



UNIVERSITÀ
degli STUDI
di CATANIA

Dipartimento
di Fisica
e Astronomia
"Ettore Majorana"



MSC PROGRAMME IN PHYSICS

CLARA SAIA

ANALYSIS OF THE AUTOMATIC RECONSTRUCTION OF PROTONS FROM
SIMULATED NEUTRINO INTERACTIONS IN THE ICARUS DETECTOR: STUDY
OF THE PERFORMANCE OF TRACK VS SHOWER DISCRIMINATION ALGORITHMS



SUPERVISOR:

BRUCE HORWARD

MENTORS:

ALICE CAMPANI

LEA DI NOTO

ITALIAN SUMMER STUDENT PROGRAM @ FNAL 2022

Contents

1	The ICARUS-T600 as a detector at FNAL	2
1.1	The Short Baseline Neutrino Program	2
1.2	Neutrino Oscillation physics and anomalies	3
1.2.1	Anomalies	4
1.3	LAr-TPC technology	6
1.3.1	ICARUS as a LAr-TPC	6
2	Event Reconstruction	9
2.1	Booster Decision Tree Algorithm: input parameters	10
2.2	CCQE events: our set of MC simulated data	11
2.2.1	Examples of events	12
2.3	Analysis of topologies	12
2.4	Selection of a reference set	13
3	Final Results	17
3.1	BDT variables plots	17
3.1.1	2d plots	19
4	Conclusions	21
	Bibliography	21

Chapter 1

The ICARUS-T600 as a detector at FNAL

ICARUS, Imaging Cosmic and Rare Underground Signals, is the far detector of the Short Baseline Neutrino (SBN) Program at Fermi National Accelerator Laboratory (FNAL) [1].

This program was born with the purpose of definitely explaining some anomalies in the field of neutrino oscillation physics, that can suggest the existence of a 4-type of neutrino: the sterile neutrino [2]. It doesn't interact directly with ordinary matter through the weak interaction, but we can observe the effect of its oscillation. In this chapter I will briefly observe:

- the physics goals of the SBN Program;
- the role of ICARUS as SBN far detector;
- the detection technology.

1.1 The Short Baseline Neutrino Program

The Short Baseline Neutrino Program is made up of three detectors: SBND, MicroBooNE and ICARUS. They are located in the beam line of the ν_μ (or $\bar{\nu}_\mu$) 8 GeV Booster Neutrino Beam, at different distances from

the target: 110 m, 470 m and 600 m respectively (Fig. 1.1).

They are all LAr-TPC (Liquid Argon Time Projection Chambers) and their dimensions are gradually larger to keep the same solid angle. The neutrino spectrum is different at the different detectors.

The SBN project has the purpose of studying neutrino oscillations and is sensitive both in the neutrino appearance and disappearance channels. It will also allow to perform studies on neutrino-argon scattering and Beyond the Standard Model physics.

The ICARUS detector had its first life in Italy at the Laboratori Nazionali del Gran Sasso (LNGS) under the CERN Neutrinos to Gran Sasso (CNGS) beam line. By operating as a far detector at shallow depth, ICARUS was surrounded by a Cosmic Ray Tagger to help with the discrimination from neutrino signal and cosmic background [1] [2] [3].

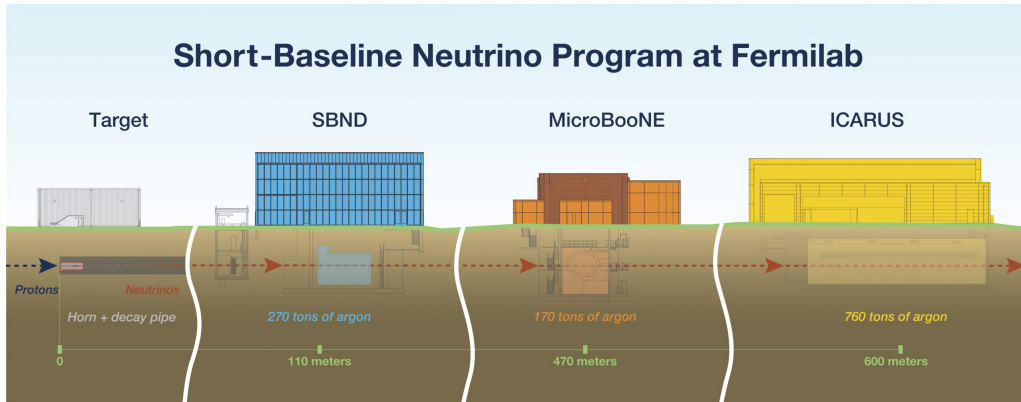


Figure 1.1: ICARUS is the largest and farthest detector in the Short-Baseline Neutrino program at Fermilab, which examines neutrino oscillations over short distances and looks for hints of elusive sterile neutrinos. Graphic: Fermilab [4]

1.2 Neutrino Oscillation physics and anomalies

In the Standard Model (SM), neutrinos exist in three states of flavour: ν_e , ν_μ and ν_τ . Neutrinos weakly interact through charged-current (CC), ex-

changing W^\pm bosons and neutral-current (NC), exchanging Z^0 boson.

In 1957 Pontecorvo proposed a first theory on neutrino oscillation, later developed by Maki, Nakagawa and Sakata in 1962.

If we assume that the neutrino has a mass, we can write the three flavor eigenstates $|\nu_e\rangle$, $|\nu_\mu\rangle$ and $|\nu_\tau\rangle$ as a linear combination of three mass eigenstates $|\nu_1\rangle$, $|\nu_2\rangle$ and $|\nu_3\rangle$, according to the following relationship:

$$\begin{pmatrix} |\nu_e\rangle \\ |\nu_\mu\rangle \\ |\nu_\tau\rangle \end{pmatrix} = \begin{pmatrix} U_{e1} & U_{e2} & U_{e3} \\ U_{\mu1} & U_{\mu2} & U_{\mu3} \\ U_{\tau1} & U_{\tau2} & U_{\tau3} \end{pmatrix} \begin{pmatrix} |\nu_1\rangle \\ |\nu_2\rangle \\ |\nu_3\rangle \end{pmatrix} \quad (1.1)$$

The unitary matrix U , called the PMNS matrix (Pontecorvo-Maki-Nakagawa-Sakata) [5] [6].

According to their theory the probability that a $|\nu_\alpha\rangle$ flavour neutrino is detected as another flavour $|\nu_\beta\rangle$ is:

$$P_{\nu_\alpha \rightarrow \nu_\beta} = |\langle \nu_\beta | \nu_\alpha(t) \rangle|^2 = \sum_{j=1}^3 \sum_{i=1}^3 U_{\alpha i}^* U_{\beta i} U_{\alpha j} U_{\beta j}^* e^{-i(\frac{\Delta m_{ij}^2}{2} \frac{L}{E})} \quad (1.2)$$

The oscillation depends on:

- L/E ratio. L stands for Length, i.e. the distance traveled by neutrinos in the laboratory frame. E stands for neutrino Energy;
- $\Delta m_{ij}^2 = m_i^2 - m_j^2$ i.e. the mass square difference in eV^2

We note that if the neutrinos had no mass or had identical masses there would have been no oscillation [2] [6] [5] [7].

1.2.1 Anomalies

In the last decades a lot of experiments from solar, reactor and accelerator neutrinos, showed a set of parameter strongly compatible with the 3-neutrino oscillation scenario with the following as most recent parameters results: $|\Delta m_{31}^2| = 2.5 * 10^{-3} eV^2$ e $\Delta m_{21}^2 = 7.4 * 10^{-5} eV^2$ [8].

However there has been also some anomalies, in particular in the so-called short-baseline: $L/E \leq 1$ m/MeV:

- **disappearance** of low-energy ν_e from nuclear reactors: the **reactor anomaly**;
- **appearance** of ν_e in muon neutrino beams in accelerators such as LSND and MiniBooNE: the **LSND and MiniBooNE anomalies**.
- **disappearance** of ν_e in the SAGE (Soviet-American Gallium Experiment) and GALLEX (Gallium Experiment) experiments, created with the aim of measuring the flux of solar neutrinos: the **Gallium anomaly**;

None of these anomalies can be explained by the oscillation of the three neutrinos of the standard model. One possible explanation is the existence of a fourth neutrino, the **sterile neutrino**, which does not interact weakly. This neutrino would be responsible for the short-range oscillations with Δm^2 greater than at least 3 orders of magnitude with respect to the known values, with $\Delta m_{4i}^2 \sim 1 \text{ eV}^2$ ($i = 1,2,3$) [1] [2].

In these new 3+1 neutrino scenario we can rewrite the PMNS matrix as:

$$\begin{pmatrix} |\nu_e\rangle \\ |\nu_\mu\rangle \\ |\nu_\tau\rangle \\ |\nu_s\rangle \end{pmatrix} = \begin{pmatrix} U_{e1} & U_{e2} & U_{e3} & U_{e4} \\ U_{\mu1} & U_{\mu2} & U_{\mu3} & U_{\mu4} \\ U_{\tau1} & U_{\tau2} & U_{\tau3} & U_{\tau4} \\ U_{s1} & U_{s2} & U_{s3} & U_{s4} \end{pmatrix} \begin{pmatrix} |\nu_1\rangle \\ |\nu_2\rangle \\ |\nu_3\rangle \\ |\nu_4\rangle \end{pmatrix} \quad (1.3)$$

As long as $\Delta m_{41}^2 \gg |\Delta m_{31}^2|, \Delta m_{21}^2$, oscillations at short-baseline experiments can be well described by a two-flavor vacuum oscillation formula

$$P_{\nu_\alpha \rightarrow \nu_\beta} = \delta_{\alpha\beta} - 4|U_{\alpha\beta}|^2(\delta_{\alpha\beta} - |U_{\alpha\beta}|^2)\sin^2\left(\frac{\Delta m_{41}^2 L}{4E}\right) \quad (1.4)$$

The characteristic $\sin^2\left(\frac{\Delta m_{41}^2 L}{4E}\right)$ dependence may allow one to distinguish it from other possible explanations of the anomalies [2].

The ICARUS detector can fortunately do measurement both in ν_μ disappearance channel (with the BNB beam) but also in the ν_e channel: in fact it will also record events from the off-axis flux of the NuMI neutrino beam, with its higher electron neutrino content and different energy spectrum.

1.3 LAr-TPC technology

Liquid Argon Time Projection Chamber can be considered as the leading technique in the field of neutrino physics. It can join the large sensitive mass of electron detectors and the high resolution of the old bubble chambers [9].

In fact, thanks to the Liquid Argon properties, when a charged particle (e.g. produced by a neutrino interaction) arrives in the detector, it ionizes the argon: a couple ion-electron is produced.

Liquid argon allows free electrons produced by ionization to move in the liquid over distances of the order of a few meters, thanks to the following main characteristics:

- the density (1.4 g/cm^3) which guarantees greater interaction;
- the very high degree of purity, achievable thanks to an advanced purification system;
- high electronic mobility.

Furthermore, argon can be easily cooled with nitrogen. Inside the cryostat, the argon is kept at a temperature of 98 K, by means of a liquid nitrogen cooling system [1] [9].

Another property of liquid argon is the ability to generate scintillation light: the particles passing through the detector excite the molecular bands of the argon, emitting light with wavelength which assumes a peak value at $\lambda = 128 \text{ nm}$ [1].

1.3.1 ICARUS as a LAr-TPC

ICARUS basically consists of two adjacent cryostats module containing two TPCs that share a central cathode. The external dimensions are of $3.6 \times 3.9 \times 19.6 \text{ m}^3$ for each cryostat. ICARUS is filled with 760 tons of ultra-pure LAr, of which only 480 are sensitive volume.

The charged particles, produced by the neutrino interaction, are drifted by a uniform electric field ($E_{drift} = 500 \text{ V/cm}$) toward the anode

with velocity $v_{drift} = 1.6 \text{ mm}/\mu\text{s}$. They are directed towards 3 parallel wire planes, placed at the end of the sensitive volume. These planes consist of a total of nearly 54,000 9 m long threads, each oriented along one of three directions: 0° , $+60^\circ$, -60° with respect to the horizontal. They are respectively called **Induction-1**, **Induction-2** and **Collection**. The charge induces signals on the first two in a non-destructive way, and is finally collected by the last plane of wires.



Figure 1.2: Picture (left) and scheme of one of the cryostat (right) [1] [3]

The trace of the ionizing particle is reconstructed three-dimensionally in the coordinates (x, y, z) by combining the signals from the wires (x, y) with the drift time (from which the third coordinate is obtained) and allows to obtain a spatial resolution of the order of mm^3 (Fig. 1.3) [1] [3].

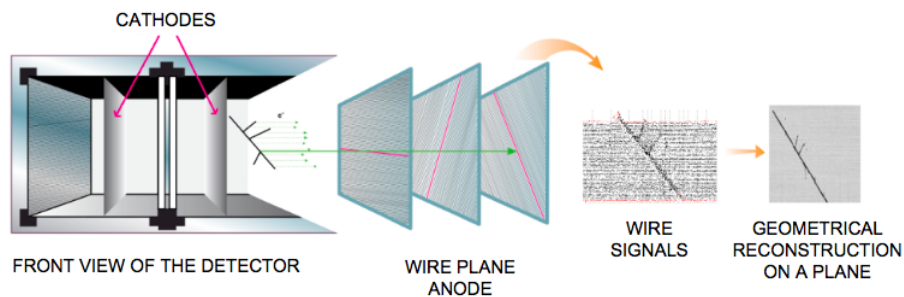


Figure 1.3: Principle of operation of ICARUS-T600: the path of a charged particle in liquid argon and its geometric reconstruction [1]

Behind the plane of wires are installed arrays of 90 photo-multipliers (PMTs), for a total of 360 PMTs, which make up the light collection system of the detector. The PMT will be used to best perform the trigger signal of the detector [1].

Event Reconstruction

In this second chapter I will focus on the details of my job; I worked with the ICARUS TPC reconstruction group. My purpose was to analyze a set of Monte Carlo (MC) simulated data and try to help in the reconstruction of protons in a Charge Current - Quasi Elastic (CCQE) neutrino interaction. As I mentioned CC means that the interaction is mediated by a charged boson. The name Quasi-Elastic recalls an elastic scattering: the neutrino interacts with a neutron in the argon; there are two particles that interact and two other particles (a lepton and an hadron) exiting from the vertex of interaction. In the Fig. 2 it is shown a scheme of an example of a muon neutrino interaction on the left picture and there is a real data candidate from the BNB on the right.

It is clear that we cannot observe directly the neutrino, but we can only see in the detector, the effects of the charged particles that exit from this interaction (e.g. the proton and the muon, as in the example). CCQE neutrino interaction events are important in my analysis because they can be a very simple reference to look for some pathology in the event reconstruction. In particular, in my analysis I tried to investigate those cases where the proton, that should be a track, is seen by the reconstruction algorithm, instead, as a shower.

The algorithm used is a Booster Decision Tree (BDT) [10], which is an algorithm that looks at some features of the signal of the particle (*input parameters*) and decides how to cluster the particle. In my particular case the BDT decides to classify the proton as a track or as a shower (misiden-

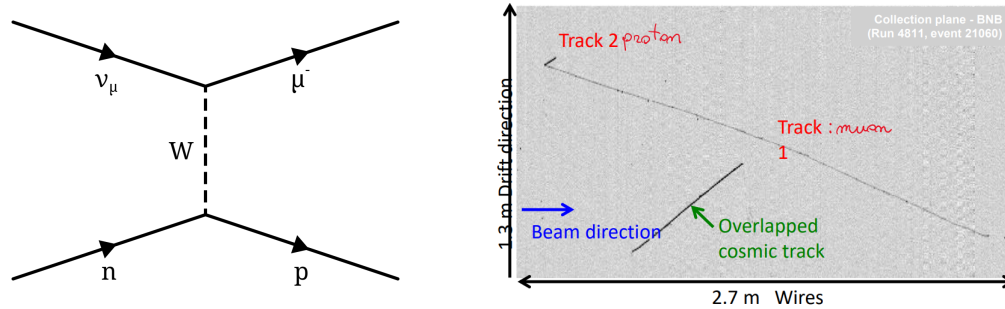


Figure 2.1: Feynman diagram of a CCQE interaction for a muon neutrino (left) and one of the first CCQE ν_μ candidate found in BNB (right).

tifying it).

2.1 Booster Decision Tree Algorithm: input parameters

The BDT uses 10 input variables [10]:

- **Length:** estimate of the length of the reconstructed particle.
- **Sliding Linear Fit:** estimate of difference with respect to a straight line (normalized to length).
- **Sliding Linear Fit:** estimate of largest gap on the 3 planes (normalized to length).
- **Sliding Linear Fit:** estimate of RMS with reference to the fit (normalized to length).
- **Vertex distance:** distance from interaction reconstructed vertex to the start of the reconstructed particle.
- **Difference in open angle:** difference in “opening” angle and “closing” angle (from 2 points at beginning and at the end of the particle).
- **Principal Component Analysis: secondary eigenvalue** (normalized to primary one)[11].

- **Principal Component Analysis: tertiary eigenvalue** (normalized to primary one)[11].
- **Charge: fractional spread** (using spread in values and mean value).
- **Charge: fraction** near the end of particle (using charge near end and total).

The chapter 3 will show the distribution of these variables for my data set.

2.2 CCQE events: our set of MC simulated data

The analysis was performed on a set of MC simulated data from the BNB. In this set there were 3738 CCQE events. I listed different topologies of reconstruction done by the BDT in the following table (if there is more than one proton, I look at the most energetic one):

Type of Protons Events	Number	Percentage
Found only in the reconstructed tracks	2990	80,0%
Found only in the reconstructed showers	158	4,2%
Found both in reconstructed tracks and showers*	463*	12,4%
Not reconstructed at all	127	3,4%
Total	3738	\

*of which **218 have most of their energy reconstructed as a shower**, meaning their best match is a shower

This means that in the 80% of the cases the proton is well reconstructed. In the rest of the cases the proton has some matches both as a shower and as a track, or even only as a shower or even worst is not reconstructed at all.

2.2.1 Examples of events

In the Fig. 2.2 there is an example of CCQE muon neutrino interaction on the event display: what we observe are the products of the interaction, the muon and the proton. This is a case where there is a well reconstructed proton. In this view there are the 2-dimensions projections on each plane of wire (Induction 1, Induction 2 and Collection). The abscissa represents that wire number and remembering that in the Induction 1 plane the wires are at 0 degree, in this view the abscissa is proportional to the vertical direction. The ordinate represents the time and it is the same on each of the views.

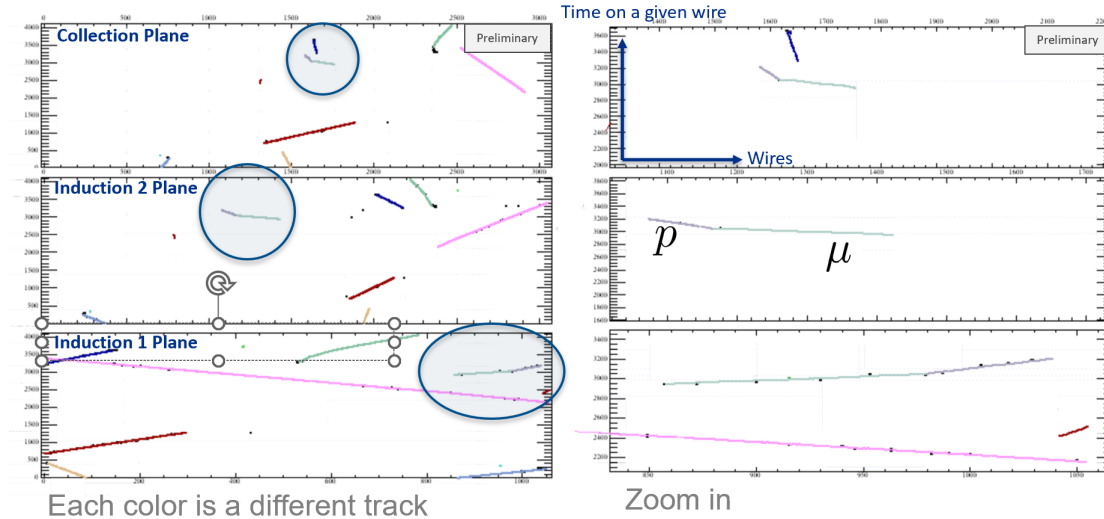


Figure 2.2: Example of a well reconstructed event

2.3 Analysis of topologies

After this first analysis of the data, I have done, for each of the topologies, a characterization of the CCQE protons (length, energy, angle between proton and muon, etc) in order to look for some pattern (e.g. all the misidentified protons are the ones with lower energy).

Following in Fig. 2.3 on the left, as an example, it is shown the recon-

structed length of a the proton in the case where it has a match only in the showers. In the plot on the right there is the ratio between the reconstructed length and the true length.

Despite the fact that the protons are identified as showers, the reconstruction itself is not so problematic as we can see in the figure regarding the ratio, which is more or less peaked at 1. But what I found out is that, if I force the reconstruction as a track, it improves. This is shown in Fig. 2.4. So the focus became understanding which (one or more) of the BDT parameters causing this misidentification.

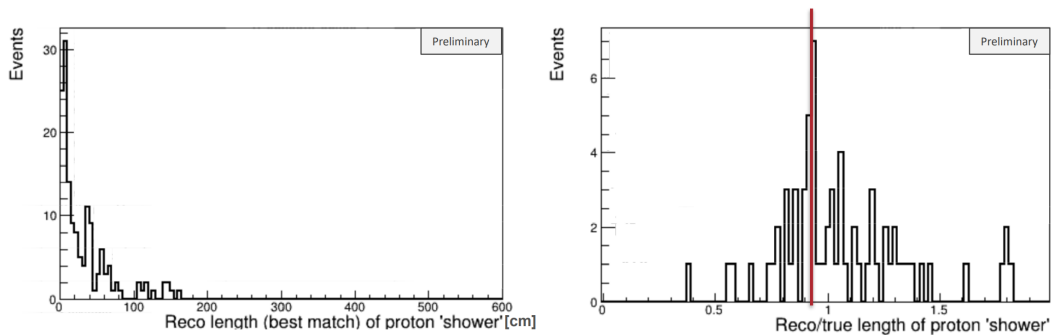


Figure 2.3: Reconstructed length as a shower of the proton with only match in the shower (left) and ratio between reconstructed and real length (right)

2.4 Selection of a reference set

At this point the best option was to identify the best set of data, in terms of reconstruction, in order to take this set as a reference for comparing the BDT variables. Starting from the events where the protons were well reconstructed (80%) I added some other constraints to this set that I am going to discuss.

The types of analysis described in the previous paragraph allowed us to identify an important pathology: in the Fig. 2.5 there are different

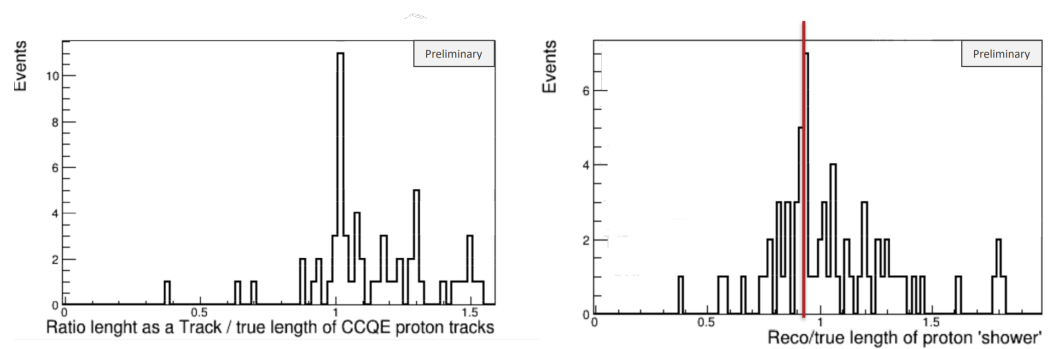


Figure 2.4: Ratio between reconstructed and real length of the CCQE protons with only match as showers: reconstruction as a shower (right) and if we force the reconstruction as tracks (left)

plots regarding this set of well reconstructed protons. In plot A there is the true length, in the plot B the reconstructed length and in the plot D the ratio. Despite the fact that the ratio is peaked at 1, the plot B seems to be a little bit different from the one A. I discovered that this is due to the fact that sometimes the algorithm mixed the reconstruction of the proton with the muon one and assigns to both of them the same track length. In plot C there is the true length of the muon of the same CCQE events. This usually happens for instance when the proton is collinear to the muon or a very small proton track is normal to the muon track.

So in the final reference set ("Good Events") I removed all the cases where the reconstructed proton length was equal to the muon one.

Furthermore I added other constraints to have cleaner sample:

- removed all the events where the CCQE muon was not reconstructed;
- excluded the events where the CCQE protons had purity and completeness $< 90\%$. The purity is the ratio between the matched hits for a given particle and the hits that are actually found in the reconstructed tracks for that particle. The completeness is the ratio between the same matched hits and the total number of hits belonging to the track in truth;

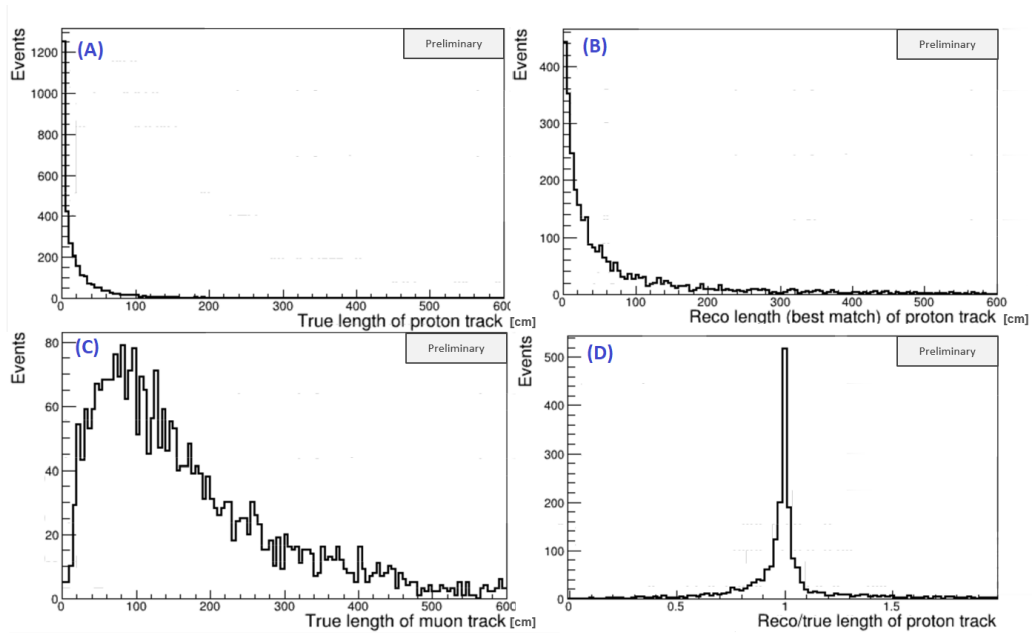
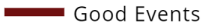


Figure 2.5: CCQE events with only match in the tracks. True length of the protons (A). Reconstructed length of the protons (B). True length of the muons (C). Ratio between A and B (D).

- excluded the events where the CCQE muons had purity and completeness $< 90\%$.

In the following table I listed the numbers of events that I removed; the good reference sample is in the end 1019 events (went from 80,0% to 27,3% of the total numbers of events).

Type of Protons Events	Number	
Found only in the reconstructed tracks	2990	-
Reconstructed proton length == reconstructed muon length	883	-
Muon not reconstructed as a track	25	-
Completeness and Purity < 90% for proton	899	-
Completeness and Purity < 90% for muon	164	=
Good Sample:	1019	

Chapter 3

Final Results

As a final result I have analysed all the BDT variables for the good sample (in the plot indicated with a brown line) comparing those to the "worst sample": the events where the proton is reconstructed only as a shower (blue line). I also added the events where the protons have both matches in the reconstructed showers and tracks as an intermediate case (green line). The general idea was to understand which of the BDT variables influence more the particle identification.

3.1 BDT variables plots

As mentioned in section 2.1, this chapter will show the distribution of these variables for my data set.

In the Fig. 3.1 there are the BDT variables related to the linear fit. The linear fit length distribution seems to be very similar for all the cases and this is consistent with the fact that in general there is not a problem in the length reconstruction, as we have seen. All the other variables related to the fit (gap length, linear fit difference, rms) instead show a different trends in the three samples. In particular, for the "blue line" the type of behavior that we have is the one that you can expect for particles that are real shower. And also, the case that I choose as intermediate shows actually an intermediate behavior in the distribution (green line).

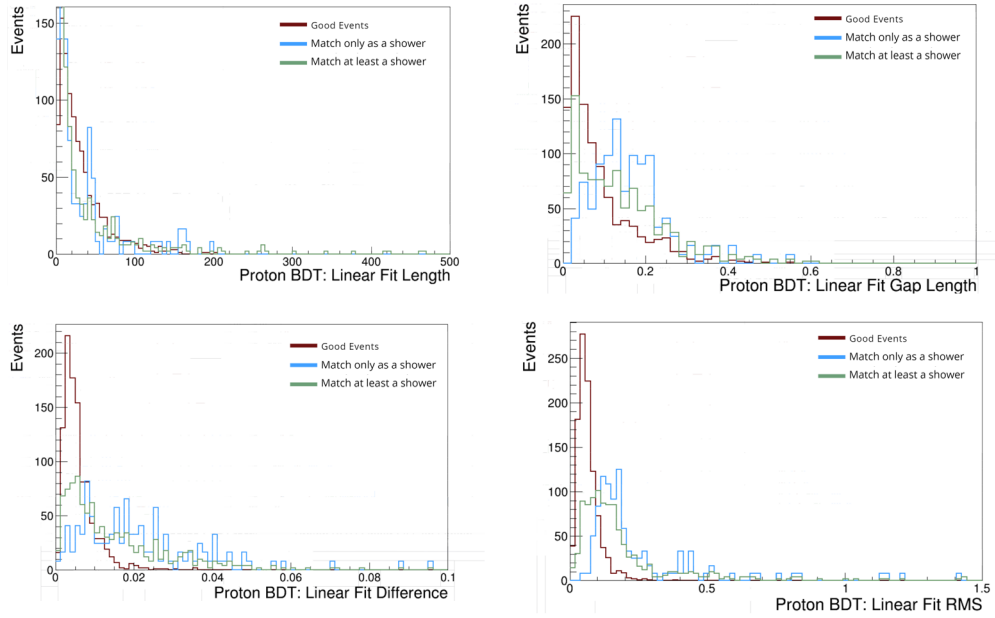


Figure 3.1: BDT variables related to the linear fit. The distributions are normalized to the good sample one

In Fig. 3.2 there are the BDT variables related to the charge. The behavior is similar to the one that I have described above the linear fit but it is even more emphasized, in particular for the charge fractional spread.

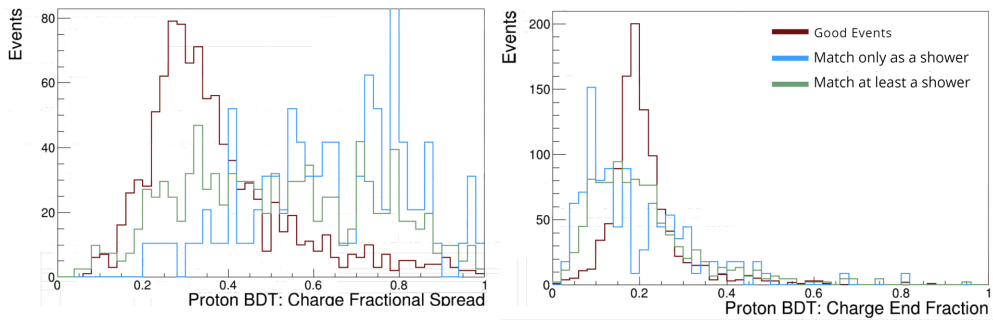


Figure 3.2: BDT variables related to charge. The distributions are normalized to the good sample one

As last, Fig. 3.3 shows the distributions of the remaining BDT variables. In such cases, some differences in the tails of the distribution are evident, but less significant with respect to the linear fit and charge.

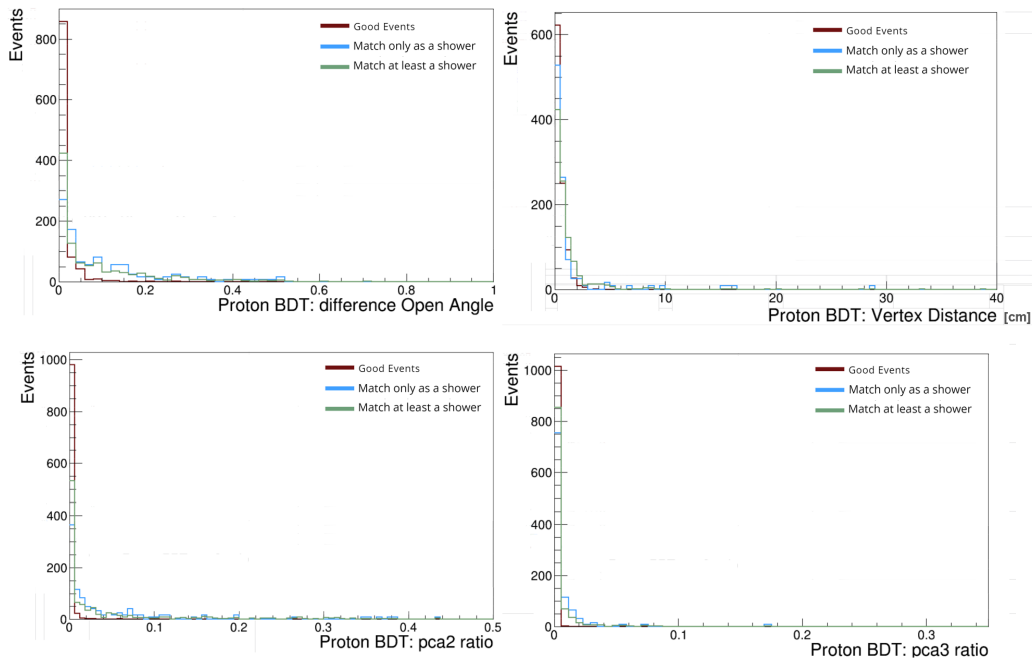


Figure 3.3: BDT variables related to pca and position in the space. The distributions are normalized to the good sample one

3.1.1 2d plots

In the analysis of BDT variables, I also included some 2d plots that could potentially highlight correlations between them. Some examples are shown in Fig. 3.4. Plots (A) in particular are the 2d plots between the two BDT variables related with the charge that, as seen, are the distributions that seems to be more different in the 1d plots. In general, all the plots convey the message that for the badly reconstructed events the distribution is in general wider: also plots (B) and (C) shown the same trend. Sometimes the distribution is not only wider but covers different areas of the parameters space with respect to the good events. An example of this can be the plots (C).

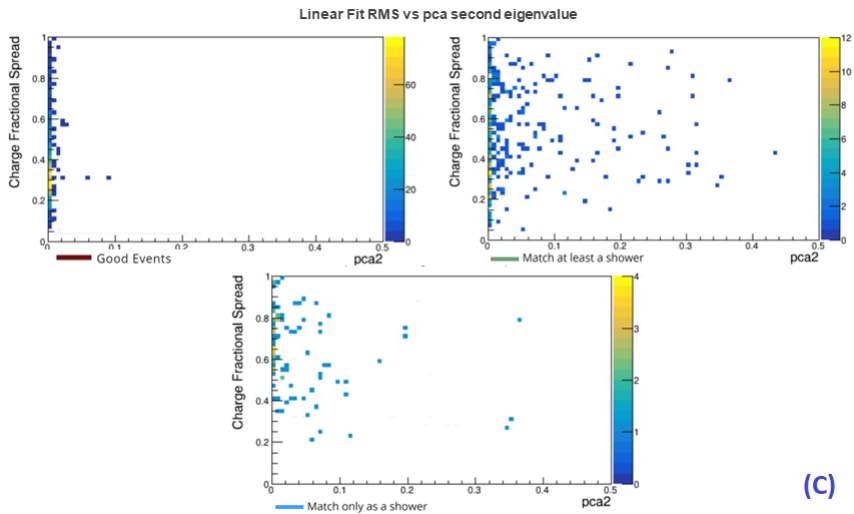
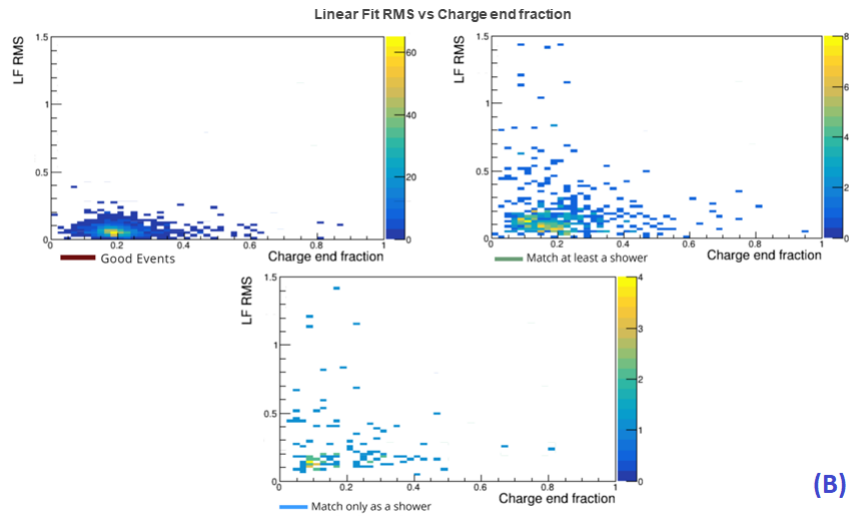
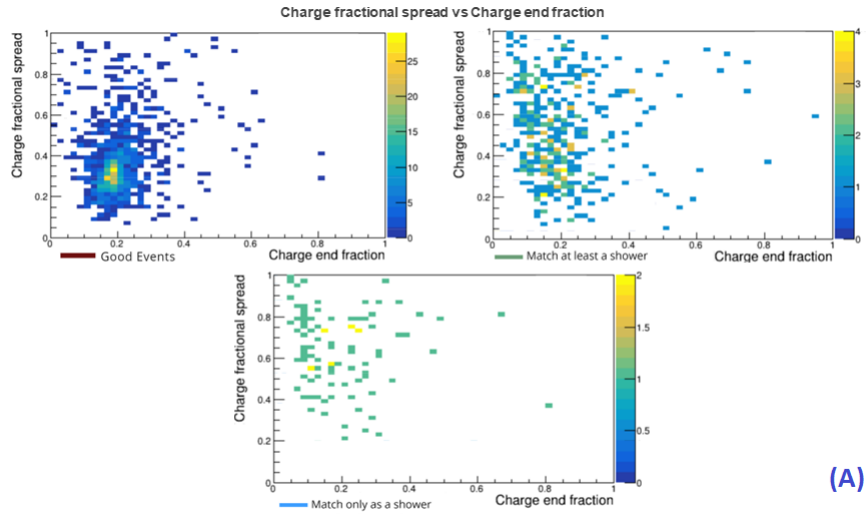


Figure 3.4: Some 2d plot of the BDT variables

Conclusions

The aim of this analysis was to identify a clean sample of well reconstructed events and improve the performance of the BDT for track vs shower discrimination for the protons, especially for the analysis of real ICARUS data. The work that has been done in order to obtain our purpose was made up of three parts:

- First of all, I studied the set of MC simulated data in order to classify the protons that were not identified as tracks, as it should be.
- Secondly, I looked for a best set of data (i.e. well reconstructed) using that as a reference.
- Finally, I compared the BDT variables of the badly reconstructed protons with those extracted from the best set taken as a reference.

As shown in chapter 3 I identified the BDT variables related to the linear fit (excluded the length) and to the charge as the main problematic ones. This analysis gave strong starting input for the further work of the reconstruction group that has to implement the reconstruction system.

Bibliography

- [1] R. Acciarri et al., *A Proposal for a Three Detector Short-Baseline Neutrino Oscillation Program in the Fermilab Booster Neutrino Beam*, (2015).
- [2] Pedro A. N. Machado et al., *The Short-Baseline Neutrino Program at Fermilab*, <https://doi.org/10.1146/> (2019).
- [3] ICARUS, <http://icarus.lngs.infn.it/DetectorOverview.php> .
- [4] Brianna Barbu, *ICARUS gets ready to fly*, <https://news.fnal.gov/2021/05/icarus-gets-ready-to-fly> (2021).
- [5] Sakata Maki, Nakagawa, *Remarks on the Unified Model of Elementary Particles* (Progress of Theoretical Physics, Oxford, 1962), Vol. 28.
- [6] B. Povh et al., *Particles and Nuclei* (Springer-Verlag, Berlin Heidelberg, 2015).
- [7] Yury Kudenko, *New results and perspectives in neutrino physics*, EPJ Web of Conferences 212, 01005, <https://doi.org/10.1051/epjconf/201921201005> (2019).
- [8] Joachim Kopp Basudeb Dasgupta, *Sterile Neutrinos*, Phys. Rept. 928 <https://doi.org/10.1016/j.physrep.2021.06.002> (2021).
- [9] Carlo Rubbia, *The Liquid Argon Time Projection Chamber: A New Concept for Neutrino Detectors*, CERN-EP-INT-77-08 (1977).

- [10] Bruce Howard, *Brief update on track vs. shower BDT scores*, FNAL, internal note (2022).
- [11] Age K. Rasmus Bro, *Principal component analysis*, *Anal. Methods*, 6, 2812-2831 <https://doi.org/10.1039/C3AY41907J> (2014).