\rangle$, associated with eigenvalues m_1 , m_2 .

We write the mixing between the two basis as:

$$\begin{aligned} |\nu_e\rangle &= \cos \theta |\nu_1\rangle + \sin \theta |\nu_2\rangle \\ |\nu_\mu\rangle &= -\sin \theta |\nu_1\rangle + \cos \theta |\nu_2\rangle \end{aligned}$$

Let us derive the propagation of a state $|\psi(t)\rangle$ which starts as $|\psi(0)\rangle = |\nu_\mu\rangle$.

$$\begin{aligned} |\psi(t)\rangle &= \sum_{i=1}^2 \langle \nu_i | \psi(t) \rangle |\nu_i\rangle \\ &= \sum_{i=1}^2 e^{-iE_i t} \langle \nu_i | \psi(0) \rangle |\nu_i\rangle \\ &= -\sin \theta e^{-iE_1 L} + \cos \theta e^{-iE_2 L} \end{aligned}$$

We suppose our neutrinos to be ultra-relativistic, so that

$$E_i = \sqrt{m_i^2 + p^2} \simeq p + \frac{m_i^2}{2p^2} \Rightarrow E_2 - E_1 = \frac{m_2^2 - m_1^2}{2p} \equiv \frac{\Delta m^2}{2E}. \quad (1.2)$$

The probability to measure an electron at time $t = L$ is then:

$$P_{\nu_\mu \rightarrow \nu_e}(L) = |\langle \nu_e | \psi(L) \rangle|^2, \quad (1.3)$$

where

$$\begin{aligned} \langle \nu_e | \psi(L) \rangle &= \sum_{i=1}^2 e^{-iE_i L} \langle \nu_i | \psi(0) \rangle \langle \nu_e | \nu_i \rangle \\ &= -\sin \theta \cos \theta e^{-iE_1 L} + \cos \theta \sin \theta e^{-iE_2 L} \\ &= \cos \theta \sin \theta e^{-iE_1 L} \left(e^{-i(E_2 - E_1)L} - 1 \right) \\ &= \cos \theta \sin \theta e^{-iE_1 L} \left(e^{-i\frac{\Delta m^2}{2E}L} - 1 \right) \end{aligned}$$

Then:

$$P_{\nu_\mu \rightarrow \nu_e}(L/E) = \cos^2 \theta \sin^2 \theta \left| e^{-i\frac{\Delta m^2}{2E}L} - 1 \right|^2$$

That is :

$$\boxed{P_{\nu_\mu \rightarrow \nu_e}(L/E) = \sin^2 2\theta \sin^2 \left(\frac{\Delta m^2}{4E} L \right)} \quad (1.4)$$

It is plotted in Figure 1.1.

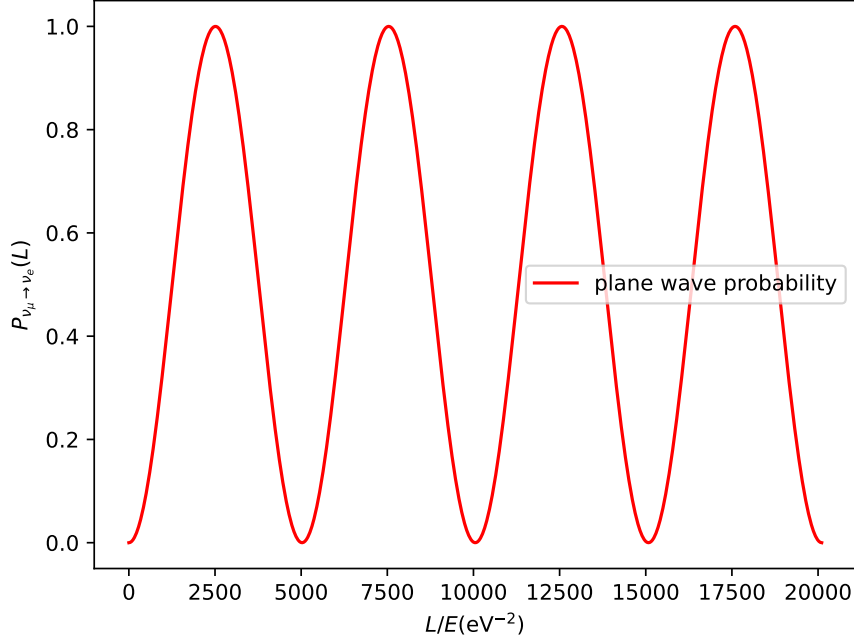


Figure 1.1: Vacuum appearance probability as a function of L/E for $\Delta m^2 = 2.5 \text{ meV}^2$

We can derive the disappearance probability as:

$$P_{\nu_\mu \rightarrow \nu_\mu}(L/E) = 1 - \sin^2 2\theta \sin^2 \left(\frac{\Delta m^2}{4E} L \right) \quad (1.5)$$

The oscillation length L_0 is then defined as the smallest L such that $\sin^2 \left(\frac{\Delta m^2}{4E} L \right) = 0$, that is:

$$L_0 = \frac{4\pi E}{\Delta m^2}$$

The approach taken here is not really helpful, for neutrinos come with a certain distribution of energy. We must therefore average our probability over this distribution of energy.

A good model for an energy distribution is a Gaussian distribution.

$$\rho(E) = \frac{1}{\sqrt{2\pi}\sigma} \exp \left(-\frac{(E - E_{\text{mean}})^2}{2\sigma^2} \right) \quad (1.6)$$

We can then compute the probability averaged over the energy in function of L :

$$\langle P_{\nu_\mu \rightarrow \nu_e}(L/E_{\text{mean}}) \rangle = \int_{-\infty}^{\infty} dE \rho(E) P_{\nu_\mu \rightarrow \nu_e}(L/E) \quad (1.7)$$

The $\sin(1/E)$ can be troublesome at 0. We pick values of E_{mean} and σ such that the peak of the Gaussian is located far from 0, and we can restrict the integral to an interval of a few σ 's around the mean of the Gaussian. Python can easily take care of this integral for us. The plot is given in Figure 1.2.

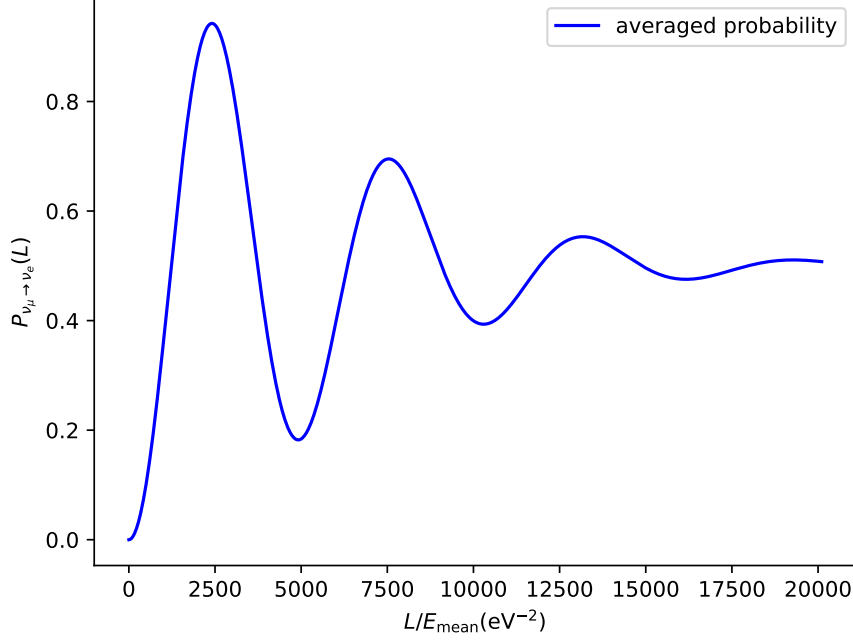


Figure 1.2: Averaged vacuum appearance probability as a function of L/E_{mean} for $\Delta m^2 = 2.5 \text{ meV}^2$, $E_{\text{mean}} = 600 \text{ MeV}$, $\sigma_E = 0.15 E_{\text{mean}}$

Converting the maximum of the averaged appearance probability to SI units, we have $L_{\text{max}} \simeq 300 \text{ km}$, which is roughly the distance between the J-PARC accelerator, where a neutrino beam is produced, and Super-Kamiokande, where the neutrinos are detected, in the T2K experiment.

1.2 Detecting neutrinos at Super-Kamiokande

Super-Kamiokande is cylindrical tank of about 40 meters in diameter times 40 meters in height, originally filled with ultra-pure water (gadolinium has been added recently to improve the detection of some particular events), whose walls are covered with photo-multipliers (PMTs for short).

1.2.1 Neutrino reactions

Neutrinos interact with the nucleons of water in the following channels, depending on the energy of the incoming neutrino [4]:

- Coherent Scattering: at low energy the neutrino just bounces off the nucleus.
- Quasi-Elastic Scattering (QE): at higher energies, the neutrino might be energetic enough to penetrate the nucleus and interact with the nucleons. There are two channels, interaction through charged current (CC) and neutral current (NC).

$$\begin{aligned}
 \text{(CC)} \quad & \left\{ \begin{array}{l} \bar{\nu}_l + p \rightarrow l^+ + n \\ \nu_l + n \rightarrow l^- + p \end{array} \right. \\
 \text{(NC)} \quad & \nu_l + p(n) \rightarrow \nu_l + p(n)
 \end{aligned}$$

The nucleon can be promoted to an excited state (Δ of different charges) by resonance (RES), which will eventually decay into pions. Charged pions can easily be confused with charged leptons in the detector, and must therefore be treated carefully.

- Deep Inelastic Scattering (DIS): at the highest energies, the neutrino breaks the nucleon and we end up with a hadronic shower in the final state, as well as a charged lepton.

The cross sections for these channels, highly depending on the energy of the incoming neutrino, are summarised in Figure 1.3.

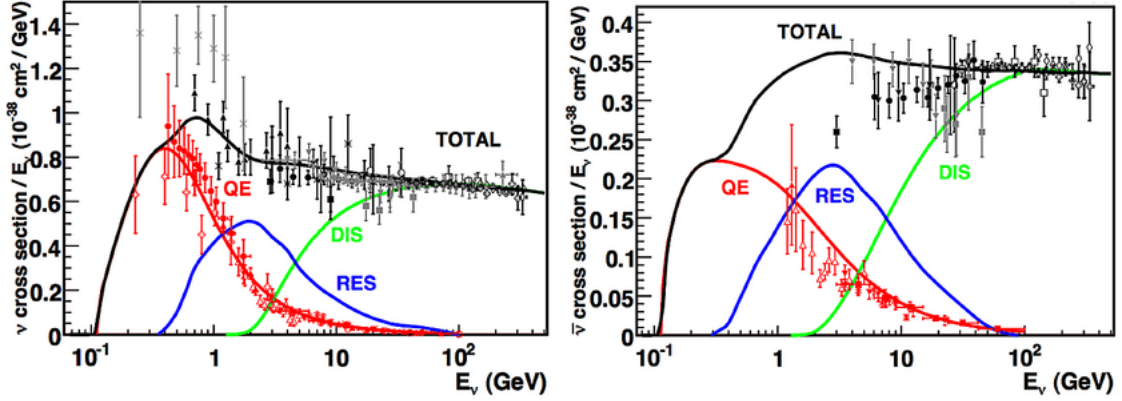


Figure 1.3: Cross-sections for the interaction of neutrinos and anti-neutrinos with nucleons as a function of the energy of the incoming neutrino [5].

This project will focus on the CC channel of quasi-elastic scattering, which we often call Inverse Beta Decay (IBD). One can indeed notice that the final states depends on the flavour of the incoming neutrino, and are therefore relevant for oscillation experiment for instance. Furthermore, a charged lepton (muon or electron) is far less elusive than an neutrino, that is to say much easier to detect.

1.2.2 Cherenkov radiation

If the neutrino is energetic enough, the produced lepton can go faster than light in the medium of interaction, henceforth emitting a Cherenkov cone along its direction of propagation (see diagram of Figure 1.4).

The angle θ_c of the cone emitted by a particle of speed β in a medium of refraction index n is

$$\cos \theta_c = \frac{1}{n\beta} \quad (1.8)$$

In Super-Kamiokande ($n = 1.33$, pure water), for a particle of speed c ($\beta = 1$), $\theta_c \simeq 0.72 \text{ rad} \simeq 41^\circ$.

As the particle propagates in water, it loses energy, β decreases, $\cos \theta_c$ increases and therefore θ_c decreases. Once the particle is slowed to a speed lower than c/n , it no longer emits Cherenkov radiation. We therefore expect to see in our detector a ring, whose outer edge corresponds to a the emission of Cherenkov when the charged lepton is created at the interaction point, an inner edge to when the lepton stopped emitting light.

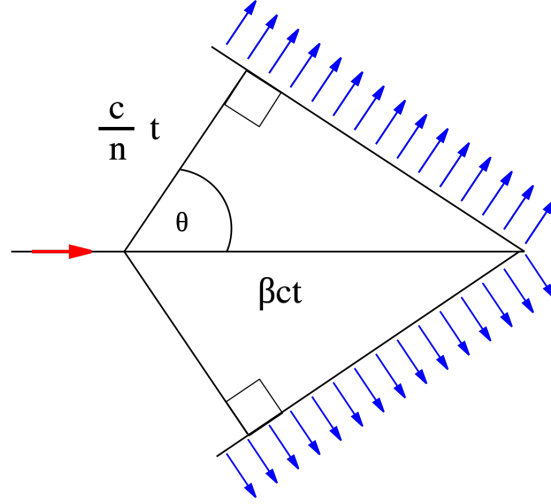


Figure 1.4: Cherenkov effect [12].

1.2.3 Dispersion of the Cherenkov light

However, in a normal optical material, the refractive index n is a function of the wavelength λ , or equivalently, of the frequency ω . The angle of the Cherenkov cone is therefore still written :

$$\cos \theta(\omega) = \frac{1}{\beta n(\omega)}$$

Which can be also expressed around a given ω_0 as [7] :

$$\tan \theta(\omega) = \tan \theta_c + \left(\frac{dn}{d\omega} \right)_0 \frac{\omega_0 \beta}{\sin \theta_c}$$

On the other hand, the Frank-Tamm formula [10] allows to express the energy emitted by Cherenkov light by unit of distance and frequency :

$$\frac{d^2 E}{dx d\omega} = \frac{e^2}{4\pi} \mu(\omega) \omega \left(1 - \frac{1}{\beta^2 n^2(\omega)} \right) = \frac{e^2}{4\pi} \mu(\omega) \omega \sin^2 \theta_c(\omega)$$

where μ is the permeability of the medium, and the formula is correct only for $\beta > \frac{1}{n(\omega)}$ by definition of Cherenkov radiation.

Integration over ω gives :

$$\frac{dE}{dx} = \frac{e^2}{4\pi} \int_{n(\omega) > \frac{1}{\beta}} \mu(\omega) \omega \sin \theta_c(\omega)$$

Therefore, we see that the larger β is, the more energy is emitted per unit length, and the greater the energy ω of the emitted photons, and with that, the more the dispersion angle θ deviates from the Cherenkov angle θ_c .

Finally, since positrons are lighter than muons, at the same energy, positrons are faster, and so we expect that Cherenkov light emitted by positrons is more scattered than that emitted by muons.

Chapter 2

Data handling

In this chapter, we present all the preliminary work on the event data to make them usable for flavor reconstruction. We present our experimental setup, including the nature of the raw event data, and then the whole process of transforming these data, based on the theoretical background developed in the previous chapter, especially the Cherenkov effect. This process ends with the extraction of a few distributions, which we hope will contain as explicitly and exhaustively as possible the information necessary for flavor reconstruction.

2.1 Experimental setup

In all our work, we use two datasets of one thousand Super-Kamiokande events (one of muonic events and the other of positronic events). These datasets were simulated with WCSIM water Cherenkov simulation package [11] which employs the Geant4 simulation framework [3] [1] [2]. All the events are noised, and occur in an energy range of 200 keV to 1 GeV. As some of our methods use machine learning methods, we have split these datasets into two new sets :

- A training set, composed of 500 positronic events and 500 muonic events
- A test set, composed of the remaining 500 positronic events and 500 muonic events. When necessary, 20% of the test set is used as validation test during the training phase.

The description of one event is given as follows. For each hit :

- The index of the PMT that was hit (from which we can deduce the spatial position of the PMT)
- The charge received by the PMT
- The time of the hit

To these informations is added the position of the interaction vertex.

All our experiments have been done with *Python*.

- *plotly* library was used for all 2D and 3D plots
- *numpy* library was used for all simple calculus and linear algebra
- *pytorch* library was used for machine learning and optimization methods
- *sklearn* library was used for distribution fitting

2.2 The Event Display

The raw data of one event consists of the positions of the PMTs that detected light on the 3D cylinder in Cartesian coordinates, the time at which they first detected a photon and the total charge they detected for the duration of the event. It is fairly easy to display these PMTs using *plotly*. It may be more relevant (especially for this report, where 3D figures cannot be moved) to unfold the cylinder in two dimensions, to get a better look at the event. This is furthermore how the control panel for SK looks like for remote shifts or surveillance.

We display a muonic and electronic event in figures 2.1, 2.2.

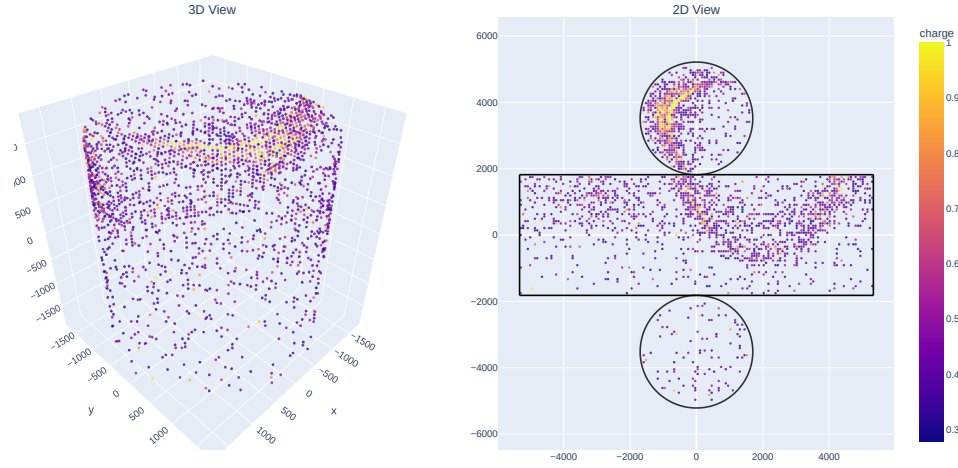


Figure 2.1: Positron event.

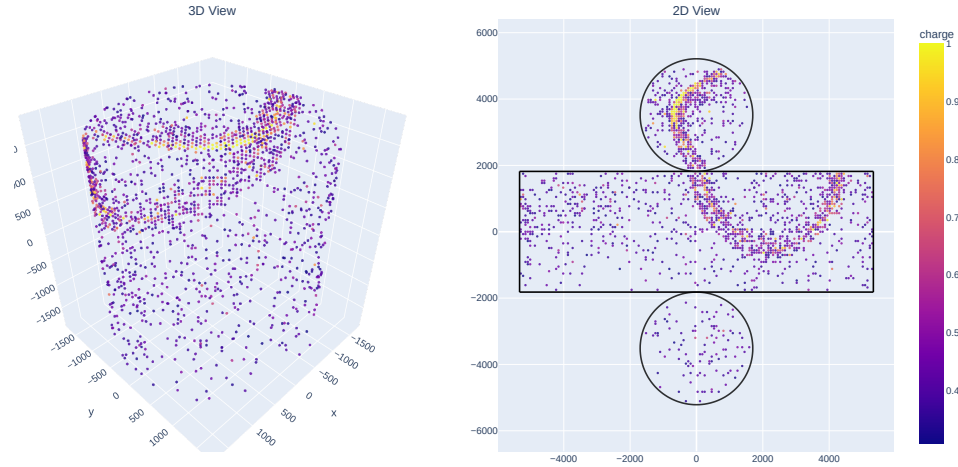


Figure 2.2: Anti-muon event.

2.3 Discrimination criteria

We want to determine whether an event in SK was induced by an anti-muon or by a positron, given the position of the PMTs which received light, the time at which they received light and the amount a light (the charge) each PMT received during the whole event. A quick look at some events (for instances figures 2.1 and 2.2) gives us some hints about how to discriminate between the two: positron rings seem "messier" than the muon ones, which is to be related to the dispersion of the emitted radiation, which is theoretically more important for positrons than for muons (see Section 1.2.3).

One sure is certain, we want to compute parameters of the charge distribution, in a way that this effect becomes significant and analysable.

2.4 Spherical projection

The 3D cylinder and its two dimensional unfolded form have the problematic inconvenient of non trivial symmetries for our problem. It would be much more convenient to work on a sphere, centred at the interaction point and of radius one. This is fairly easy to do: for each PMT on the 3D cylinder, define the vector between its position on the cylinder and the interaction point, and then divide this vector by its norm. You can see the "spherized" event in Figure 2.3. We also define a way to go back from the sphere to the 3D cylinder, which will be useful later.

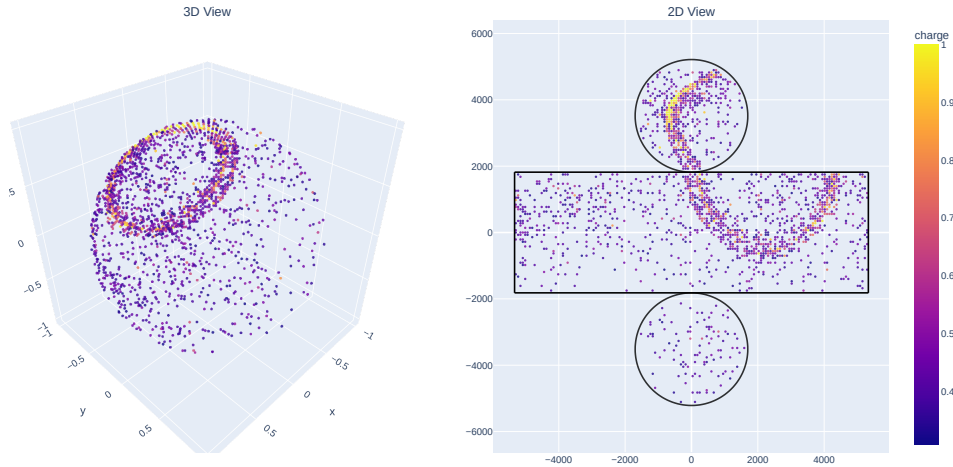


Figure 2.3: Muon event on the 3D sphere (same event as Figure 2.2).

2.5 Direction of propagation

We now need to find the centre of our ring, which corresponds to the direction of propagation of the produced lepton emitting the Cherenkov radiation. Working on the sphere, it is tempting to find the average position of all of the lit PMTs. The problem comes from the sea of lit PMTs which are not on the ring, which are either the result of the dispersion of the photons or the noise of the detectors. Since there are more PMTs outside of the ring than inside, the noise shifts the position average away from the actual centre of the ring. This problem can be solved by averaging only over the ring, where the density of PMT is higher than elsewhere in the detector. We call this operation "denoising", which is easily achieved by counting for each PMT

its number of neighbours in a small spherical mask around it. If this number is lower than an arbitrary threshold, we discard this PMT. The average point is then divided by its norm to be moved on the sphere. The denoised event in the three projection spaces is displayed in Figure 2.4

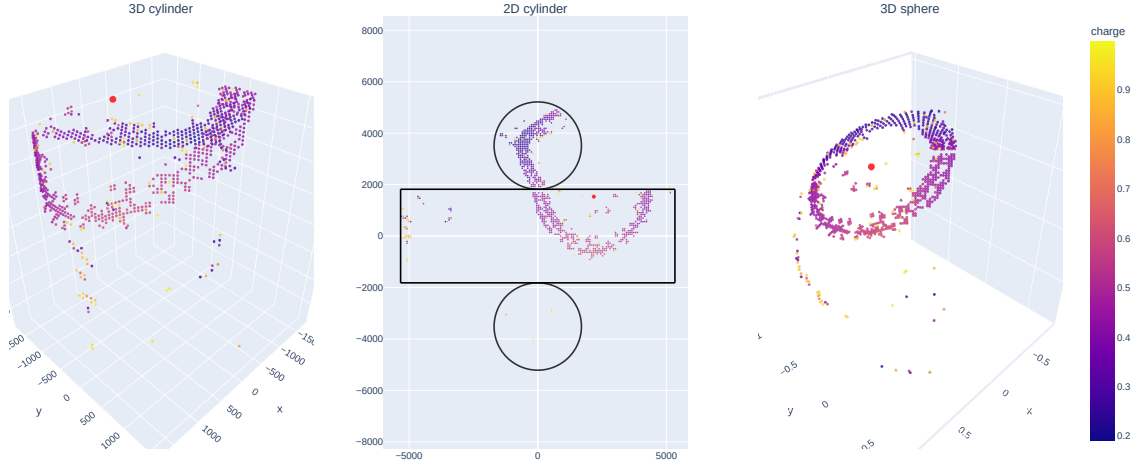


Figure 2.4: Denoised muon event (in red the centre of the ring).

2.6 Extracting distribution

Our qualitative look on the two events of Figure 2.1 and 2.2 gave us the intuition that muons and positrons differ by the shape of their Cherenkov ring. Furthermore, our understanding of Cherenkov radiation (see Subsection 1.2.2) is that the muon ring should be inside the circle defined by the Cherenkov cone emitted at the interaction point by the muon going at c , which we will call "first Cherenkov circle". We therefore want to know how the charge of the ring is dispatched around this first Cherenkov circle. We achieve this in two ways, an easy one and a more complicated one.

2.6.1 Angular distribution

We know exactly the angle on the associated cone θ_{Ch} . We can easily display this circle on the sphere around the north pole: it has spherical coordinates θ_{Ch} and $\varphi \in [0; 2\pi[$. We then need to know the spherical coordinates of the PMTs whit the centre of the ring (which we computed in the previous section) at the north pole. A bit of geometry gets us the desired angles θ and φ with respect to the direction of propagation. The event on the sphere rotated around the north pole is displayed in Figure 2.5.

We can then "unfold" the ring onto a line by plotting $\theta - \theta_{Ch}$ (to have the first circle as the reference) in function of φ . We obtain, for each phi, the distribution of charge in function of $\theta - \theta_{Ch}$. We finally average these distributions over all φ to get what we call the angular distribution of the charge around the ring. This is plotted in Figure 2.6 for the muon event. As a comparison, see in Figure 2.7 the angular distribution for the positron event of Figure 2.1.

This method has one issue: the PMTs on the hats of the cylinder seem distorted on the distribution, which might jeopardise the distinction between positron and muon rings by altering

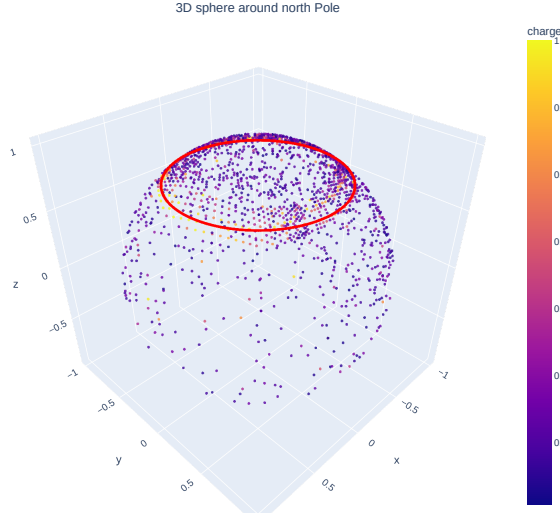


Figure 2.5: The muon event on the sphere around the north pole.

the shape of the mean distribution. This deformation seems induced by our projection on the sphere, so let us work directly on the cylinder.

2.6.2 Distances distribution

Similarly as before, we want to unfold the ring so as to obtain a distribution of charge for each $\varphi \in [0; 2\pi[$ and average over it. For each radius (that is to say a given φ), instead of displaying the angle θ of the PMTs along the radius, which induces a deformation, we want to display the distance of the PMTs to the first Cherenkov circle on the 3D cylinder. This should rid us of the artefacts of the spherical projection. The problem is that defining radii of the ring on the raw event directly (on Figure 2.8 for the muon and Figure 2.9 for the positron) is not a geometrical triviality. We have to start with our ring at the north pole of the sphere (Figure 2.5), where it is easy to define radii. We decompose our ring into N small pieces of angle $\varepsilon = 2\pi/N$ (not too big to get thin enough radii, not too small to get enough PMTs on each radius). We then basically modify the way we stock our PMTs in the code: instead of having one big array of all our PMTs, we create N smaller arrays where the array number $n \in \llbracket 0; N-1 \rrbracket$ contains all PMTs with $\varphi \in [n\varepsilon; (n+1)\varepsilon[$.

We then need to rotate every coordinates with adequate 3D rotation matrices so that the north pole is brought back to the actual direction of propagation. We can finally go back to the 3D cylinder with the inverse spherical projection. We have managed to rearrange the way we stock PMTs in our code to decompose our ring into radii on the cylinder. This is displayed in Figure 2.10.

We can now easily compute for each radius the distance of the PMTs to the first Cherenkov circle to obtain the distances distribution of the charge around the ring, then average over the ring to have the mean distances distribution of charge of the event (see Figure 2.11 for the muon event). As a comparison, see in Figure 2.12 the distances distribution for the positron event of Figure 2.1. In comparison with the previous angular distributions, we see that we managed to get rid of the possibly problematic deformation on the hats of the cylinder.

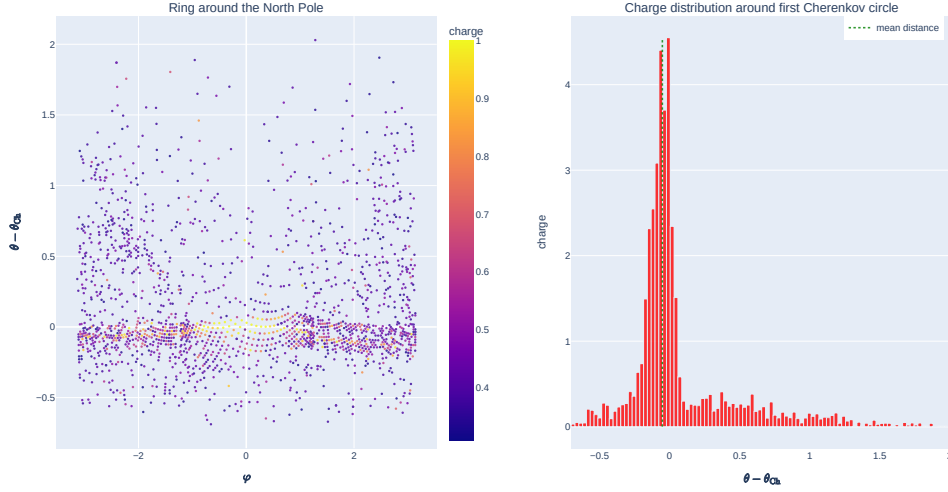


Figure 2.6: Angular distribution for the muon event.

Looking at the distributions obtained with the two methods, it seems like both yield similar qualitative differences between anti-muons and positrons: the latter have wider distribution, with a more important tail outside of the ring, in accordance with our qualitative understanding the Cherenkov radiation in Subsection 1.2.2. The mean of the distribution is computed with an arbitrary cut: we only wish to know the mean of the peak, without the tails. It is indeed relevant to know the position of the ring with respect to the first Cherenkov circle, we then do not want to be influenced by the diffusion and noise outside of the ring (always more important than inside), which will inevitably shift our mean. Compare for example the rings of the 2D cylinder for the anti-muon Figure 2.8 and the positron Figure 2.9, the muon ring fits well inside the first Cherenkov circle whereas the positron sits on both sides. Of course, taking the tail into account in the calculation of the mean of the distribution can also be relevant: since positrons display more diffusion than anti-muons, we expect the mean to be more shifted toward the outside of the ring. We can anyway test all of these parameters (width of the distribution, tails, mean with our without the tails) to discriminate between the two flavours, and see which yields the best results. This is the purpose of the next chapter.

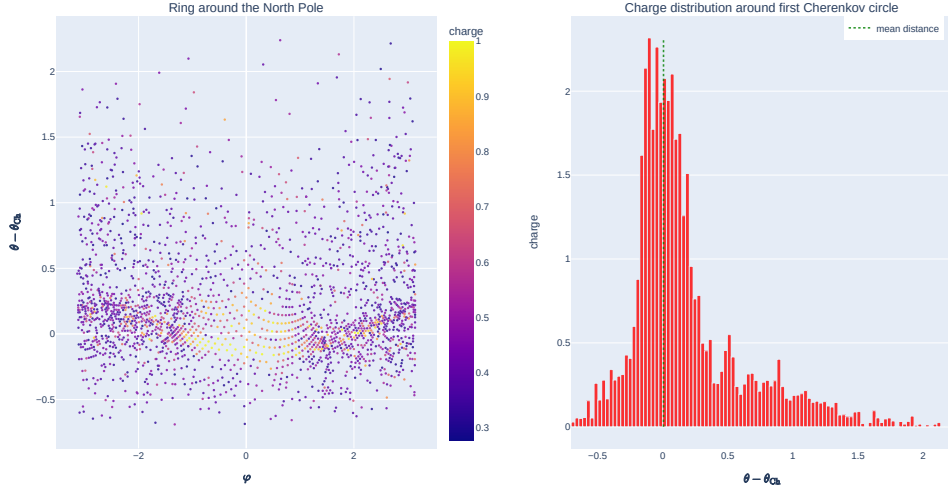


Figure 2.7: Angular distribution for the positron event.

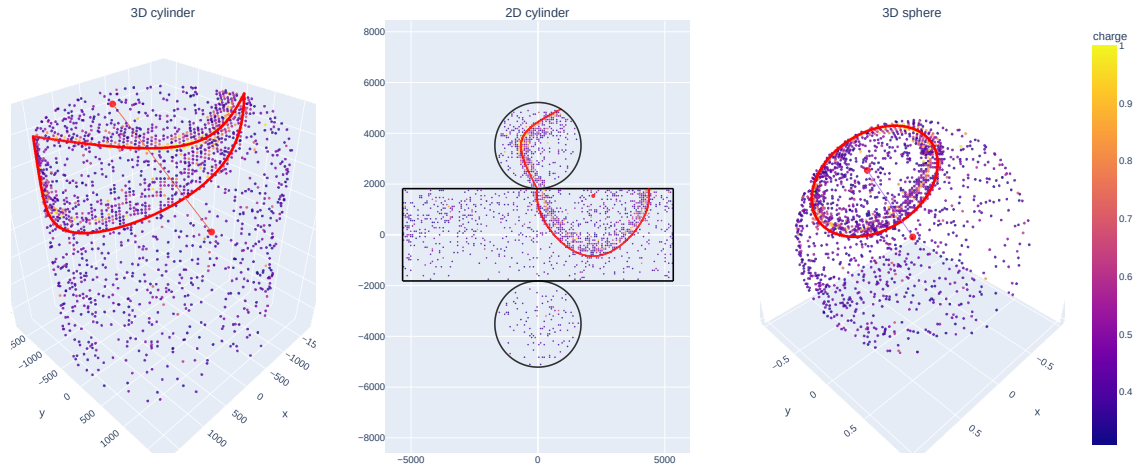


Figure 2.8: The muon event and the first Cherenkov circle (in red). The two red points are the interaction point and the centre of the ring.

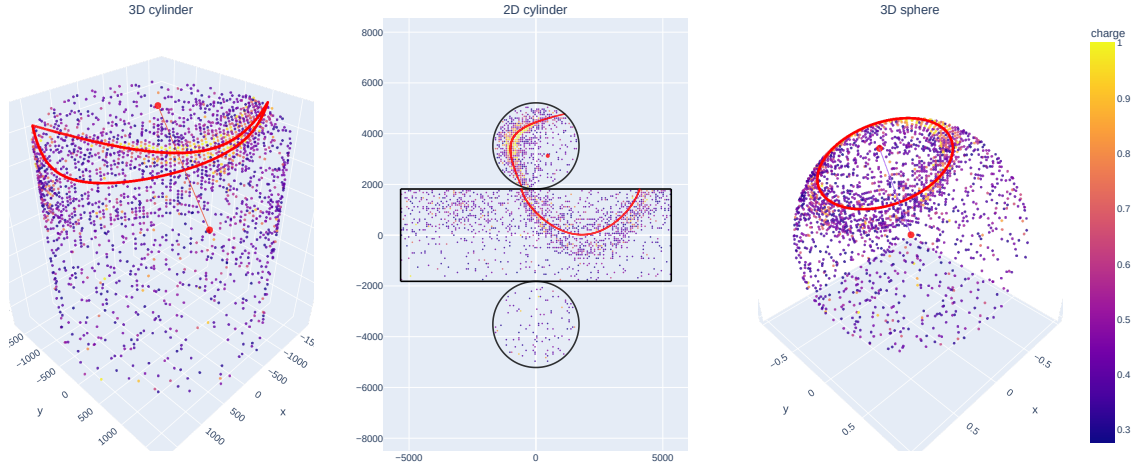


Figure 2.9: The positron event and the first Cherenkov circle (in red). The two red points are the interaction point and the centre of the ring.

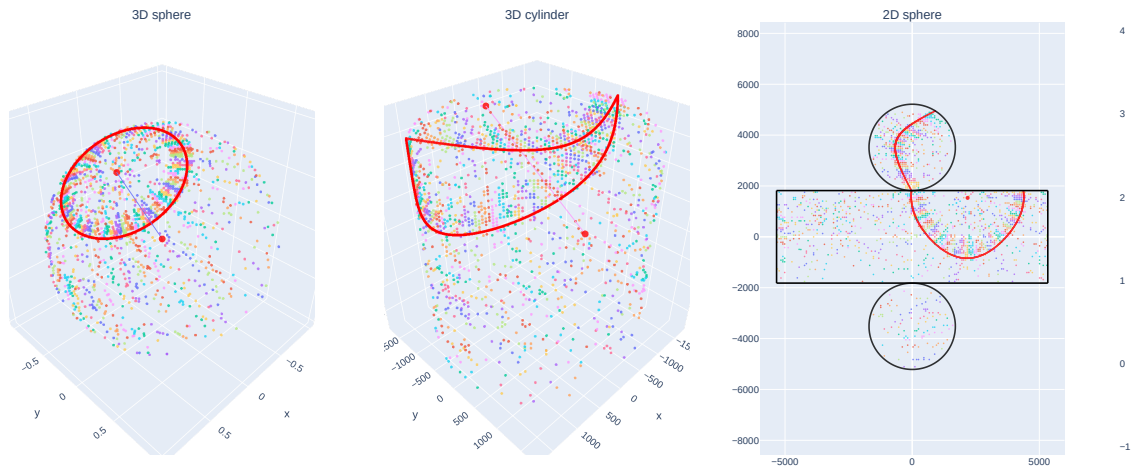


Figure 2.10: Rearranged muon event. Each colour is a different radius along which we will compute a distance distribution.

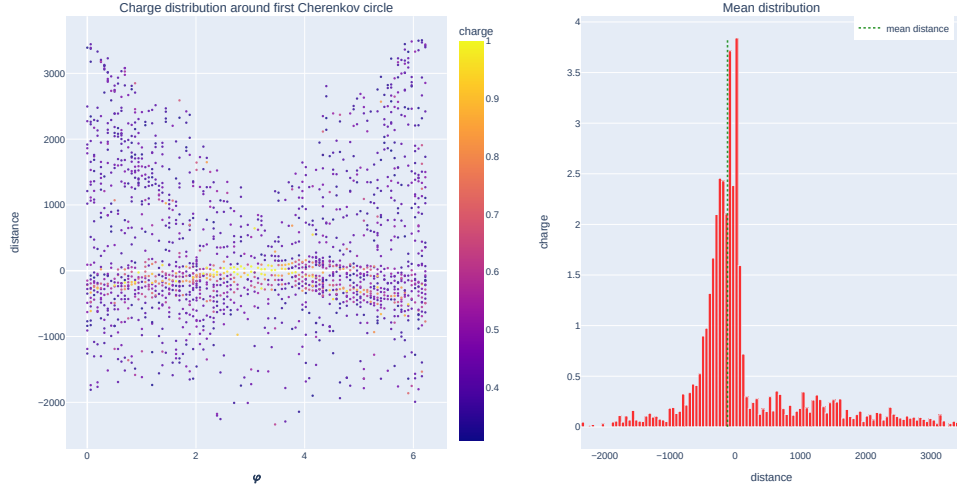


Figure 2.11: Distances distribution for the muonic event.

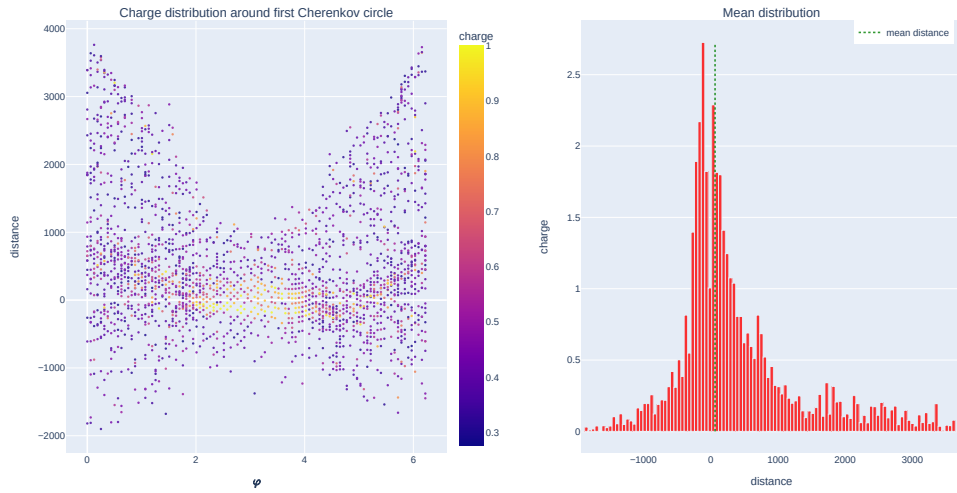


Figure 2.12: Distances distribution for the positron event.

Chapter 3

Particle identification

We now arrive at the heart of our work, that is the multiple techniques implemented to discriminate between the two flavors (electronic or muonic) for the events. Concretely, this is done by extracting scores associated to the events, which should separate the events into two groups distinguishable by this score. The evaluation of these different scores and the determination of the thresholds to be adopted is carried out thanks to standard techniques presented in Appendix A. We will essentially use the equal error rate to evaluate the performance of these scores, but different metrics could be used in other contexts.

3.1 The RISPID score

The first method we considered to discriminate electronic and muonic events is based on the RISPID score, described in the paper [6]. The method as described in the paper consists in projecting all activated PMTs on the propagation line of the particle, according to the Cherenkov cone angle at 41° . The score is then calculated according to the following formula:

$$\text{RISPID} = \frac{\int_{x_0-\delta}^{x_0} c}{\int_{x_0}^{x_0+\delta} c}$$

where the integration is performed on the propagation line, and :

- x_0 is the point where the charge distribution on the line is sharpest, that is when $\left| \frac{d^2 c}{d\lambda^2} \right|$ is maximal (λ being the affine coordinate on that line)
- δ is the minimum between the range available around x_0 inside the cylinder, and 5 meters.

We have slightly modified this method, in the following ways :

- In the paper [6], the interaction point is assumed unknown, whereas in our case, we know the interaction point. Therefore, we take x_0 equal to the interaction point instead of the sharpest point on the line
- As it gave much better results in practice, we use $\delta = +\infty$ in all cases.

The intuition we developed behind this score is the following: in an ideal case where there is no dispersion, all the PMTs actived should be inside the 41° Cherenkov cone issued from the interaction point, which would give $\text{RISPID} = 0$. But in presence of dispersion, we expect that the more dispersion there is, the greater the proportion of PMTs activated outside this cone, and the greater the RISPID. Therefore, the positrons events should have a greater RISPID than the muon events. The Figures 2.8 and 2.9 show a muon and a positron event respectively, with the 41° Cherenkov cone identified in red. It is visible that in both cases, most relevant PMTs are

inside the cone, and in greater proportion for the muon event.

The performance of this score on our test set is shown in Figure 3.1. The equal error rate obtained for this score is 0.089, which corresponds to an accuracy of 91.1% in the task of discriminating positrons and muons events.

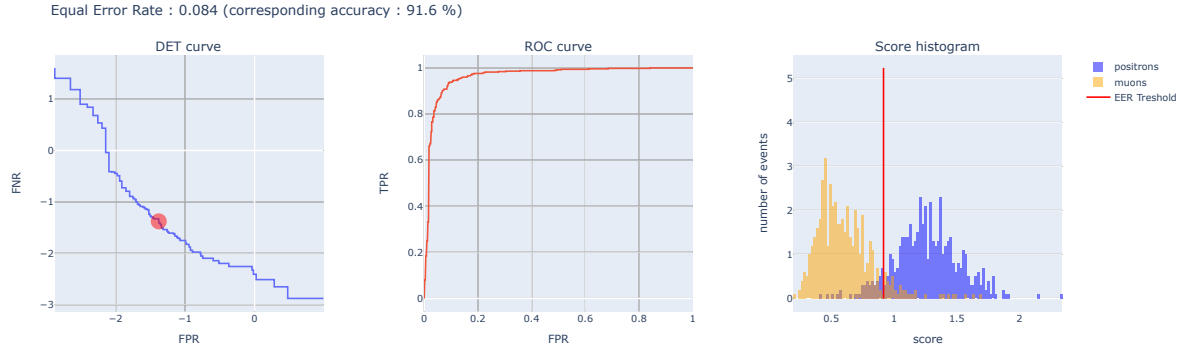


Figure 3.1: Prediction for the (modified) RISPID score.

3.2 Single-moment scores

In a second step, we decided to explore the discriminative power of the moments of the extracted distribution. More precisely, we tried to use various moments of the distribution as scores, which we can then evaluate with the usual methods.

3.2.1 Mean

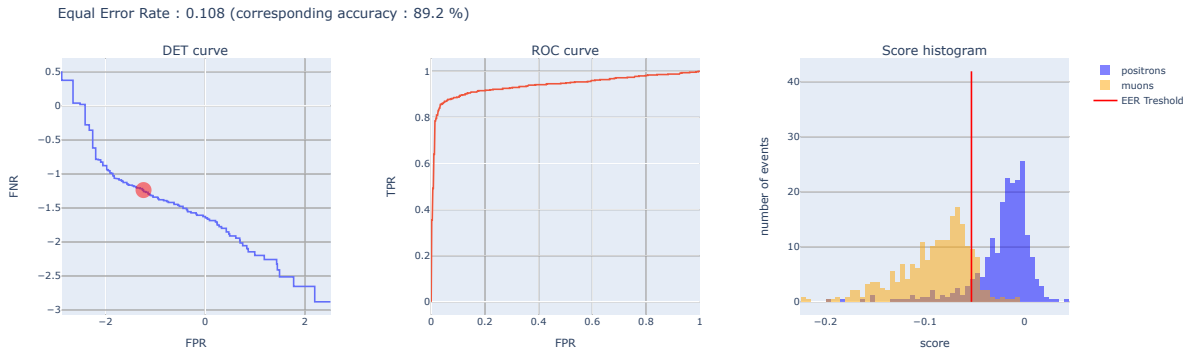


Figure 3.2: Prediction for the mean on the angular distribution

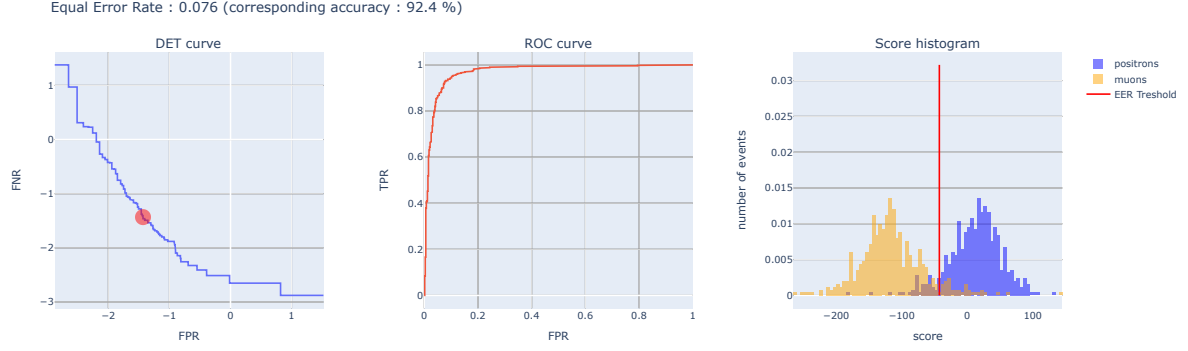


Figure 3.3: Prediction for the mean on the distances distribution

Among the four moments we tested, the mean is the one that gives the best discriminating power. The results are presented in Figures 3.2 and 3.3. The equal error rate is 0.108 using the angular distribution, and 0.076 using the distances distribution.

Our interpretation of this result is that because positronic events exhibit more dispersion, the mean of their distribution is pulled slightly more toward the center of the range. Conversely, as muonic events are less scattered, the mean of their distribution remains closer to the theoretical circle at 41° .

In addition, observe that the distances distribution gives indeed better results than the angular distribution. Unfortunately, this is not the case for the other moments (as seen in table 3.1). Therefore, for the other moments, we will present the detailed results only for the angular distribution, and the performances of all the moments and extracted distributions are presented in 3.1

3.2.2 Standard deviation

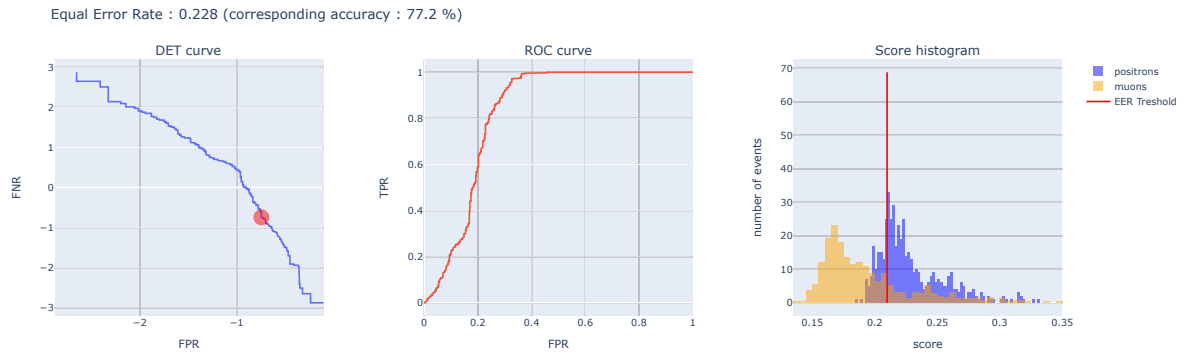


Figure 3.4: Prediction for the standard-deviation on the angular distribution

The interpretation here is that the angular distribution for positrons events is wider, as there is more dispersion, and therefore the standard-deviation is larger. However, this gives a really poor performance, in particular, much poorer than the mean.

3.2.3 Skewness

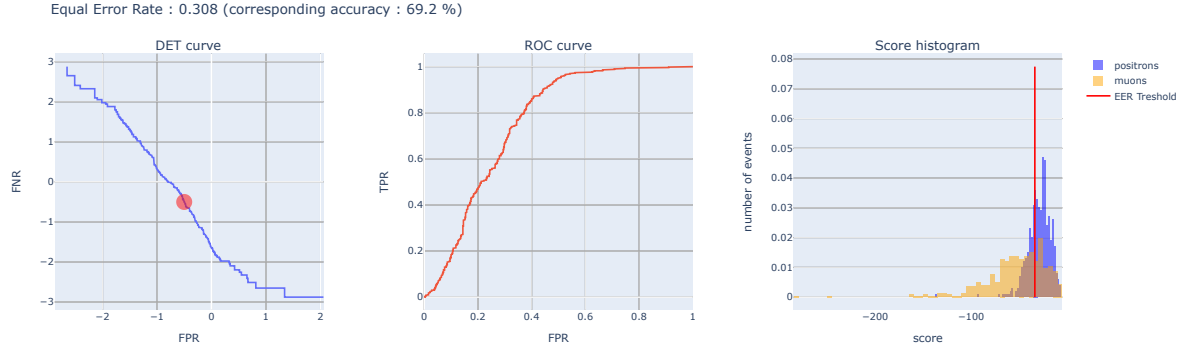


Figure 3.5: Prediction for the negative skewness on the angular distribution

We used here the negative skewness as a score, given that positrons events show a smaller skewness than muon events. This means that the positronic distribution is more shifted on the right of the median than the muonic distribution. Again, this may be explained by the fact that for positronic events, there is more dispersion, and therefore more PMT actived on the right of the distribution (that is, for large θ).

3.2.4 Kurtosis

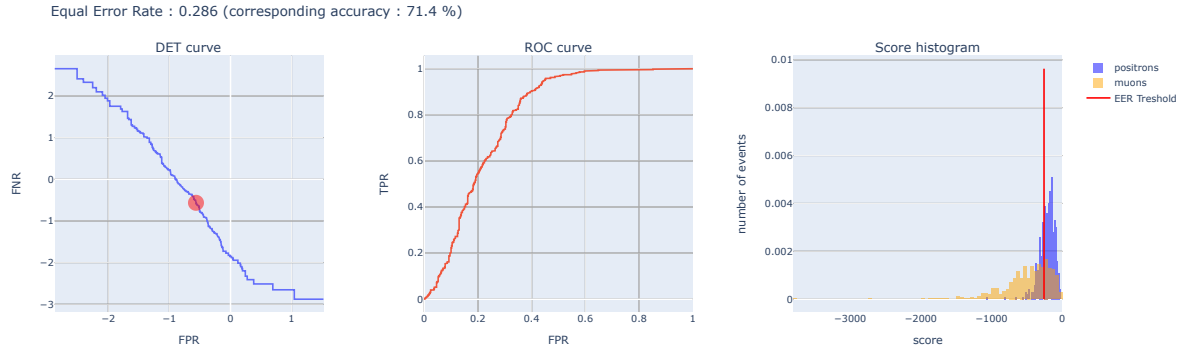


Figure 3.6: Prediction for the kurtosis on the angular distribution

Again, we used here the negative kurtosis as a score, given that positrons events show a smaller kurtosis than muon events. This means that the positronic distribution has smaller tails and is less sharp than the muonic distribution. This is a bit conterintuive. Indeed, the positrons events showing more dispersion, we expected larger tails in the positronic distributions. However, we didn't take time to investigate that point further.

3.2.5 Summary

	mean	std	skewness	kurtosis
Angular distribution	89.2%	77.2%	69.3%	71.4%
Distances distribution	92.4%	70.8%	63.6%	66.4%

Table 3.1: Accuracies for different moments as input of the model, evaluated on the test set, with the angular distribution and the distance distribution

3.3 Multi-moment scores

After studying the different moments independently, we tried to cumulate several moments at once to take advantage of the discriminating powers of each of these moments. The problem is that now, instead of having a single score on which we can place a threshold, we have a multidimensional space (2, 3 and 4 in the following examples). We must therefore build a model that takes these different moments as input and constructs a score, a function of these moments, that best discriminates between positronic and muonic events. As we do not have an objective method to combine these different moments, we have opted for a machine learning method.

3.3.1 Model

The model we have chosen to implement is a neural network with two fully connected layers. The input is composed of 2, 3, 4 or 5 (depending on the number of moments considered) features, the hidden layer is of dimension 64 and the output is a real number normalized between 0 and 1 using a sigmoid. The activation functions used after the hidden layer are Regular Linear Units (ReLU) defined by $\text{ReLU}(x) = \max(0, x)$. The described structure is schematized in Figure 3.7.

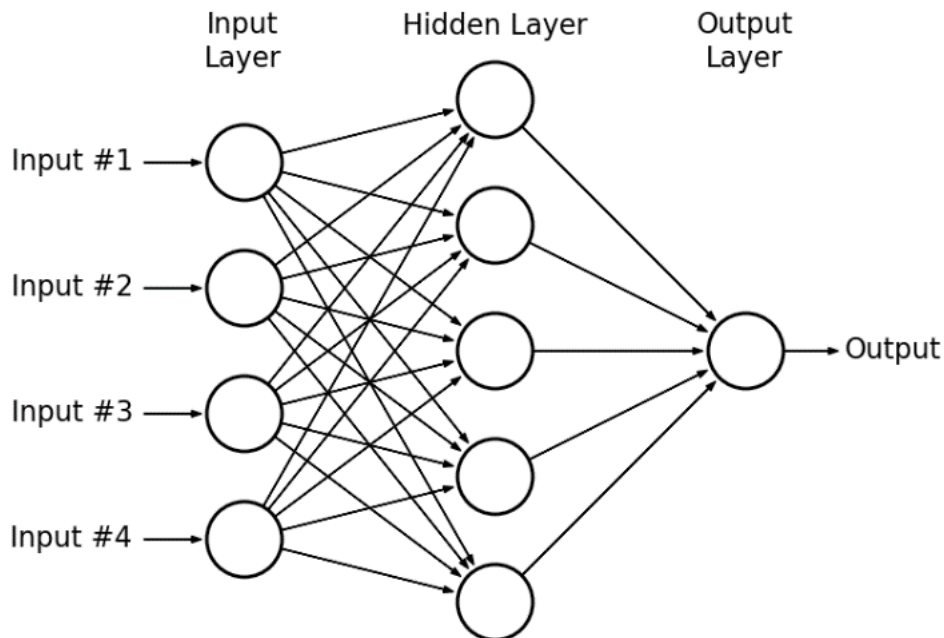


Figure 3.7: Diagram of a fully connected neural network with two layers, 4 inputs and one output. Only 5 neurons are shown here in the hidden layer, although there are 64 in our model.

We have also tried slightly different architectures, for example by increasing the number of neurons in the intermediate layer, or by adding an extra layer, but the results obtained with these

architectures are not significantly better than those obtained with the architecture presented above, so we stick to the simplest performing model.

3.3.2 Training

To train our model, we used the cross entropy loss, defined by :

$$\mathcal{L}(y_{\text{pred}}, y_{\text{true}}) = \mathbb{E} [y_{\text{true}} \log(y_{\text{pred}}) + (1 - y_{\text{true}}) \log(1 - y_{\text{pred}})]$$

where :

- y_{pred} is the output of the model (normalized by the sigmoid), that should be interpreted as a probability that the event is positronic.
- y_{true} is the true label of the event, that is $y_{\text{true}} = 0$ if the event is a muon and $y_{\text{true}} = 1$ if the event is a positron.
- The expected value $\mathbb{E}[\cdot]$ is taken over all the training set.

This loss is theoretically a way to measure the amount of information provided by the true distribution, knowing the predicted distribution. The lower this quantity is, the better the predicted distribution describes the true distribution.

The algorithm used to train this model is the "Adam" algorithm, implemented in the pytorch library, with parameters ($lr = 0.001$; $\beta_1 = 0.9$; $\beta_2 = 0.999$, $\varepsilon = 10^{-8}$).

3.3.3 Results

As mentioned above, we have tested this model with the 2, 3, 4 or 5 first moments (mean, standard-deviation, skewness, kurtosis...) of the extracted distribution as input. The results are summarized in the table 3.2

	2 moments	3 moments	4 moments	5 moments
Angular distribution	93.5%	94.5%	94.5%	95%
Distances distribution	93%	93.25%	93.25%	93.75%

Table 3.2: Equal error rates for different numbers of moments as input of the model, evaluated on the test set, with the angular distribution and the distance distribution

As expected, the combination of several moments provides much better scores, although the different moments taken independently give more or less poor performances. This shows that each moment holds fragmented information, and that by combining these moments, our model manages to take advantage of all this information to obtain a good performance.

Another interesting thing that can be deduced from these results is that in each of the two distributions, in the measure of what our model manages to extract, the kurtosis does not seem to bring any new information compared to the three previous moments, while the fifth moment still allows us to slightly increase the performance.

3.4 Maximum likelihood method

3.4.1 Principle

Finally, the last method we implemented for particle identification is a maximum likelihood method. Given an event (or observation) \mathcal{O} , which consists in the data of the charges and

activation times of the PMTs, we associate to each flavor s (positron or anti-muon) a likelihood \mathcal{L} defined by :

$$L_{\mathcal{O}}(s) = \mathbb{P}[\mathcal{O}|s]$$

Or, more conveniently in the calculations, the negative log-likelihood :

$$\mathcal{L}_{\mathcal{O}}(s) = -\log \mathbb{P}[\mathcal{O}|s]$$

The likelihood function is supposed to characterise the likelihood that the flavor of the incident particle is s , given the observations. Thus, for a positronic event for example, we expect $\mathcal{L}_{\mathcal{O}}(e)$ to be small and $\mathcal{L}_{\mathcal{O}}(\mu)$ to be large (and conversely for a muonic event). Therefore, we use the likelihoods quotient as a score, or equivalently, the log-likelihoods difference.

$$\text{score}(\mathcal{O}) = \mathcal{L}_{\mathcal{O}}(\mu) - \mathcal{L}_{\mathcal{O}}(e)$$

This score can then be subjected to the usual analyses to determine its discriminative power.

3.4.2 Probability distribution

In order to calculate the likelihood function, we need to know the probability distributions $P[\mathcal{O}|e]$ and $P[\mathcal{O}|\mu]$. To do this, we inferred these two distributions on the training set data.

Figure 3.8 shows the histograms of activated PMTs as a function of θ angle (angular distribution), for positronic events and muonic events (normalized to 1).

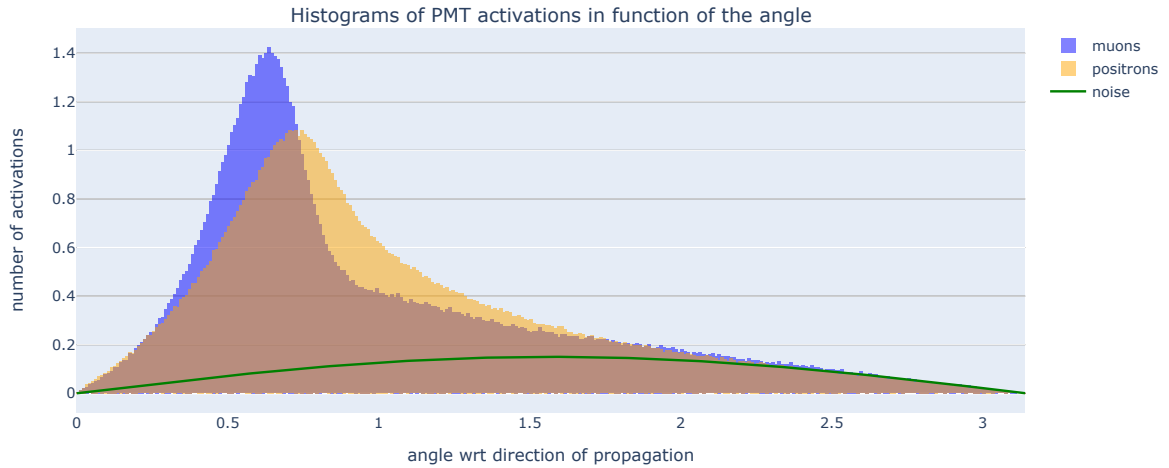


Figure 3.8: Histogram of the PMT activations in function of the angle, for the positronic events (in orange) and the muonic events (in blue)

As can be seen, the distribution has a large tail due to noise, characterized by a uniform angular distribution (i.e. a $\sin(\theta)$ distribution in spherical coordinates). However, this distribution is computed without taking into account the charge of the PMTs. Figure 3.9 shows the histogram obtained by weighting the PMT activations by their charge (also normalized to 1).

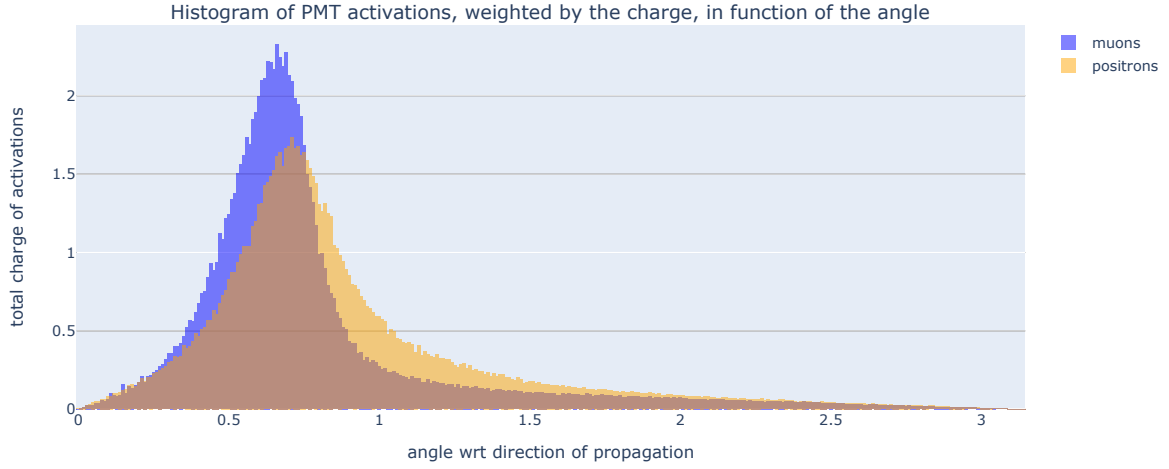


Figure 3.9: Histogram of the PMT activations, weighted by the charge, in function of the angle, for the positronic events (in orange) and the muonic events (in blue)

In addition to being more relevant from a conceptual point of view, we can see that this approach has the effect of drastically reducing the influence of noise on the distribution.

Now, we know the probability of seeing a hit at angle θ , with precision ε , the width of one bin in the histogram, arbitrary small (in practice, we took $\varepsilon = \frac{\pi}{500}$, which is more than enough). We denote this probability $p(\theta)$. Note also that we can obtain the total angular distribution (θ, φ) by simply dividing by $2\pi \sin \theta$, since the φ distribution is always uniform :

$$p(\theta, \varphi) = \frac{p(\theta)}{2\pi \sin \theta}$$

We would like to deduce from that the probability for a given PMT j , situated at angle θ_j , to be hit. We will simply note this probability p_j (j is the index of the PMT). To do this, we need to know the solid angle $\delta\Omega_j = \sin \theta_j \delta\theta \delta\varphi$ that the PMT occupies with respect to the interaction point. This can be calculated easily, knowing the interaction point, the propagation direction and the radius of a PMT. We have then :

$$p_j = p(\theta_j, \varphi_j) \delta\Omega_j$$

Finally, we have to compute the probability of observing the collection of all hits in a single event, that is, the full observation \mathcal{O} . We can estimate this probability as follows:

$$\mathbb{P}[\mathcal{O}|s] = \prod_{j \in \text{PMTs}} \binom{Q}{q_j} p_j^{q_j} (1 - p_j)^{Q - q_j}$$

Where q_j is the charge of the PMT j , and $Q = \sum_j q_j$ is the sum of all charges in the event. The dependence in s resides in the probability distribution p used to compute the p_j .

Forgetting the constant terms with respect to s , we then have :

$$\mathcal{L}_{\mathcal{O}}(s) = - \sum_j q_j \log(p_j) - \sum_j (Q - q_j) \log(1 - p_j)$$

3.4.3 Results

The performance of this score is presented in figure 3.10.

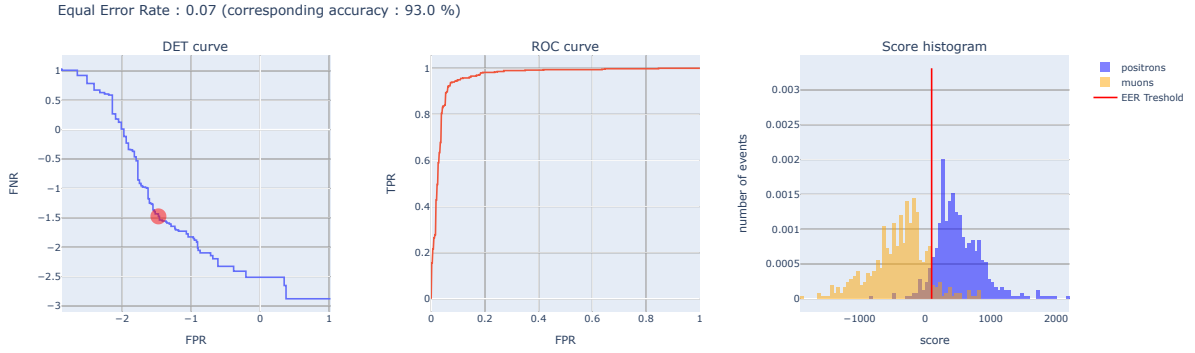


Figure 3.10: Prediction for the likelihood quotient on the angular distribution

This method has the advantage of providing almost the same performances than the multi-moments scores, while being fully canonical and understandable. Further development of this method could imply using wisely the denoised distribution to have better precision, but one should take care of not destroying the useful information in this process.

Chapter 4

Vertex reconstruction

Finally, we decided to experiment with vertex reconstruction. This part has two purposes. First, for the global coherence of the project, since all our particle identification methods use the position of the interaction vertex, it seemed relevant to try to implement vertex reconstruction methods. Second, since we used a maximum likelihood method for particle identification, we wanted to reuse this formalism to apply it to vertex reconstruction at the same time.

4.1 Likelihood method

Let us briefly recall the results developed in the Section 3.4. The negative log-likelihood function is given by :

$$\mathcal{L}(\vec{x}_0, \vec{u}) = - \sum_j q_j \log(p_j) - \sum_j (Q - q_j) \log(1 - p_j)$$

where this time, the negative log-likelihood function is a function of the position of the interacting vertex \vec{x}_0 , and of the propagation direction \vec{u} , which are the parameters we are trying to determine. The dependence in these parameters is involved in the computation of p_j (j being the indices of the PMTs), which is expressed as :

$$p_j = p(\theta_j, \varphi_j) \delta\theta \sin\theta_j \delta\varphi$$

where the angles θ_j depend on \vec{x}_0 and \vec{u} (cf. Section 2.6.1), and the probability distribution p is now the averaged distributions on the two flavours e and μ (as we assume we don't know the flavour when applying vertex reconstruction).

Thus, the method we have implemented consists in minimising this function $\mathcal{L}(\vec{x}_0, \vec{u})$ by gradient descent on the six scalar parameters (\vec{x}_0, \vec{u}) .

4.2 Initialisation with heuristic

To make the gradient descent more efficient, we initialise it with a heuristic on \vec{x}_0 and \vec{u} . The heuristic on \vec{x}_0 simply consists in taking the average of the hits positions, weighted by the charge. From this \vec{x}_0 , we compute the direction \vec{u} with the method used previously (see Section 2.5). Finally, we slightly improve the heuristic on \vec{x}_0 by moving \vec{x}_0 along the \vec{u} direction so that the hits are already best distributed according to the $p(\theta)$ distribution.

4.3 Results

On the test set, the average distance obtained between the reconstructed vertex and the true vertex is about 3m. Further investigations show multiple events on which the method seems to

work quite well, and others on which it fails completely. We interpret this by the fact that in the cases where the heuristic is bad, the gradient descent gets stuck in a local minimum far away from the true vertex of interaction.

We are aware that this performance is well below the state of the art, which manages to obtain an accuracy of about 30cm on single rings [9]. However, we did not have time to develop the method further given the time constraints.

Further development could involve taking into account the activation times of PMTs, which seems to be a very rich information for the vertex reconstruction task, which we do not take advantage of at all.

Conclusion

Only armed with a Python notebook and a few megabytes of simulated data, we had to dive head first in the unknown (for us) world of experimental physics. The results are quite satisfactory, we have managed to discriminate between anti-muons and positrons with a respectable precision, close enough to the literature state of the art [9] given our time constraints and our limited experience in data analysis. We have explored the typical features of the iconic particle physics detector that is Super-Kamiokande, namely analysing the Cherenkov rings of atmospheric neutrinos. Of course things could be further refined, we still test plethora of scores and tweak arbitrary parameters (cuts to the distributions, thresholds for denoising, neural network architectures, optimization algorithms...), but the main channel for improvement would be an estimation of the energy of the interacting neutrino, which defines the angle of the first Cherenkov ring (the outer edge of the ring if there was no dispersion), at the heart of our discrimination between positrons and anti-muons. Approximating the speed of the charged leptons to be c is a bit rough for the lowest energy events of our data set (muon at ~ 200 MeV $\Rightarrow \beta \sim 0.85$). You can see in Appendix B examples of unforgiving events where our methods fail, probably due to this energy approximation, or to the way we extract the centre of the ring. Last but not least, further investigations into how to use the temporal distribution of PMT hits could also lead to significant improvements in the various reconstruction methods. All this work could of course be more than improved with further documentation and implementation of state-of-the-art methods, but for the sake of discovery and experimentation, the project has been deliberately framed to a limited use of documentation. In the end, the whole project was a great experience. We were given a certain independence in our work, which let us experiment and -to be honest- have fun. The nice atmosphere of the LLR neutrino team has almost convinced us to reject theoretical physics and embrace experimental physics (almost...). After all these histograms, a touch a poetry is much needed, delicacy of an inspired physicist [7].

The Lord of the Rings

Photons from ice and sea under the sky,
Photons from vast water tanks in halls of stone,
Photons from the atmosphere in an insect's eye,
Photons from aerogels, light, clear, blown,
Photons from liquids, gases, crystals flying by,
Photons from fused silica expanding on a cone.
In RICH detectors where PID truths lie.
One Ring to rule them all, One Ring to find them,
One Ring to bring them all, correlate, and bind them
In RICH detectors where PID truths lie.

Bibliography

- [1] S Agostinelli. “GEANT4-a simulation tool kit Nuclear Instruments and Methods in Physics Research. 2003”. In: *A506* (), pp. 250–303.
- [2] J Allison et al. *Recent developments in Geant4 Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment 835: 186-225*. 2016.
- [3] John Allison et al. “Geant4 developments and applications”. In: *IEEE Transactions on nuclear science* 53.1 (2006), pp. 270–278.
- [4] Alice Coffani. “New studies on cosmogenic induced spallation background for Supernova relic neutrino search in the Super-Kamiokande experiment”. Theses. Institut Polytechnique de Paris, Dec. 2021. URL: <https://theses.hal.science/tel-03591741>.
- [5] J. A. Formaggio and G. P. Zeller. “From eV to EeV: Neutrino cross sections across energy scales”. In: *Rev. Mod. Phys.* 84 (3 Sept. 2012), pp. 1307–1341. DOI: 10.1103/RevModPhys.84.1307. URL: <https://link.aps.org/doi/10.1103/RevModPhys.84.1307>.
- [6] Karl Florian Goebel. “A study of particle identification with the Super-Kamiokande detector”. MA thesis. Citeseer, 1996.
- [7] Blair N Ratcliff. “Imaging rings in ring imaging Cherenkov counters”. In: *Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment* 502.1 (2003), pp. 211–221.
- [8] André Rougé. *Introduction à la physique subatomique*. fre. Palaiseau: les Éditions de l’École polytechnique, 2005. ISBN: 2-7302-1231-0.
- [9] M Shiozawa. “Reconstruction algorithms in the Super-Kamiokande large water Cherenkov detector”. In: *Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment* 433.1-2 (1999), pp. 240–246.
- [10] Ig Tamm. “Radiation emitted by uniformly moving electrons”. In: *Selected Papers*. Springer, 1991, pp. 37–53.
- [11] *WCSTim git repository* : <https://github.com/WCSTim/WCSTim>.
- [12] Wikipedia. *Cherenkov radiation*. URL: https://en.wikipedia.org/wiki/Cherenkov_radiation.

APPENDIX

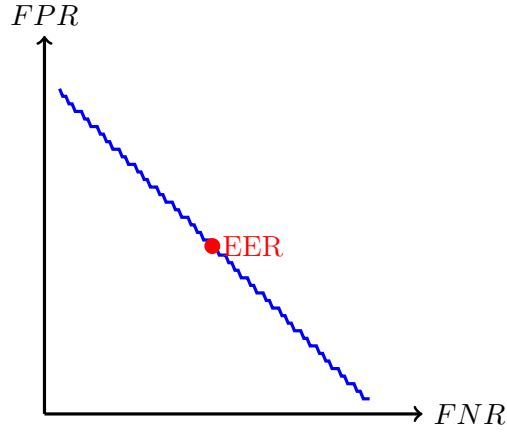
A Prediction analysis

The purpose of this appendix is to present some methods used to evaluate classification models, typically here the two-class classification task for events: positronic and muonic.

A very basic evaluation of a classification model consists in calculating the accuracy, i.e. the number of examples where the network predicts the right class out of the total number of examples. But it is often necessary to have more advanced statistics to evaluate the model. To do so, we consider a given class and we take into account the score assigned to this class by the model for each example of the test set. We then try to determine a threshold at which the example will be considered as belonging to the class if the score exceeds the threshold, and as being excluded from the class otherwise. Given a threshold, we can then calculate the following values:

- Number of True Positives (TP): Number of examples predicted to belong to the class and actually belonging to it.
- Number of True Negatives (TN): Number of examples predicted as not belonging to the class and actually not belonging to the class.
- Number of False Positives (FP): Number of examples predicted to belong to the class when they do not.
- Number of False Negatives (FN): Number of examples predicted as not belonging to the class when they do belong.
- False Positive Rate (FPR) : $FPR = \frac{FP}{FP + TN}$
- False Negative Rate (FNR) : $FNR = \frac{FN}{TP + FN}$

By varying the threshold, we can then vary these different quantities, and we can for example draw the parametric curve $(FNR(t), FPR(t))$ where t is the threshold. In an adapted scale ($x = \phi_{\mathcal{N}}^{-1}(FNR(t))$ and $y = \phi_{\mathcal{N}}^{-1}(FPR(t))$ where $\phi_{\mathcal{N}}$ is the cumulative distribution function of the standard normal distribution), we obtain a curve with the following global appearance:



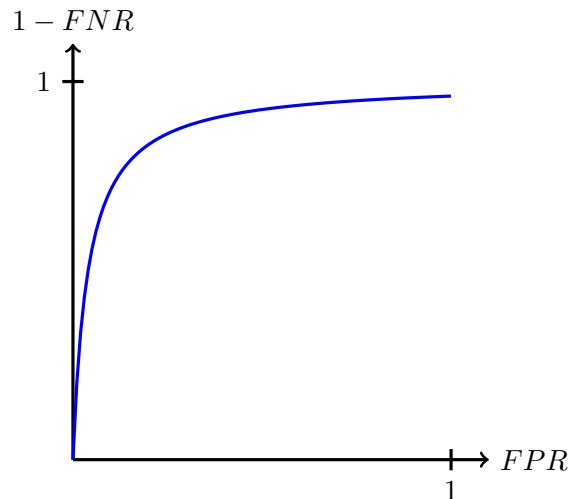
This curve is called the DET (Detection Error Tradeoff) curve. The EER (Equal Error Rate) is the value at the point $FNR = FPR$, and can be a measure of the quality of the system. However, the part of the curve to look at will depend on the use case. Indeed, depending on the case, one may wish to minimize the FPR , or the FNR in priority. Moreover, we can also weight the decision by the *a priori* probability that the events have to occur. If, for example, it is much more likely *a priori* that an example will be positive than negative (or in more practical words, if we expect to see more positrons than muons in a certain case), one can require a higher degree of certainty from the system to predict a negative, which results in a lower threshold. These requirements can be translated into the minimisation of the following function, called DCF (Decision Cost Function):

$$DCF = p \times C_N \times FNR + (1 - p) \times C_P \times FPR$$

where p denotes the *a priori* probability that an example is positive and C_N and C_P are the costs assigned to predicting false negatives and false positives respectively.

The value of the minimum, $\min DCF$, then gives a measure of the quality of the model to meet the requirements in a particular use case, and the threshold for which this minimum is reached gives an indication of the threshold to choose for making predictions.

Another curve that can be drawn is the ROC (Receiver Operating Characteristic) curve, which is the parametric curve $(FPR(t), 1 - FNR(t))$. It is useful to evaluate the quality of the model as a whole. The relevant measures can then be the integral of the curve over $[0, 1]$, which is larger the more qualitative the model is (1 being the maximum). Keeping the linear scale, the ROC curve usually looks like this:



B Bloopers

We have made the strong assumption that the produced charged leptons were emitted at c , implying a certain Cherenkov angle for the outer edge of the ring (without dispersion). We plot in Figure 1 the dependence of this angle to the energy of the particle (muon here). We see that our assumption could be problematic for the lowest energies of our data-set ($\sim 200\text{-}300\text{MeV}$).

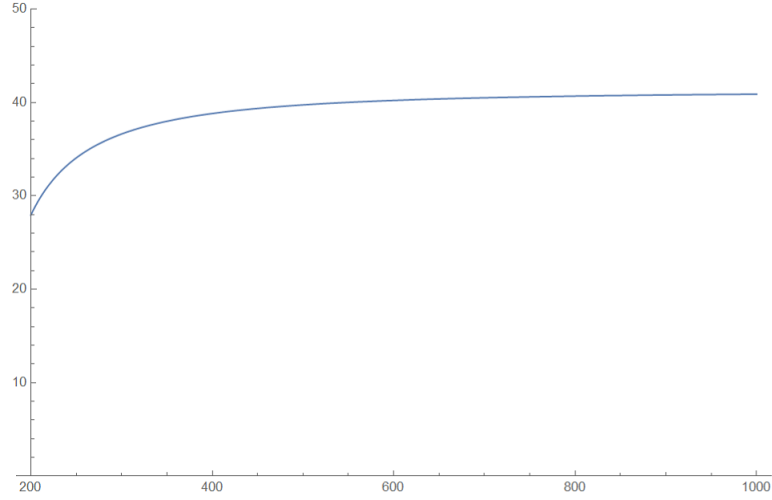


Figure 1: Dependence of the muon Cherenkov angle θ_c on the energy E (in MeV).

The following Figures are a compilation of examples where our methods failed, either because the events were at low energy, or for reasons we cannot really explain.

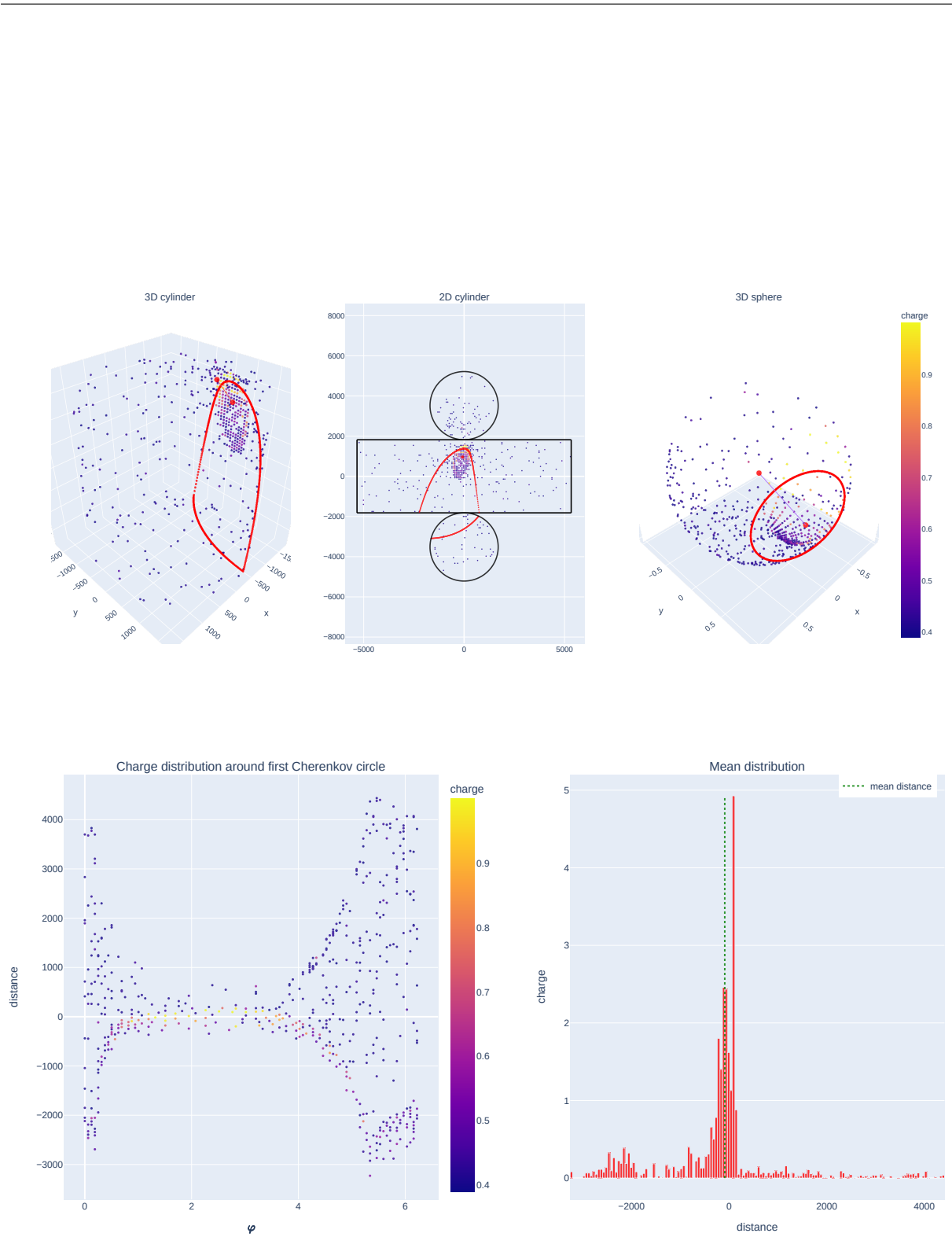


Figure 2: An important failure.

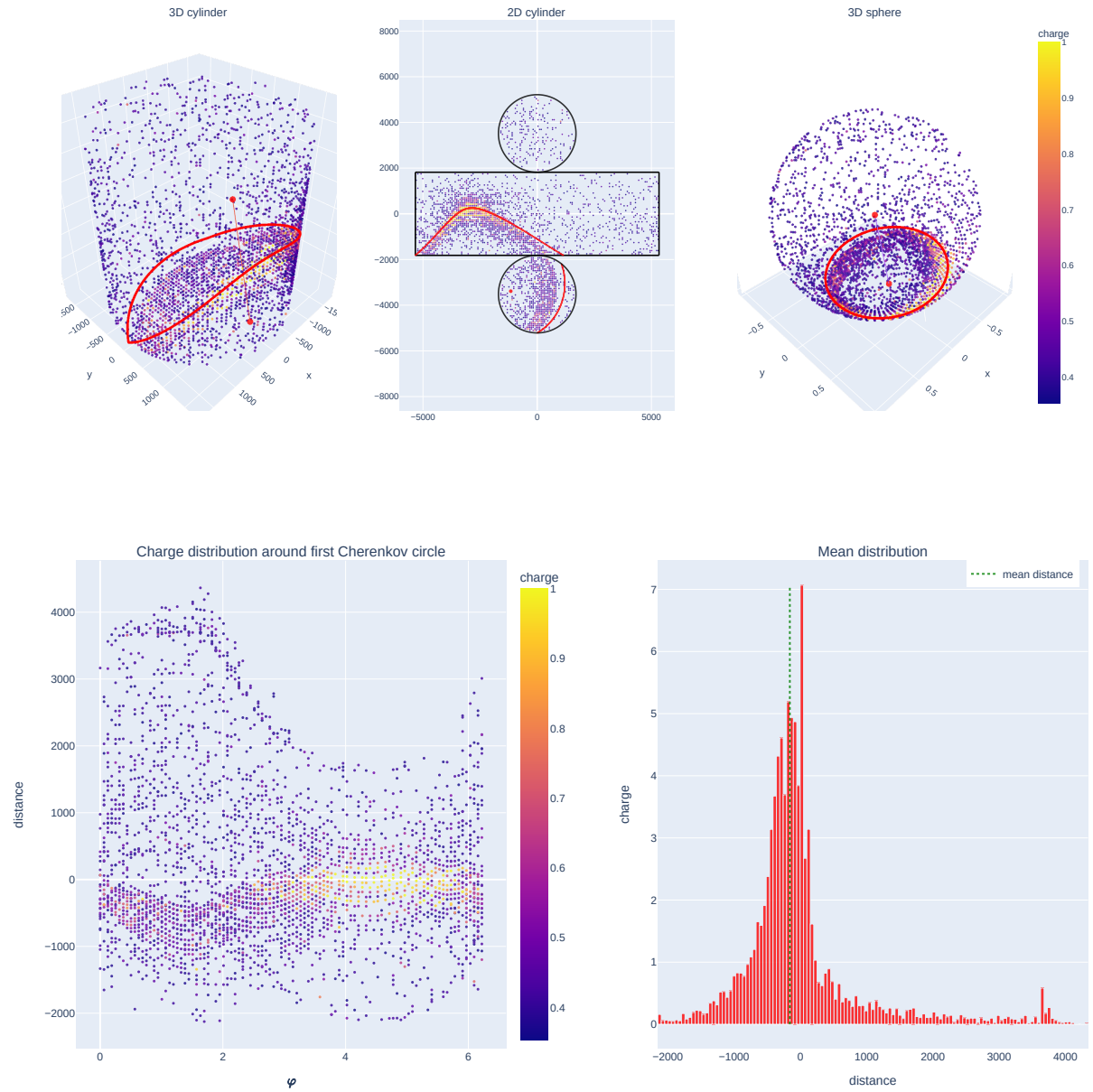


Figure 3: A slight issue, probably a low energy event, or a bad evaluation of the centre of the ring.

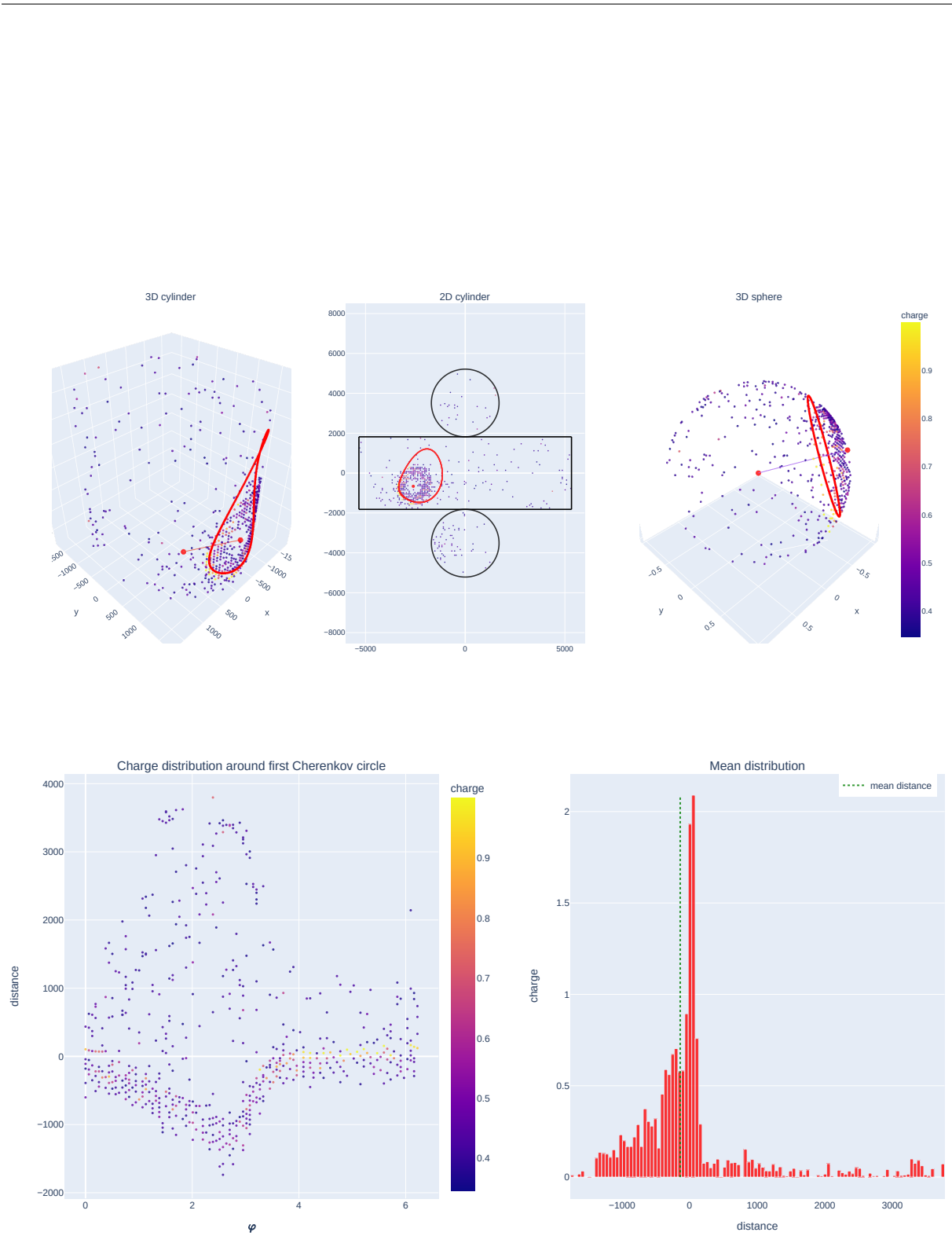


Figure 4: Another slight issue.