FERMILAB-THESIS-2024-22

Dimuon production in the MicroBooNE detector

by

Norman Amilkar Martinez Figueroa

M.S. Universidad de Puerto Rico, UPRM, 2018

AN ABSTRACT OF A DISSERTATION

submitted in partial fulfillment of the
requirements for the degree

DOCTOR OF PHILOSOPHY

Department of Physics
College of Arts and Sciences

KANSAS STATE UNIVERSITY
Manhattan, Kansas

2025

Abstract

This analysis studies the production of the rare signal with two muons (dimuon) in the final state at μB using the Booster Neutrino Beam (BNB) as source of neutrino interactions. The main background source for the dimuon signal are pions. Differentiating muons and pions has been challenging at μB because the tracks left by these particles are almost identical. This analysis faces this problem using a supervised classifier algorithm called boosted decision tree (BDT). A Monte Carlo simulation (MC) enhanced dimuon signal sample was produced to train a BDT capable of recognizing this signal. A set of variables that show some discrimination characteristics was used to perform the training. A MC neutrino interaction simulation sample was used to estimate the background. Estimations of dimuon production in real data samples were made for different cases. Additionally, an analysis of the systematic uncertainties, fake data studies, and BDT performance checks were obtained.

Dimuon production in the MicroBooNE detector

by

Norman Amilkar Martinez Figueroa

M.S. Universidad de Puerto Rico, UPRM, 2018

A DISSERTATION

submitted in partial fulfillment of the
requirements for the degree

DOCTOR OF PHILOSOPHY

Department of Physics
College of Arts and Sciences

KANSAS STATE UNIVERSITY
Manhattan, Kansas

2025

Approved by:

Major Professor
Tim Bolton

Copyright

Limited use of artificial intelligence was employed for grammar correction and generation of Root macros, including Grammarly and ChatGPT.

Abstract

This analysis studies the production of the rare signal with two muons (dimuon) in the final state at μB using the Booster Neutrino Beam (BNB) as source of neutrino interactions. The main background source for the dimuon signal are pions. Differentiating muons and pions has been challenging at μB because the tracks left by these particles are almost identical. This analysis faces this problem using a supervised classifier algorithm called boosted decision tree (BDT). A Monte Carlo simulation (MC) enhanced dimuon signal sample was produced to train a BDT capable of recognizing this signal. A set of variables that show some discrimination characteristics was used to perform the training. A MC neutrino interaction simulation sample was used to estimate the background. Estimations of dimuon production in real data samples were made for different cases. Additionally, an analysis of the systematic uncertainties, fake data studies, and BDT performance checks were obtained.

Table of Contents

List of Figures

List of Tables

Nomenclature

MicroBooNE	Micro Booster neutrino experiment.
LArTPC	Liquid argon time projection chamber.
FullMC	Full Monte-Carlo neutrino interaction simulation.
Mctruth	Monte-Carlo simulated information.
GENIE	Event generator and global analysis of neutrino scattering data.
Fermilab	Fermi National Accelerator Laboratory.
BNB	Booster neutrino beam.
Root	Scientific software designed for particle physics analysis.
POT	Protons on target.
TMVA	Multivariable tool kit in Root.
BDT	Boosted decision tree.
CV	Central value sample.

Acknowledgments

This analysis would not have been possible without the support of Dr. Tim Bolton, who was my advisor for more than five years, his contributions to my professional and personal life are of huge value. Thanks to Kansas State University for being my second home for six years, to the Physics department and the whole University staff. To Dr. Horton-Smith, Dr. Ivanov and, Dr. Korten for being part of my supervisory committee. I would like to extend my thanks to Fermilab and especially to the MicroBooNE collaboration for the help provided to this study, to Mark Ross-Lonergan, Jairo Rodriguez, Afro Papadopoulou, Keng Lin, Diego Andrade, Luciano Arellano, Adil Hussain, Nabila Majeed, Daniel Barrow, Elena Gramellini, and others who provided feedback to this project.

Thanks to my friends in Manhattan, Kendall Borthwick, Lois Borthwick, Julian Amaya, Raquel Amaya, Julian E. Amaya, Sergio Carvajal, Geraldin Sierra, Marvin Tzirin and others who shared hundreds of beautiful moments with me.

Dedication

Dedicated to my Mom Maria Eugenia Figueroa Arellano, my Dad Edgar Enrique Martinez Montaño, my sister Dania Gisela Martinez Figueroa and my brother-in-law Elliot Diaz, my family and my most beautiful source of motivation.

Chapter 1

Introduction

During the last decades, liquid argon time projection chamber (LArTPC) detectors have been used for many neutrino experiments worldwide. For example, in 2015, a proposal for a short-baseline neutrino (SBN) program was submitted to Fermilab in which three detectors would be working with the LArTPC technology. These experiments are MicroBooNE (μB), ICARUS, and SBND [1]. Additionally, the future deep underground neutrino experiment (DUNE) will use a large scale of this kind of detector in its near and far detector [2]. DUNE will be the flagship in neutrino physics studies (among others) during the following several decades in the world.

Many experiments like those in the SBN program have been producing measurements of all kinds of neutrino cross-section interactions and particle productions that will be useful for DUNE.

Events with two muons (μ) in the final state are known as dimuons. This study aims to measure the production of dimuon events in μB by doing a selection of dimuon candidates and comparing real data with simulations. This production will cast in the form:

$$\sigma(\mu\mu) = X\sigma_{SM}(\mu\mu), \tag{1.1}$$

where $\sigma(\mu\mu)$ is the dimuon production measurement made by this study, $\sigma_{SM}(\mu\mu)$ is the standard model (SM) prediction, as implemented through the GENIE event generator for the flux-average dimuon production [3], and X the parameter that is going to be constrained by this analysis. The parameter X is how many times this analysis observed the GENIE prediction. A value of X near 1 would indicate a measurement consistent with the SM expectations. $X > 1$ would indicate dimuon production by non-standard model expectations. The study will not be able to constrain or observe values for this parameter near or smaller than 1 due to the very small size of the SM cross-section for dimuons. Any measurement of a larger value of X would thus be evidence for dimuon production via mechanisms beyond the SM. In the case that no evidence of dimuon signal is observed, this study will measure an upper limit for X. Chapter 2 will explain in detail the reasons why this signal is rare.

DUNE will be exposed to a more intense neutrino flux that can lead to a more evident dimuon production. Being able to measure the dimuon production in μB could show the potential of studying this signal in future projects.

In the present chapter, the next sections will introduce the MicroBooNE (μB) experiment and a basic theoretical framework of the physics behind this thesis. Chapter 2 describes the main focus of this study, the concept of dimuon, and details about the enhanced dimuon MC sample. Chapter 3 will explain the concept of a classifier based on machine learning techniques and the tools needed for searching the dimuon signal.

Chapter 4 explores the different sources of uncertainties present in this problem and how to calculate the errors associated with those. Chapter 5 explains the analysis made using a larger data set. Chapter 6 evaluates the performance of the BDT classifier, the no signal BDT region, and fake data studies. Chapter 7 shows the final conclusions of this study.

1.1 MicroBooNE

The Micro Booster Neutrino Experiment (MicroBooNE or μB) was a 170-ton liquid argon time projection chamber (LArTPC) neutrino detector. It was designed to perform short-baseline neutrino (ν) cross-section studies. One of the first objectives was to explain the anomalous electron-neutrino (ν_e) excess of low energy events observed by MiniBooNE [4]. It used the potential of liquefied noble gases (argon (Ar) in this case) as a detection medium for neutrino interactions. MicroBooNE combined the advantages of high spatial resolution and calorimetry for excellent particle identification with the potential to scale to large volumes [5].

Figure 1.1 shows the liquid argon time projection chamber (LArTPC) dimensions, 2.56 m of height by 2.32 m of width by 10.36 m. It was located at Fermi National Lab (Fermilab) in Batavia, IL, USA Figure 1.2. MicroBooNE took data from 2015 to 2021.

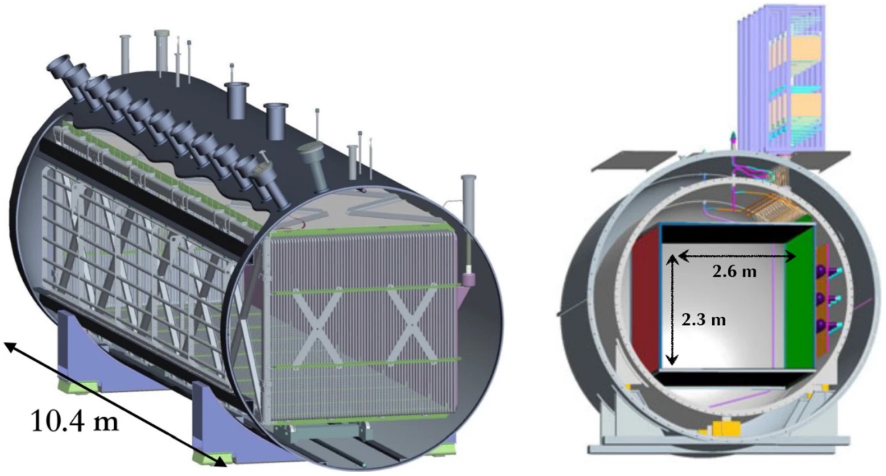


Figure 1.1: *Left: Perspective vision of μB detector, Right: transversal cut with sizes, picture taken from [6].*

In the map (see Figure 1.2) is possible to identify the location of μB in the Fermilab campus. The arrows and solid line show the trajectory that the accelerated protons follow inside the booster neutrino beam (BNB) until they collide with the beryllium target (among others) to produce the resulting neutrino beam. The detector is right in front of the BNB collinear (z-axis) with the beam trajectory.

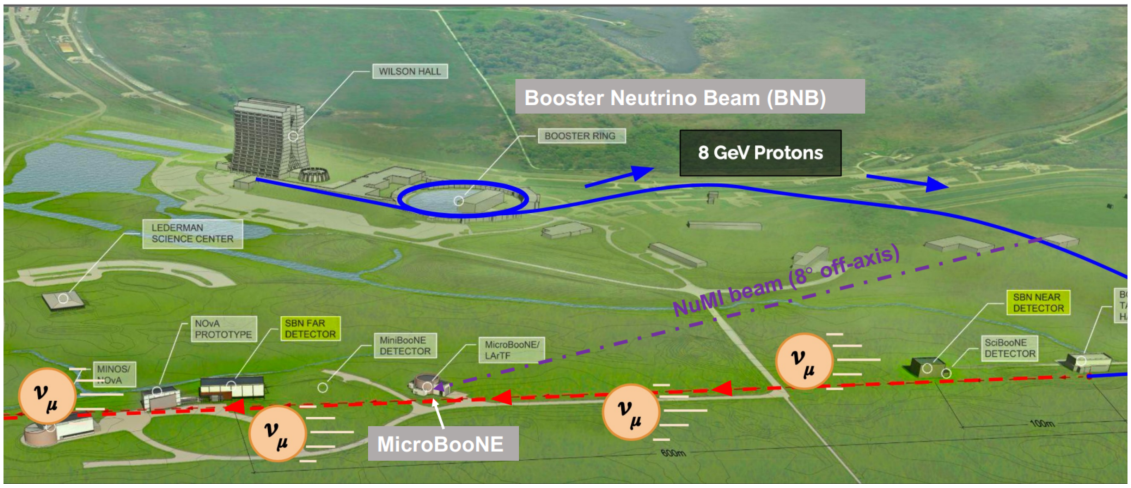


Figure 1.2: *Detector location inside FermiLab campus [5], modified picture from [7].*

1.2 LArTPC

1.2.1 Liquid argon

The argon (Ar) element has 24 isotopes from ^{30}Ar to ^{53}Ar. Three of those isotopes are stables, ^{36}Ar, ^{38}Ar and ^{40}Ar. The most abundant one is ^{40}Ar (around 99.6%) and is relatively economical, which is good for projects where hundreds of tons are required. Liquid argon is used in particle detectors because of its properties. Argon is a noble gas; it almost never interacts with other elements and electrons can move through it without being captured. It is also a scintillator, which means that it produces photons when a charged particle excites its atom [8].

1.2.2 LArTPC

LArTPC is a type of particle detector that uses a combination of an electric field and a sensitive volume of liquid argon to reconstruct particle interactions in three dimensions. Particles with charge passing through a volume of liquid argon leave a series of ionized electrons (e) in their trail. These ionized electrons are not re-captured by the argon atoms because they have the last level of energy fully occupied by electrons (valence shell fully occupied, see Equation 1.2), allowing long drift distances. It is easy to obtain and purify [8].

$$1s^2\ 2s^2\ 2p^6\ 3s^2\ 3p^6. \tag{1.2}$$

Using highly pure liquid argon allows a reconstruction of the trail by capturing those electrons using the drift of an electric field. The ionized electrons are transported to one side of the detector, which has sensing planes (anode plane). On the other side of the detector, there is a cathode plane. The potential difference between the anode and cathode plane creates the drift electric field.

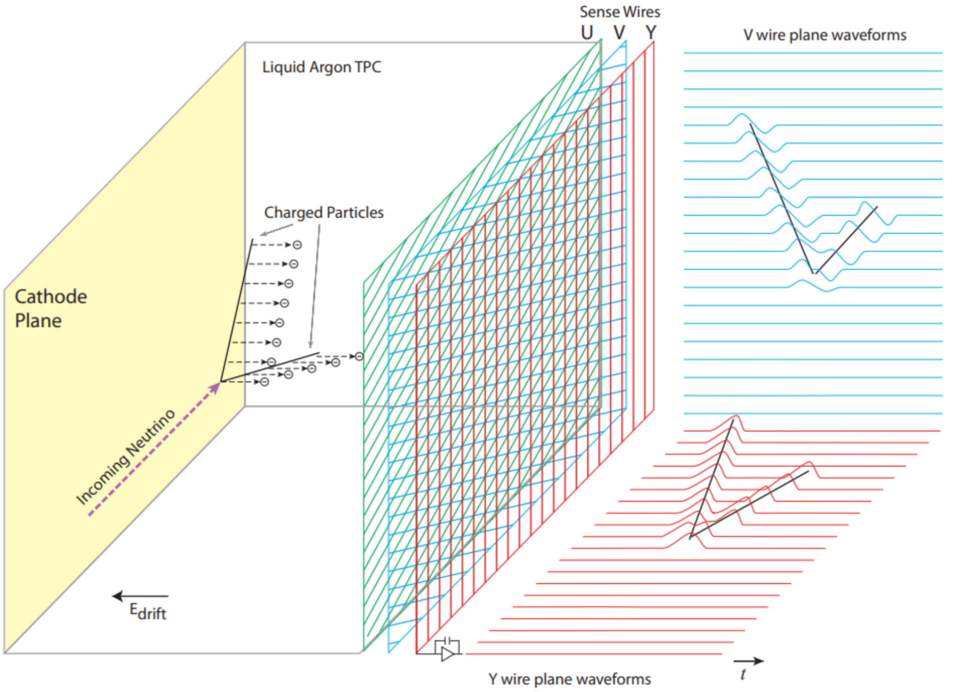


Figure 1.3: *Diagram of operational principle of μB LArTPC [5]. The green and blue wire planes are the induction planes; the red wire plane is the collection plane. On the right side, it is shown the wire waveforms before the reconstruction of the tracks [9].*

The anode plane is composed of three planes (U, V, Y). Each plane has wires that capture the drifted electrons (see Figure 1.3). Planes U, V, and Y differ in the orientation of the composing wires. This allows a grid distribution that helps with the reconstruction of each trail, from now on called track. Depending on the location of the capturing wire in each plane, it is possible to perform a 2-dimensional reconstruction of the track.

During the "flight" of the charged particle, scintillation photons are also produced due to the interaction with the argon. These photons can be captured as well by photomultipliers (PMTs). The difference between the PMT capture and the wire capture provides the time of flight or drift time. This time is used with the drift velocity to calculate the third coordinate of each hit. This completes the 3D reconstruction of the track [5]. Extra details about some parameters of μB LArTPC are shown in detail in Table 1.1.

Parameter	**Value or operative description**
Anode planes	3 (U, V, Y)
Anode planes spacing	3 mm
Wire pitch	3 mm
Wire coating	2μm Cu, 0.1μm Ag
Design wire tension	6.9N
Wires	8256
Induction plane 0 (U) wires	2400
Induction plane 1 (V) wires	2400
Collection plane (Y) wires	3456
Wire orientation	+60°,-60°,0°(U,V,Y)
Cathode voltage (nominal)	-128 kV
Bias voltages (U,V,Y)	-200 V, 0 V, +440 V
Drift-field	500 V/cm
Max drift time, cathode to U	1.6 ms
Field-cage steps	64
Ring to ring voltage steps	2 kV

Table 1.1: *Parameters of μB LArTPC [5].*

1.3 Booster neutrino beam (BNB)

Fermilab's accelerator complex has several particle accelerators that provide proton and neutrino beams for different experiments. The booster neutrino beam (BNB) was used in this study. To produce this beam, protons accelerated (to about 400 MeV) originally in Fermilab's linear accelerator (Linac) enter to Fermilab's booster synchrotron. There, they reach energies of about 8 GeV. These protons are oriented to a Beryllium target, smashing it and producing mesons (pions/kaons) in this interaction [10].

These mesons are then focused using a magnetic horn. Depending on the flavor desired in the neutrino beam, the current that produces the horn's magnetic field could be switched. For example, the horn must allow only positive mesons to pass through it to obtain a muon neutrino beam. For instance, this meson could be π^+. π^+ decays almost exclusively to $\mu^+\nu_\mu$ ($\mathrm{BF}(\pi^+ \to \mu^+\ \nu_\mu) \approx 100\%$) (see Equation 1.3), then a muon-neutrino will be obtained.

$$\begin{aligned} p\ \mathrm{Be} &\to \pi^+\ \mathrm{X}, \\ &\quad \searrow \pi^+ \to \mu^+\ \nu_\mu. \end{aligned} \tag{1.3}$$

This will be the neutrino beam (mainly composed by muon-neutrinos) used for this study. Switching the current and therefore switching the magnetic field, would allow to obtain an antineutrino beam. Additionally, the muons in the final state should be removed. To do that, absorbers are located in the decay pipe to filter out any particle but the neutrinos (which very rarely will interact with it) (see Figure 1.4).

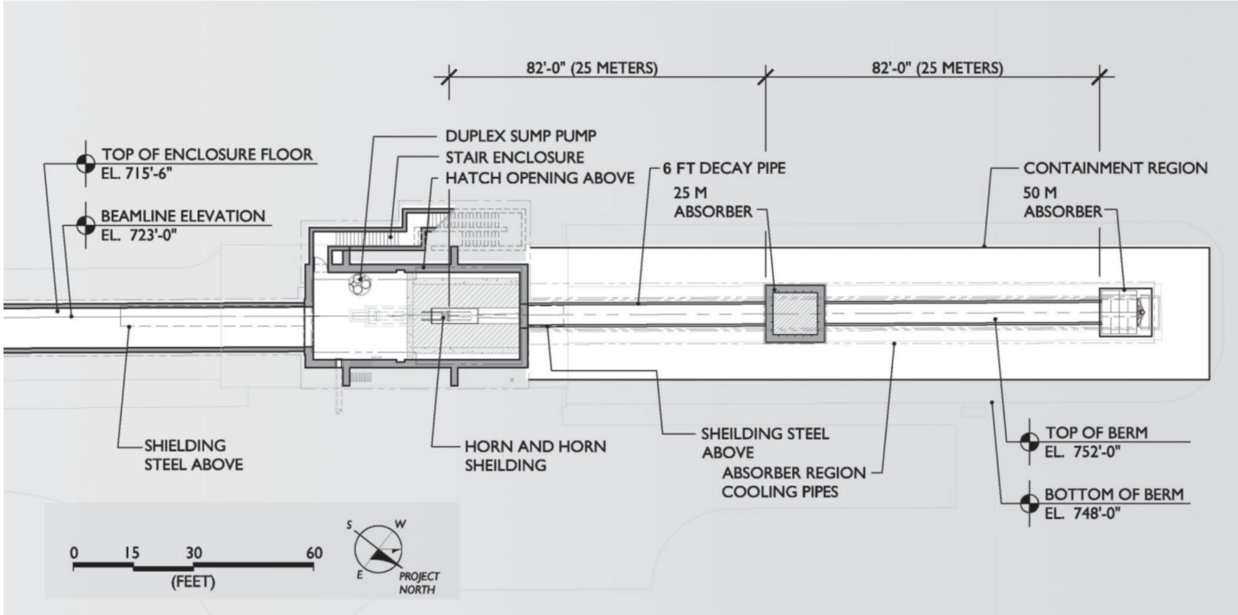


Figure 1.4: *Protons coming for the booster enter from the left and smash the Be target located right before the focusing horn. Any particle but the neutrino will be filtered out by the absorber regions [11].*

The target is a cylinder of about 70 cm in length and 0.5 cm in radius. Beryllium was chosen motivated by residual radioactivity issues under the hypothetical event of a replacement and energy loss considerations. The focusing horn is a pulsed toroidal electromagnet of an aluminum alloy. It operates with a nominal peak current of 170 kA. The magnitude of the magnetic field is inversely proportional to the radius inside the horn, which explains the fact that its value is maximum in the narrowest part of the horn with a value of 1.5 T [11]. Additionally, a collimator right outside the horn absorbs hadrons that are not useful for the neutrino beam production. The decay pipe has 50 m of length and 1 m of radius. At the end of the decay pipe an absorber of steel and concrete is located, it additionally can count any long-lived muon that may pass through for beam purity analysis.

1.4 Standard Model of Particle Physics (SM)

The standard model (SM) of particle physics is one of the biggest achievements in the history of modern physics. It classifies the known elemental particles in three families, leptons, quarks, and bosons (see Figure 1.5) with respect to electromagnetic, weak (later unified in an electroweak theory), and strong interactions (quantum chromodynamics). It is a model that looks to describe experimental data using deep theoretical ideas. The main elements of the SM are [12]:

- Dirac equation of relativistic quantum mechanics, describing dynamics of the fermions,
- Quantum field theory (QFT), with a description of the particles and how they interact,
- Higgs mechanism, explaining how particles acquire their masses,

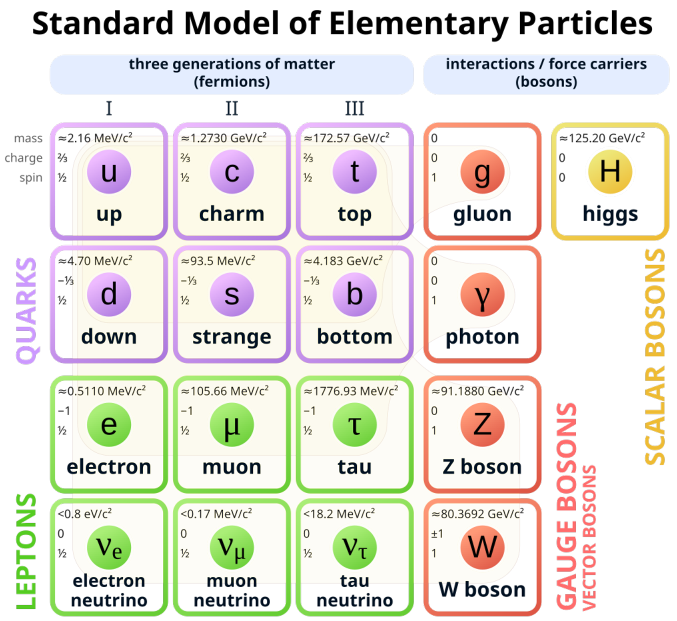


Figure 1.5: *Standard model of particle physics, picture from Wikimedia commons.*

In 2012 Higgs boson was detected for the first time at the European organization for nuclear research (CERN) [13] completing the observation of the 12 fermions and 5 bosons that conform the SM.

The SM has been very successful and has predicted many experimental observations, however, there are still several questions related to it without answer at the moment [14]:

- Why are there three elemental forces? (gravity is not included in the SM, given that it is not quantized, among other reasons).
- Why are there three quark and lepton generations and no more?
- In general, particles of different generations share the same characteristics, but why are their masses that unalike?
- What are the masses of the neutrinos? Are they Majorana particles (its own anti-particle)?
- What is the dark matter?
- What is the origin of charge-parity (CP) violation?

These and many more questions have been tried to be explained with beyond standard model (BSM) theories.

1.5 Charged particles in μB

1.5.1 Muons in μB

Muons have a minimum ionizing particle (MIP) behavior. Their energy loss could be consider almost constant throughout their displacement in the detector. An approximated value is:

$$\left(\frac{dE}{dX}\right)_{\mu} \approx 2.1 \text{ MeV/cm}. \tag{1.4}$$

This value increases sharply in the last few centimeters of the track, producing a Bragg peak. Muons move in relatively straight lines that, because of multiple Coulomb scattering (MCS), can get some smooth curvature. Also, at the end of their track, they can decay:

$$\begin{aligned} \mu^- &\rightarrow \nu_\mu \; e^- \; \overline{\nu}_e, \\ \mu^+ &\rightarrow \overline{\nu}_\mu \; e^+ \; \nu_e. \end{aligned} \tag{1.5}$$

From this decay the resulting Michel has an energy in the 10 - 50 MeV range. In this range, electron is also a MIP for most of its trajectory and could travel 5 - 25 cm. Negative muons can be captured by Ar nucleus in the process:

$$\mu^- \; p \rightarrow \nu_\mu \; n. \tag{1.6}$$

This usually breaks the nucleus apart, producing extra energy at the end of the track with no Michel electron coming out from the capture. That does not happen with positive muons.

1.5.2 Other charged particles in μB

With rare exceptions (kaons and nuclear fragments like deuterons) all charged particles seen in the μB detector are either electrons, muons, pions or protons.

Electrons tend to produce electromagnetic showers when their energies are over 50 MeV. This study is not interested in showers, so one first requirement in the next section would be to work exclusively with tracks.

Protons lose energy at a much higher rate than muons and pions. They also experience hadronic interactions that produce large energy losses and changes in their trajectory producing not smooth tracks.

Pions as will be seen in Section 2.1.3, are the main source of background for this study. This is because when they do not suffer hadronic interactions, they look just like muons. Still, they have a wide interaction cross-sections over usual energies at μB. Their tracks experience changes in direction from nuclear scattering. The production of secondary particles and energy losses are larger compared to the muon case. Pions with energies above 150 MeV are more likely to nuclear scatter than to produce long tracks. Despite this, pions and muons are so closely related in μB that their separation is the most important aspect of this study.

1.6 Neutrino interactions

In the SM, the neutrino is an elementary particle with spin 1/2 that can interact only through weak and gravity forces. In general, there are four main charged current neutrino interactions. These are explained in the next sections, with a fifth case (coherent pion production) important for this study because it has a small contribution to the enhanced dimuon signal sample.

1.6.1 Quasi-Elastic (QE)

The first detection of a neutrino (antineutrino in this case) was performed in 1956 in the so-called Cowan–Reines neutrino experiment [15]. The main idea was to study the inverse beta decay:

$$\overline{\nu}_e + p \rightarrow n + e^+. \tag{1.7}$$

This first detection is an example of a quasi-elastic neutrino interaction. The incoming neutrino energy required for one case like this to happen is relatively low. The term "quasi" comes from the fact that the interaction is almost elastic because a target nucleon is "impacted" and a nucleon is got in the final state. The difference is that in the leptonic side of the interaction, the charged-current produces a different lepton than the original (similarly the hadron in the hadronic side of the interaction).

MicroBooNE is dominated by the quasi-elastic mode with around a 45% of cross-section (CC case) (see Figure 1.6):

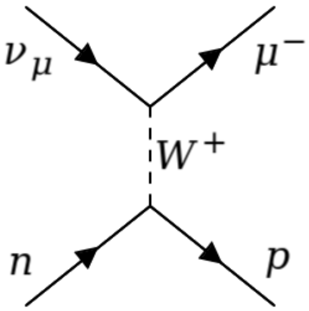


Figure 1.6: *Example of a QE neutrino interaction mode.*

The incoming ν energy required for the process in Figure 1.6 to happen is $E_\nu >\approx 113$MeV ($E_\nu >\approx 110$MeV for $\nu_\mu\ p \rightarrow \mu^+\ n$).

1.6.2 Meson exchange current (MEC)

The MEC case is similar to QE, but the neutrino interacts with a correlated pair of nucleons (instead of a single nucleon in QE mode). It is also called “2p2h” scattering [17]. In this case, a weak meson is exchanged between the incoming neutrino and a pair of nucleons (see Figure 1.7), producing the emission of two nucleons.

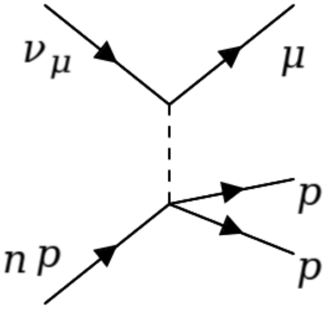


Figure 1.7: *Example of MEC neutrino interaction mode.*

1.6.3 Resonance (RES)

It can occur as well that a nucleon (a proton or a neutron) can be excited by the neutrino interaction; therefore it can create different particles in the de-excitation process.
Some examples of the particular case of delta resonance are:

$$\begin{aligned} \nu_\mu + p \to \mu^- \ \Delta^{++} \to p + \pi^+, \\ \nu_\mu + n \to \mu^- \ \Delta^{+} \to p + \pi^0, \\ \nu_\mu + n \to \mu^- \ \Delta^{+} \to n + \pi^+. \end{aligned} \tag{1.8}$$

The last mode in Equation 1.8 is shown as a Feynman diagram in Figure 1.8.

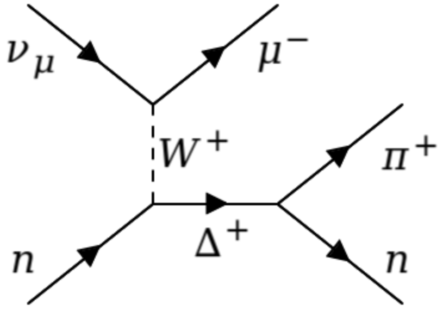


Figure 1.8: *Example of RES neutrino interaction mode.*

One particular case is η production. The particle η is a meson made of up, down and strange quarks and their correspondent antiquarks. It is an important case for this study because it can lead to dimuon production (see Section 2.1.2). In this case, a heavy resonance baryon is produced, for example, N(1535):

$$\nu_\mu \ N \to \mu^- \ N(1535) \to N \ \eta. \tag{1.9}$$

1.6.4 Deep inelastic scattering (DIS)

In this case, an energetic enough incoming neutrino is able to break up the bound state of a nucleus. As a result, different hadrons are obtained in the final state. The incoming neutrino energy required for this case to happen is relatively high.

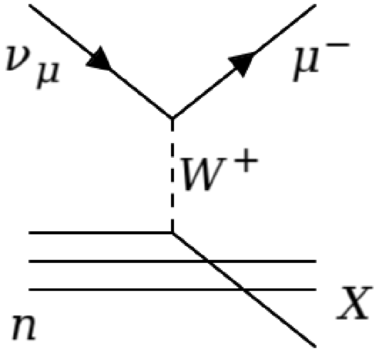


Figure 1.9: *Example of DIS neutrino interaction mode.*

Similarly to Section 1.6.1, the DIS mode can produce charmed particles. It requires at least $E_{\nu,DIS} > 2802$MeV [16]. This value is higher than the threshold for the QE charm production given that in DIS mode additionally to the baryon in the final state, at least a meson is produced. Charm production represents a significant fraction of the charged current (CC) cross-section [18] for DIS mode. Section 2.1.2 will explore more about this production and its relation to the dimuon production at μB.

1.6.5 Coherent (Coh)

Another mode for neutrino interaction is the coherent neutrino-nucleus scattering mode. Not many measurements have been performed due to its small cross-section. In this process, the nucleus is not broken, and mesons can be obtained in the final state (see Figure 1.10).

As mentioned before, GENIE is the main theoretical framework used in this study. For the case of coherent production, it uses the PCAC theorem with the Rein-Sehgal model [19], which produces pions in the same direction that the neutrino beam.

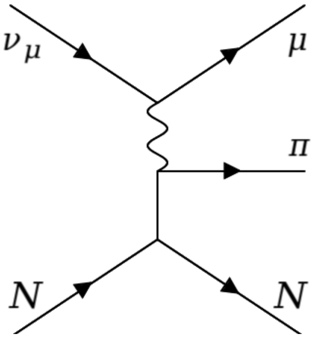


Figure 1.10: *Example of Coh neutrino interaction mode.*

1.6.6 Distribution of ν interaction modes in MC simulations

The distribution of ν interaction modes in a sample with simulated neutrino interaction events for μB shows the following behavior:

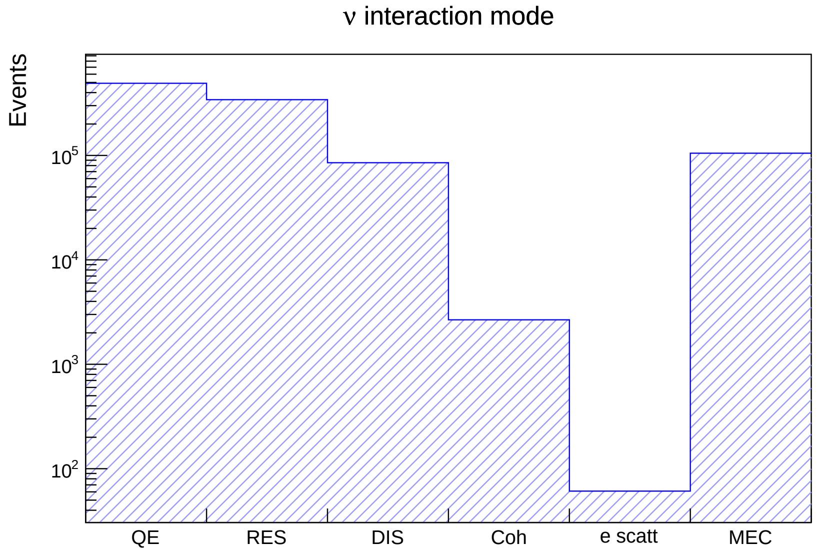


Figure 1.11: *ν interaction modes for events simulated at μB (Log scale in y-axis).*

Where ν is the correspondent symbol for neutrino, and "e scatt" is the abbreviation of electron scattering, a kind of neutrino-electron interaction. This plot was generated using available truth information from GENIE. Section 2.4.2 will introduce the concept of full Monte-Carlo neutrino interaction simulation for run 1. This is the sample used for producing Figure 1.11.

As expected, the QE mode dominates at μB energies, followed by the RES, MEC, and DIS modes. Additionally, there is a very small contribution from the coherent and electron scattering modes. Electron scattering may occur, but its cross-section is tiny. In this case, the neutrino exchanges a weak force boson with the electron, producing this last to scatter.

Chapter 2

Dimuon signal, Monte Carlo, and data samples at μB

2.1 Dimuon signal at μB

2.1.1 Signal definition

For this study a dimuon event is defined as a final state in a neutrino interaction in which there are at least two muons.

$$\begin{aligned} &\nu_\mu \text{ Ar} \rightarrow \mu\ \mu\ \text{X}, \\ &\overline{\nu}_\mu \text{ Ar} \rightarrow \mu\ \mu\ \text{X}. \end{aligned} \tag{2.1}$$

Where Ar is the Argon (Ar from now on) nucleus and “X” could be any particle.
It is a rare mode at MicroBooNE (μB). In the full Monte-Carlo (MC) neutrino interaction simulation for run 1 of μB (fullMC henceforth), a total of 10 dimuon events are observed out of ~1 million events.

In Figure 2.1, a generated dimuon event is shown. The neutrino interaction produces a μ^- in the leptonic side of the interaction and a Σ_c^{++} charmed baryon is produced and decays to $\Sigma_c^{++} \rightarrow \Lambda_c^+ \pi^+$ and subsequently $\Lambda_c^+ \rightarrow \Lambda^0 \mu^+ \nu_\mu$, which results in a dimuon in the final state. Two long tracks are observed that come from the same vertex forming a big angle. This event also contains a small track corresponding to the π^+ generated in the initial Σ_c^{++} decay. In general, this is going to be the geometry that this study is going to be pursuing.

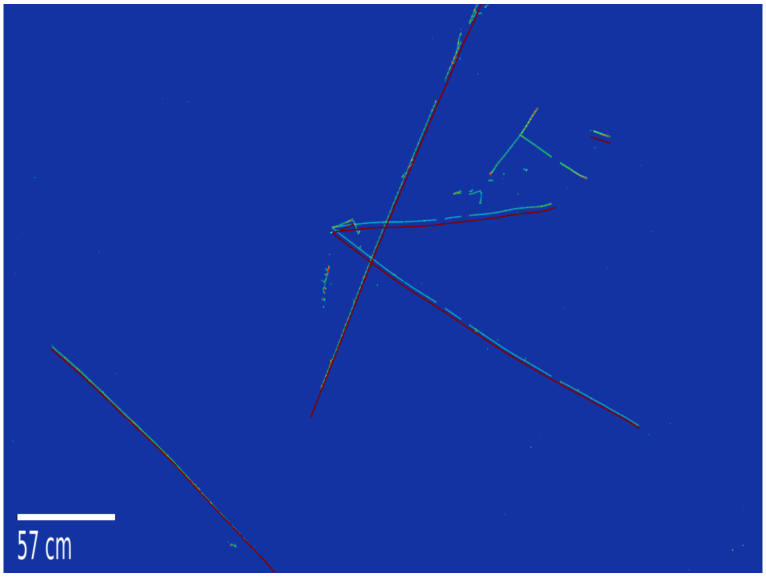


Figure 2.1: *Event from dimuon signal sample, 73355 run.subrun: 7054.1467.*

The main objective of this study is to investigate whether dimuon production is consistent with the standard model (SM) predictions. Finding an excess of dimuon production could potentially lead to beyond standard model (BSM) studies related to this signal. This would strongly depend on there being no significant backgrounds. The topology that this study looks for is:

- At least two tracks.
- Tracks that look like muons:
  - Long tracks.
  - Low energy loss (dEdx).
  - Moderate scattering.

2.1.2 Dimuon production Processes in ν interactions at μB

The dimuon signal is rare because most of the processes that can generate it share the fact that the baryon or meson in the hadronic side of the interaction is heavy, which requires high energy from the incoming neutrino in order to be produced. The main known SM modes of dimuon production are listed as follows:

Dimuons from charm modes production

The Λ_c^+ is the lightest isospin I = 0 baryon with a single charm quark ($u\ d\ c$), while the Σ_c baryons are I = 1, baryon with a single charm quark.

At μB energies, dimuon production is dominated by the production of charmed particles like Λ_c^+, Σ_c^+,Σ_c^{++} etc (see Figure 2.2). For example:

$$\begin{aligned} &\nu_\mu \text{ Ar} \to \mu^- \ \Lambda_c^+, \\ &\qquad\qquad \hookrightarrow \Lambda_c^+ \to \Lambda^0 \ \mu^+ \ \nu_\mu. \end{aligned} \tag{2.2}$$

From the Particle Data Group (PDG) [20] the lifetime of Λ_c^+ is $\tau_{\Lambda_c^+} = (2.00\pm0.06)\times10^{-13}s$. This explains why the tracks in Figure 2.1 are visually coming from the same vertex because the charmed particle itself lives for a very short period of time, leaving no track.

To produce this interaction, the incoming neutrino energy E_ν must be at least 2573 MeV (although nucleon motion within the Ar nucleus can lower it). Comparing that value with the mean Booster Neutrino Beam (BNB) energy of 800 MeV explains why this channel (and therefore dimuon production) is widely suppressed compared to non-charmed modes.

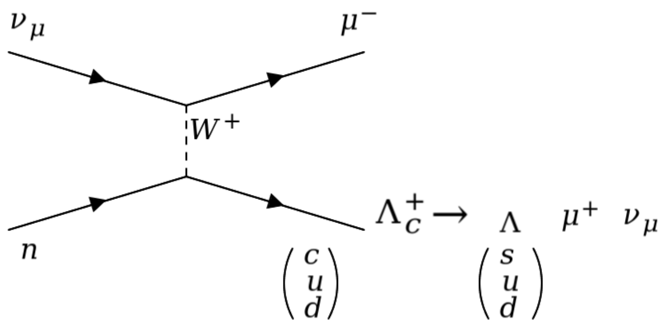


Figure 2.2: *Feynman's diagram for Λ_c^+ production and subsequent dimuon obtention.*

A second reason for a small production rate is the CKM suppression of the d-quark to c-quark transition. This is similar to the suppression of Λ baryon production in antineutrino scattering [21].

Additionally, this channel is also suppressed (like all SM modes involving dimuon production) because the branching fraction (BF) of Λ_c^+ to produce the second muon is small [20],

$$BF(\Lambda_c^+ \rightarrow \Lambda^0 \ \mu^+ \ \nu_\mu) = 3.5\%, \tag{2.3}$$

which makes even smaller the probability of getting a dimuon final state. There are other possible final states for Λ_c^+ decay that includes muons, but their contribution is negligible. The mode shown in Equation 2.3 dominates [20].

Dimuons from η production

Reference [22] shows that η production has been observed at μB. The η meson has a lifetime of $\tau_\eta = (5.0 \pm 0.3) \times 10^{-19} s$ and it usually decays to final states containing photons or pions, but it has a small BF to a dimuon pair $\eta \rightarrow \mu^+ \mu^-$:

$$BF(\eta \rightarrow \mu^+ \mu^-) = 5.8 \times 10^{-6}\%. \tag{2.4}$$

While the dimuon BF is small, the relatively low 548 MeV η mass results in this dimuon mode being non-negligible compared to charmed baryon modes.

the η production is assumed to be dominated by a resonance channel, although details are poorly known.

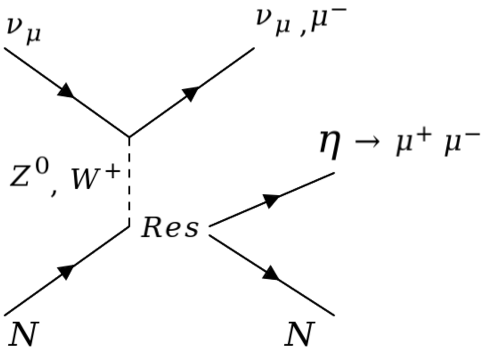


Figure 2.3: *Feynman's diagram for η production and subsequent dimuon obtention.*

The neutral current (NC) neutrino interaction mode is:

$$\begin{aligned} &\nu_\mu\ N \rightarrow \nu_\mu\ N(1535), \\ &\qquad \hookrightarrow N(1535) \rightarrow N\ \eta, \\ &\qquad\qquad \hookrightarrow \eta \rightarrow \mu^+\ \mu^-. \end{aligned} \tag{2.5}$$

The charged current (CC) neutrino interaction mode is:

$$\begin{aligned} &\nu_\mu \; N \rightarrow \mu^- \; N(1535), \\ &\qquad\qquad\qquad \looparrowright N(1535) \rightarrow N \; \eta, \\ &\qquad\qquad\qquad\qquad\qquad\qquad \looparrowright \eta \rightarrow \mu^+ \; \mu^-. \end{aligned} \tag{2.6}$$

Figure 2.3 shows that in the CC case, a final state composed of three muons is possible. The case for N(1535) is indicated here because is the one present in the events in the fullMC with $\eta \rightarrow \mu^+\mu^-$. N(1535) is a baryon resonance particle, an excited baryon state with a short lifetime. The number in parenthesis indicates its mass in MeVs. Figure 2.3 shows these resonances as “Res” for the cases with different baryon resonances that produce η particles.

Dimuon from D^+ meson

The D mesons contain charm quarks. In μB, these are likely produced in deep inelastic scattering (DIS) processes at ν high energy interactions. D^+ is composed by ($c\ \overline{d}$) quarks and can decay to μ^+ + X (anything). Then, it is possible to obtain dimuons with the following mode:

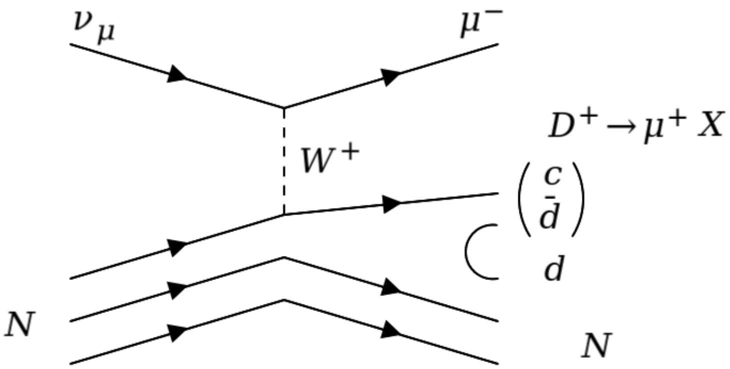


Figure 2.4: *Feynman's diagram for D^+ production and subsequent dimuon obtention.*

Where the quark d in a nucleon goes to c and then recombines with a $\overline{d}$ from a pair $d\overline{d}$ spontaneously produced from vacuum. That produces a $D^+ \to \mu^+\ X$, where X can be any other particles. Then a dimuon is present in the final state. The following equation presents the BF for the D^+ meson decay mode that may produce a muon.

$$BF(D^+ \to \mu^+\ X) = 17.6\%. \tag{2.7}$$

The muon from the leptonic side of the interaction and the muon from D^+ decay configures a dimuon.

Dimuon from D^0 meson

The D^0 is composed by ($c\ \overline{u}$). Similarly to the D^+, is also likely to be produced by DIS processes in μB at ν high energy interactions. Dimuons can be produced with the process:

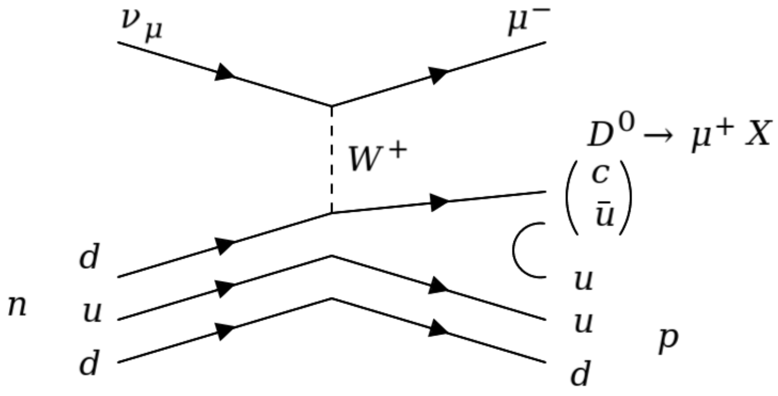


Figure 2.5: *Feynman's diagram for D^0 production and subsequent dimuon obtention.*

Where a d-quark from a neutron, for example, goes to c and recombines with a $\overline{u}$ from a pair $u\overline{u}$ spontaneously produced, obtaining a D^0 that can decay to μ and more particles. This process has a BF:

$$BF(D^0 \to \mu^+\ X) = 6.8\%. \tag{2.8}$$

The muon from the leptonic side of the interaction and the muon from D^0 configures a dimuon.

Dimuon from D_s^+ meson

D_s^+ is the pseudoscalar $(c\ \bar{s})$ meson. Its lifetime is $\tau_{D_s} = (5.00 \pm 0.07) \times 10^{-13} s$ [20]. Similar to the first case of dimuon production (see Section 2.1.2), one μ^- is obtained from the leptonic side of the interaction, and the second one comes from the last decay of D_s^+ in the hadronic side.

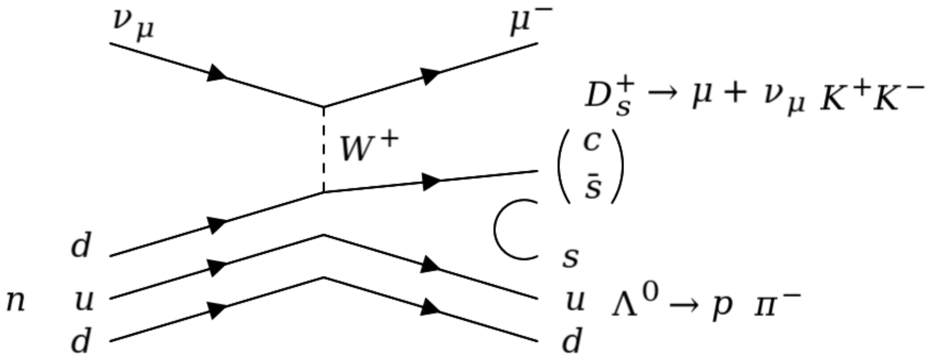


Figure 2.6: *Feynman's diagram for D_s^+ production and subsequent dimuon obtention.*

The relatively high charmed meson mass should strongly suppress this mode. However, it is not CKM suppressed. This mode is present in the GENIE menu of generation modes, and there are several events in the dimuon signal sample (see Table 2.4).

Dimuons from coherent D_s^{*+} production

D_s^{*+} is the excited vector ($c\ \bar{s}$) meson. The following interaction can produce dimuons:

$$\begin{aligned} &\nu_\mu \text{ Ar} \to \mu^- \text{ D}_\text{s}^{*+} \text{ Ar}, \\ &\qquad \looparrowright D_s^{*+} \to \gamma\ D_s^+, \\ &\qquad\qquad \looparrowright D_s^+ \to \mu^+\ \nu_\mu\ X. \end{aligned} \tag{2.9}$$

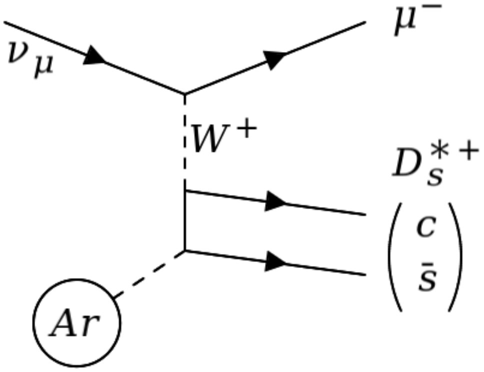


Figure 2.7: *Dimuons from coherent D_s^* production.*

The probabilities for the processes in equation 2.9 to occur are:

$$BF(D_s^{*+} \to \gamma\ D_s^+) = 100\%, \tag{2.10}$$

$$BF(D_s^+ \to \mu^+\ \nu_\mu\ X) = 6.3\%. \tag{2.11}$$

This mode is not present in the GENIE menu.

Dimuons from coherent/incoherent neutral vector boson production

A set of dimuon modes could occur through coherent or incoherent neutral vector boson V^0 production:

$$\begin{array}{cc} \textbf{Coherent} & \textbf{Incoherent} \\ \nu_\mu \text{ Ar} \rightarrow \nu_\mu \text{ V}^0 \text{ Ar} & \nu_\mu \text{ Ar} \rightarrow \nu_\mu \text{ V}^0 \text{ X} \\ \quad\quad \hookrightarrow V^0 \rightarrow \mu^+ \ \mu^- & \quad\quad \hookrightarrow V^0 \rightarrow \mu^+ \ \mu^-. \end{array} \tag{2.12}$$

Possible V^0 are ρ^0, ω, ϕ and J/ψ. The BFs to dimuons are very small, though (see Equations 2.13, 2.14 and 2.15). The only one that is not small is for J/ψ case (see equation 2.16), but its mass is large $M_{J/\psi} = 3.1 GeV$, implying the incoming neutrino energy required would be very high, compared to the BNB mean energy (similar to the charm production case, section 2.1.2).

$$BF(\rho^0 \rightarrow \mu^+ \ \mu^-) = 4.7 \times 10^{-5}\%. \tag{2.13}$$

$$BF(\omega \rightarrow \mu^+ \ \mu^-) = 7.4 \times 10^{-5}\%. \tag{2.14}$$

$$BF(\phi \rightarrow \mu^+ \ \mu^-) = 2.9 \times 10^{-4}\%. \tag{2.15}$$

$$BF(J/\psi \rightarrow \mu^+ \ \mu^-) = 6\%. \tag{2.16}$$

Dimuons from tridents

Tridents from higher-order electroweak processes where there is no nuclear breakup of the Ar atom can produce dimuons (see Figure 2.8). These processes have a very small cross section [23].

$$\nu_\mu \text{ Ar} \rightarrow \mu^+ \ \mu^- \ \nu_\mu \text{ Ar}, \qquad \overline{\nu}_\mu \text{ Ar} \rightarrow \mu^+ \ \mu^- \ \overline{\nu}_\mu \text{ Ar}. \tag{2.17}$$

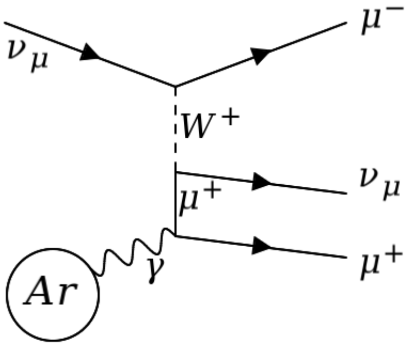


Figure 2.8: *Higher order electroweak process trident.*

Trident production has yet to be observed at μB with BNB energies.

As a summary, Table 2.1 shows the different BFs for the modes where a muon can be obtained from the hadronic side of the neutrino interaction.

Mode	**Branching ratio (%)**
$\lambda_c^+ \to \lambda^0\ \mu^+\ \nu_\mu$	3.5
$\eta \to \mu^+\ \mu^-$	5.8×10^{-6}
$D^+ \to \mu^+\ X$	17.6
$D^0 \to \mu^+\ X$	6.8
$D_s^+ \to \mu^+\ \nu_\mu\ X$	6.3

Table 2.1: *Branching fractions for the modes in Section 2.1.2, PDG [20].*

The approximated energy required for each mode is presented in Figure 2.10. The first peak corresponds to at least ~ 1 GeV to get η production. The remaining modes are obtained starting from ~ 2.6 GeV.

2.1.3 Backgrounds

Backgrounds could involve final states with secondary muons, where the primary muon comes from the neutrino interaction and the secondary from a pion or kaon decay. Also, a muon from the neutrino interaction could be paired with a proton or, more commonly, with a pion that is misidentified as a muon. Finally, cosmic ray events could be wrongly miss-reconstructed as dimuons under some circumstances. The following sections list the main sources of background for this study:

Protons (p)

While protons have a very different signature than muons, their sheer number requires they be treated carefully. Section 3.1 will show the amount of protons at μB. Protons are the most abundant particle at μB, given that the QE mode (MEC mode as well, see Figures 1.6 and 1.7) neutrino interaction is dominant at μB with over 45% of the total probability of events being CCQE (see Section 1.6).

$$\nu_\mu \ \mathrm{Ar} \rightarrow \mu^- \ \mathrm{p}. \tag{2.18}$$

Pions (π)

The main source of background for this study will be pions. Pion mass is so close to the muon mass that electromagnetic interactions like ionization and multiple Coulomb scattering (MCS) are almost identical. Then, it is probable that the reconstructed tracks left by pions can be very similar to the ones left by muons, this makes differentiate these particles very difficult.

But the pion, unlike the muon, has strong interactions, which can cause both wide angle scatters and inelastic interactions in the detector. Pion can decay in flight $\pi^+ \rightarrow \mu^+\nu_\mu$ or $\pi^- \rightarrow \mu^-\overline{\nu}_\mu$, it basically turns into a muon. This is relatively rare due to the long pion lifetime, but can produce a secondary muon that could be misidentified as part of a dimuon.

$$\begin{aligned} &\nu_\mu \text{ Ar} \rightarrow \mu^- \; \pi^\pm \text{ X}, \\ &\qquad\qquad \searrow \pi^+ \rightarrow \mu^+ \; \nu_\mu, \\ &\qquad\qquad \searrow \pi^- \rightarrow \mu^- \; \overline{\nu}_\mu. \end{aligned} \tag{2.19}$$

Alternatively, it is possible to misidentify the π track as a μ while it re-interacts if the pion does not have a strong interaction.

$$\begin{aligned} &\nu_\mu \text{ Ar} \rightarrow \mu^- \; \pi^\pm \text{ X}, \\ &\qquad\qquad \searrow \pi^\pm \text{ Ar} \rightarrow \text{X}. \end{aligned} \tag{2.20}$$

If the pion comes to rest through ranging out, it looks just like a muon except for details at the end of the track, which may be hard to observe. The ability to discriminate pions from muons comes from the pion strong interaction. Because the pion-argon scattering cross section is large, secondary interactions are common.

In both cases, pions would create an excess of muons detected from the hadronic side of the interaction and, therefore, many wrongly assigned dimuons (given that charged currents will already have a muon in the leptonic side). Section 3.3.3 will show how to mitigate this source of background.

Cosmic rays

Cosmic rays can produce misidentified dimuons when a single muon track is “split” in the middle into two different tracks as a misreconstruction. These tracks would appear as “back to back” muons and would be rejected using one requirement (see Section 3.2) that does not allow events where the two forming tracks form a large angle to be considered dimuons. Section 3.2 shows the exact cut applied to guarantee this. Also, in Section 2.4.3, there are details about the sample used to study this effect and how to include it in the MC in order to compare it with real data.

2.2 Dimuon, Monte Carlo, and data samples

Initially, five samples were used for this analysis:

Sample	**Number of events**
Dimuon signal MC sample 1	15286
Dimuon signal MC sample 2	4082
Full ν interaction MC run 1 sample	1026120
Open data set run 1 sample	176429
Cosmic rays (off-beam) run 1 sample	820416
Dirt MC run 1 sample	104572

Table 2.2: *Number of events in the initial samples.*

With the following proton on target (POT) and spills:

Sample	**POT**
Dimuon signal MC sample 1	5.480298×10^{24}
Dimuon signal MC sample 2	1.419532×10^{24}
Full ν interaction MC run 1 sample	1.28344×10^{21}
Open data set run 1 sample	4.566×10^{19}
Dirt MC run 1 sample	3.23362×10^{20}
Sample	**Spills**
Cosmic rays (off-beam) run 1 sample	28190365
Open data set run 1 sample	10127594

Table 2.3: *POTs for each initial sample. For the off-beam cosmic ray sample, the number of spills is used.*

The POT is a measurement of the number of protons that were used at the early stage of proton acceleration that subsequently produces the neutrino beam used for the neutrino interaction with the detector. This amount is important because it describes how proportional a sample can be with respect to others and how it is possible to normalize samples in order to be able to compare them (see Section 3.3.7). The spills are periods of time happening during the exposition of the detector. This amount characterizes the off-beam sample because it gives a magnitude of comparison with data (see Table 2.3, last two rows).

The open data set run 1 and cosmic rays (off-beam), often called EXT samples, were available and provided by the gLEE group [24]. The dimuon signal and the fullMC ν interaction (run 1) samples had to be produced; in the case of the dimuon sample, it had to be from the beginning, starting with a GENIE filter through all the steps of generation (see Appendix B Sections A.2 - A.4.4).

2.3 Dimuon sample description

The enhanced dimuon MC is a sample of simulated events with only dimuons. It was produced because the available MC samples have just a few events. The sample sample has the following dimuon channel production distribution:

Hadron side production	Number of events
Λ_c^+	8756
Σ_c^+ or Σ_c^{++}	4604
D^+	1025
η	901
D^0	487
ρ^0	40
D_s^+	11

Table 2.4: *Channels produced in the dimuon MC sample.*

With the following distribution of neutrino interaction:

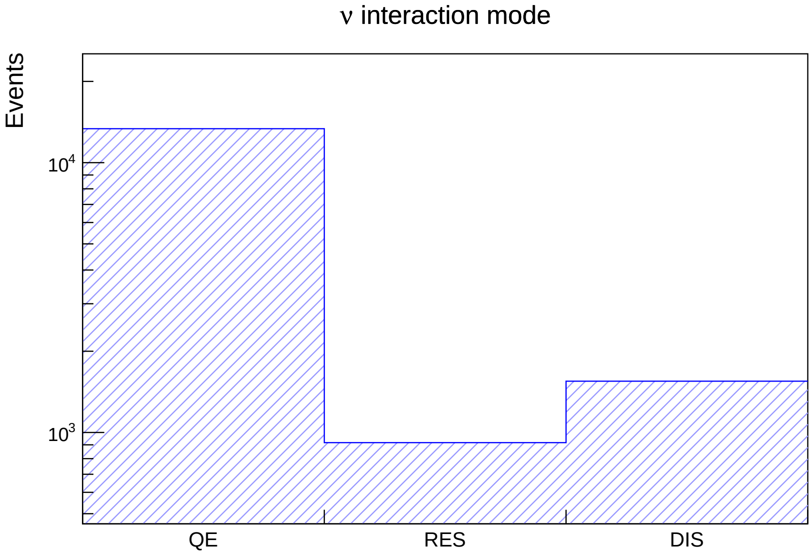


Figure 2.9: ν *interaction modes for dimuon MC sample.*

And the energy distribution of the incoming neutrino is presented in Figure 2.10. It is interesting to see the two peaks around 1.5 GeV and 3.6 GeV. This split shows the two different values required for having dimuon production through η and charmed production, respectively.

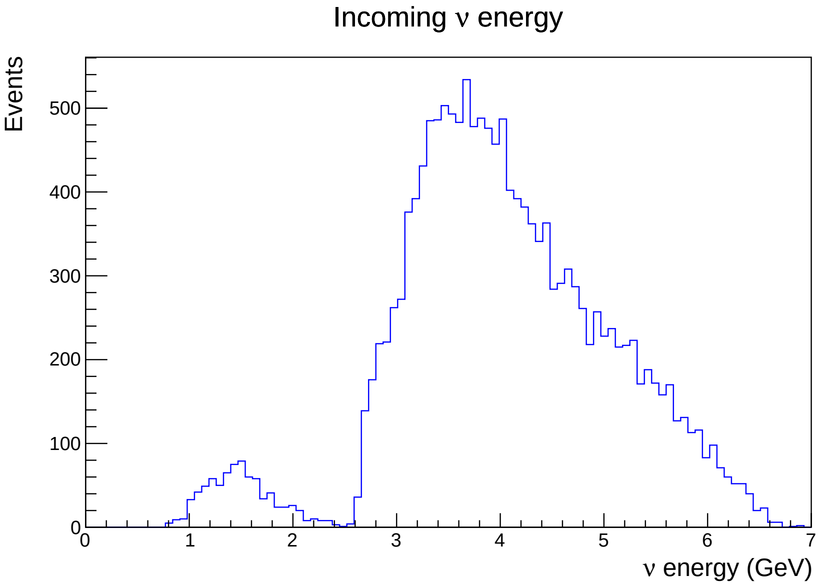


Figure 2.10: *Incoming ν energies for dimuon MC sample.*

As mentioned in Section 1.6.1 the dominant mode is QE (see Figure 2.11), which produces Λ_c^+ particles that leads to dimuon production. The other charmed baryons like Σ_c^+ and Σ_c^{++} are resonances. The next most abundant is the DIS mode (present on D^+, D^0 and D_s^+ production). It is also possible to get charm production from DIS interaction. Finally, the less abundant mode is the RES mode in general for η production (see Table 2.4).

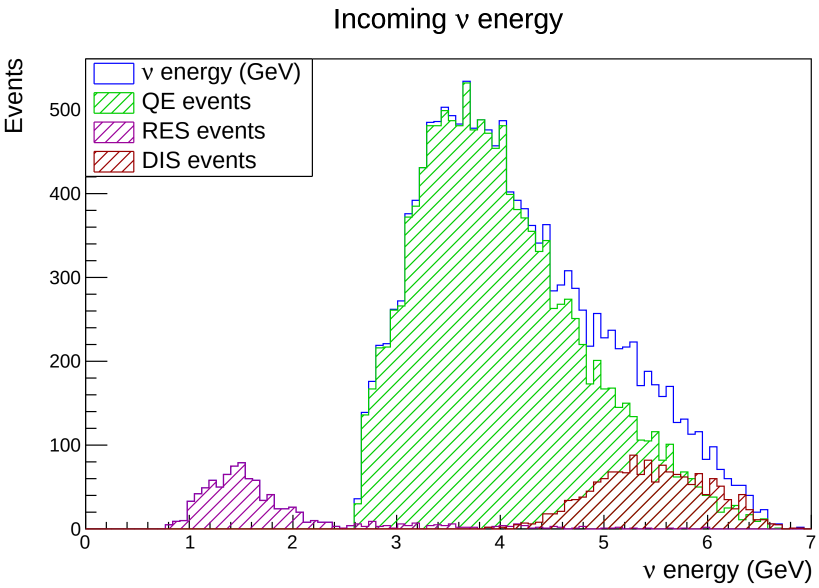


Figure 2.11: *Incoming ν energies for dimuon MC sample for each interaction mode.*

2.4 Open Data, MC, Off-beam and, dirt samples

2.4.1 Open Data Set

The first approach of this study was done using the open data set of μB for run 1. The sample was taken from the updated gLEE group samples repository. It was created with a POT of 4.566×10^{19}. It corresponds to around 5% of the total full data set (including all the runs)). This sample is usually used in μB analysis at the test stage before running with more data to ensure blindness conditions, avoiding bias.

2.4.2 Full MC Central Value sample

The main MC ν interaction simulation in μB for run 1 is essential for this study because, among other reasons, it allows the creation of estimations of the dimuon background. Usually abbreviated as fullMC or central value (CV) sample because it is the MC sample chosen to be the main simulation of ν interaction at μB. Its parameters can be varied to create different samples (usually called detector variations) with different parameters that describe the detector from a different perspective (see Section 4.2). This is useful for the study of systematic uncertainties related to the detector design.

This sample had to be produced to include weights for the additional uncertainty studies for other sources of error different than the detector itself, like the beam flux and the GENIE interaction cross-section models (see Section 4.5).

2.4.3 Off beam or cosmic rays sample

An updated gLEE sample was used for the off-beam or external sample. This sample has to be added to complete the fullMC CV sample (besides the dirt sample) in order to be able to compare the MC with the real data. Section 3.3.7 will explain how to treat the MC events and how to normalize its counting with respect to real data events.

2.4.4 Dirt sample

The dirt sample simulates the set of events that occur outside the detector fiducial volume product of neutrino interactions happening in the surroundings in the cryostat. The products of these events, in a very small fraction, can be collected and reconstructed inside the fiducial volume. For this study, the number of events of this kind that can be reconstructed is not negligible (see Table 3.3) and should be added to the fullMC sample in order to compare with data sets. The reason for the name is historical and from previous experiments.

Chapter 3

Monte Carlo and boosted decision tree analysis

3.1 MC truth analysis

3.1.1 Distribution of the longest MC truth tracks for fullMC

Section 2.1.3 presented the main sources of background for this study. This Section uses the available truth information from the fullMC (MC-truth) for the correspondent run 1 of μB to double-check those backgrounds. All the information from the simulation is available, listing of the particles, energies, and, in particular, the PDG particle ID code. This code is very useful because allows identifying the primary particles involved in each event. Just for this section the MC-truth variables were used. This variables are obtained during the GENIE stage in the MC simulation.

The approach is to identify the two longest tracks in each event and produce a scatterplot that presents the main combinations (using only the most abundant particles) and the proportions. To get the length of the truth track the initial kinetic energy was used, along with a parametrization of the range R in the detector for each particle type as a function of the particle kinetic energy K. Equation 3.1 presents how to make this calculation [25].

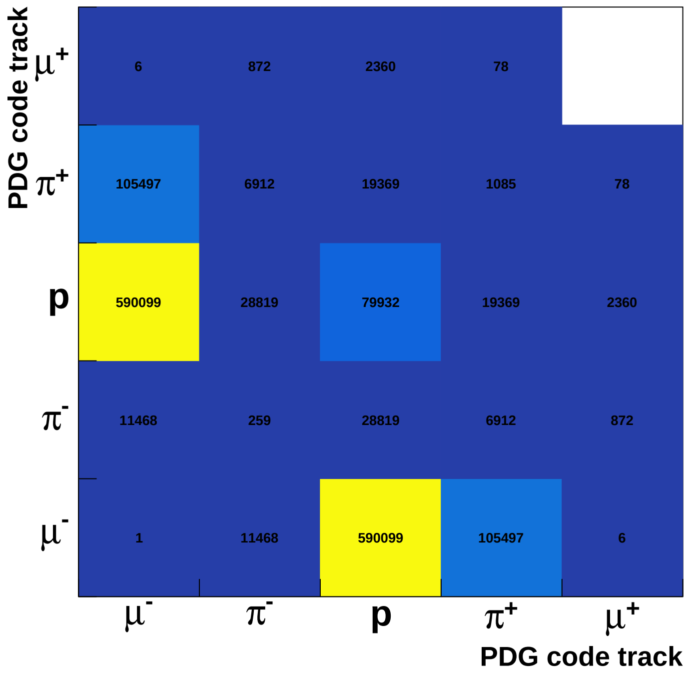
$$R = \left(\frac{b+1}{A} \times K \right)^{\frac{1}{b+1}}, \tag{3.1}$$

where R is the length, K is the kinetic energy, A and b parameters (see Table 3.1).

Particle	A(MeV/cm^{1-b})	b
Pion	8	-0.37
Muon	8	-0.37
Kaon	14	-0.41
Proton	17	-0.42
Deuteron	25	-0.43

Table 3.1: *A and b parameters values for equation 3.1.*

Figure 3.1 shows the scatter-plot of the distribution of events for the fullMC run 1. The plot considers just muons, pions, and protons (the more abundant particles in μB) for the two longest tracks. The numbers in the Figure show how many events have the corresponding pair of particles. The matrix is symmetric because it shows the possible combinations.



	μ^-	π^-	p	π^+	μ^+
μ^+	6	872	2360	78	
π^+	105497	6912	19369	1085	78
p	590099	28819	79932	19369	2360
π^-	11468	259	28819	6912	872
μ^-	1	11468	590099	105497	6

Figure 3.1: *Distribution of the two longest tracks, fullMC sample, without any cuts.*

A large amount of muon-proton pairs are present from QE scattering. In contrast to that, a tiny number of dimuon events are observed in the scatter-plot (see Figure 3.1).
The lengths of the tracks are some of the main variables, as will be discussed in section 3.3.3.

As a test in this part of the analysis different cuts were explored. For example cutting in both lengths, allowing just events with the tracks longer than 40 cm (see Figure 3.2) produces a drastic reduction in the number of protons present.

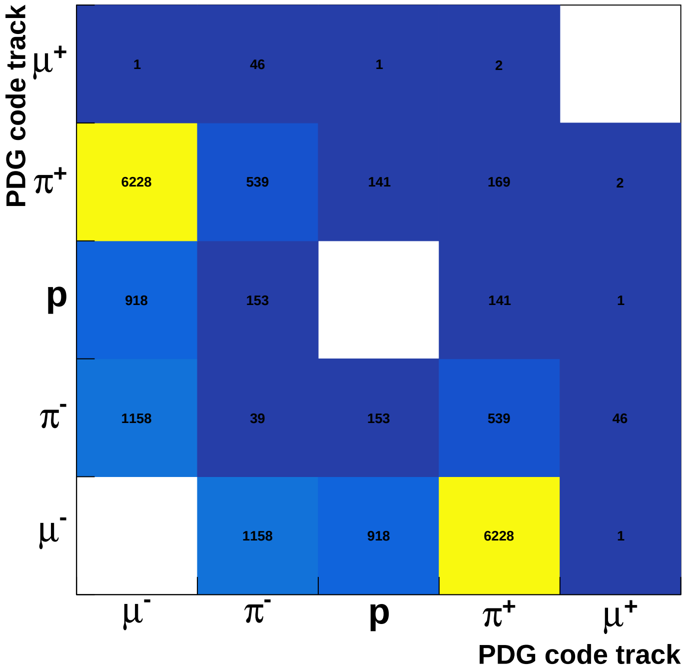
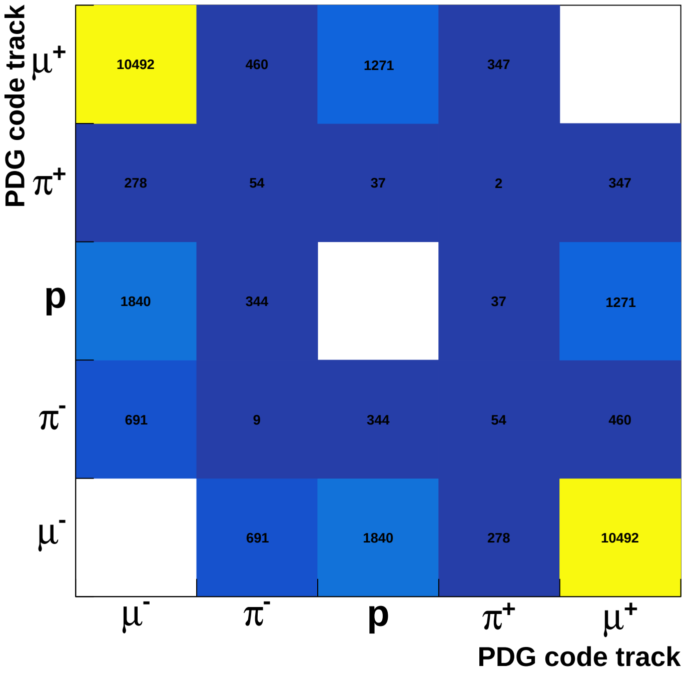


Figure 3.2: *Distribution of the two longest tracks, fullMC sample, cut > 40 cm.*

These cuts reduce the amount of protons to more than 99.8% and the amount of pions by 94%, but the muon-proton and muon-pion pairs still dwarf the SM dimuon signal. This means that just using this variable would not be enough, and more variables are going to be necessary to eliminate the remaining protons and pions. The main difficulty is that, in general, the tracks produced by muons and pions are very similar. If the pion does not interact and comes to rest, it produces a track that is indistinguishable from a muon track. This configures an irreducible background for this study.

3.1.2 Distribution of the longest MC truth tracks for dimuon sample

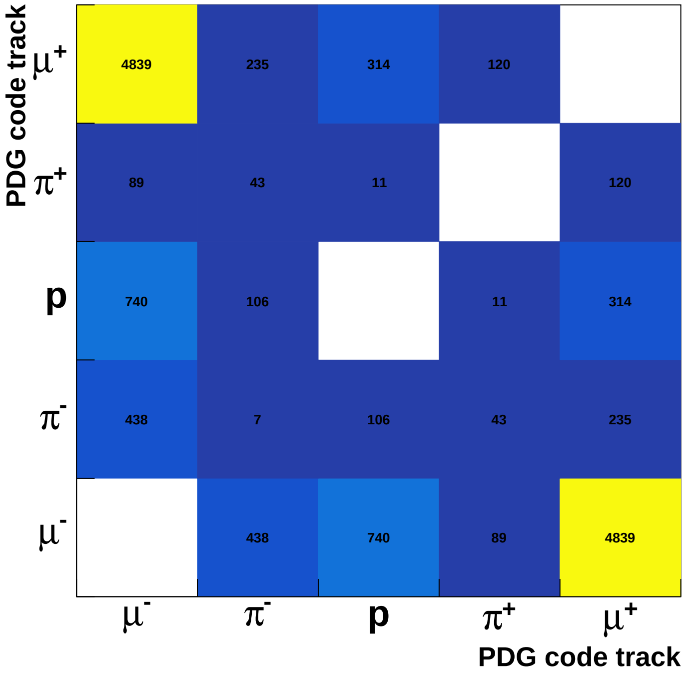
The particle distribution for the two longest MC truth tracks in the dimuon sample is presented as a comparison. As expected, the two longest tracks correspond to the muons in most of the events.



	μ^-	π^-	p	π^+	μ^+
μ^+	10492	460	1271	347	
π^+	278	54	37	2	347
p	1840	344		37	1271
π^-	691	9	344	54	460
μ^-		691	1840	278	10492

Figure 3.3: *Distribution of the two longest tracks, dimuon sample, without any cuts.*

Around 70% of the events are expected to have the muons as the two longest tracks (see Figure 3.3). Similarly to Section 3.1.1, a cut is applied to the length of the tracks, and most of the background events are filtered, which helps to conclude that these variables are key for this study.



	μ^-	π^-	p	π^+	μ^+
μ^+	4839	235	314	120	
π^+	89	43	11		120
p	740	106		11	314
π^-	438	7	106	43	235
μ^-		438	740	89	4839

Figure 3.4: *Distribution of the two longest tracks, dimuon sample, cut* > 40 *cm.*

In contrast to the fullMC sample case, many of the dimuons events are still present after the cut, and most of the protons and pions are filtered out. That concludes once again that the length of the particles will be one of the main variables for the training of a BDT capable of classifying the dimuon events (see Section 3.3). The cuts applied in this Section were not implemented during the pre-selection and selection of dimuon candidates in the reconstruction analysis.

3.2 Analysis strategy

3.2.1 Working with the reconstruction

From now on, everything used in the analysis will come from the Pandora reconstruction. The plan is to use a set of reconstructed variables from events that satisfy some conditions (see Section 3.2.2) to perform a classification (see Section 3.3). To accomplish that classification, it is key to have an enhanced signal dimuon MC sample. Section 2.3 explained the main characteristics of this sample. The classifier should be able to learn those characteristics and identify events in different test samples that share similar characteristics to dimuons. Additionally, a sample that models the background in the data samples is needed (BNB beam on MC simulation or BNB other), a sample with the cosmic rays (off-beam), a sample with the contribution from events happening outside the physical boundaries of the detector and that might be reconstructed inside the fiducial volume (dirt) and real data (open data set). That sample from now on is considered to be the fullMC sample.

The original number of events in the samples used for this study is shown in Table 2.2. The volume of the detector that is going to be used for this study is called fiducial volume. The fiducial volume was selected following the reference [26] with some variations (see Section 3.2.2). This is needed because some MC samples contain events in the very edge of the detector, even sometimes over passing the walls. Also, events that are close to the border experience different electromagnetic effects due to non-uniformities in the physical limit of the wall. Events like this are not interesting for this study. Any event used for this study must have an interaction vertex and the two longest tracks fully contained in the fiducial volume. Events should have the expected topology for dimuons (Figure 3.5), with the vertex and the start of the two longest tracks being close to each other (see Section 3.2.2).

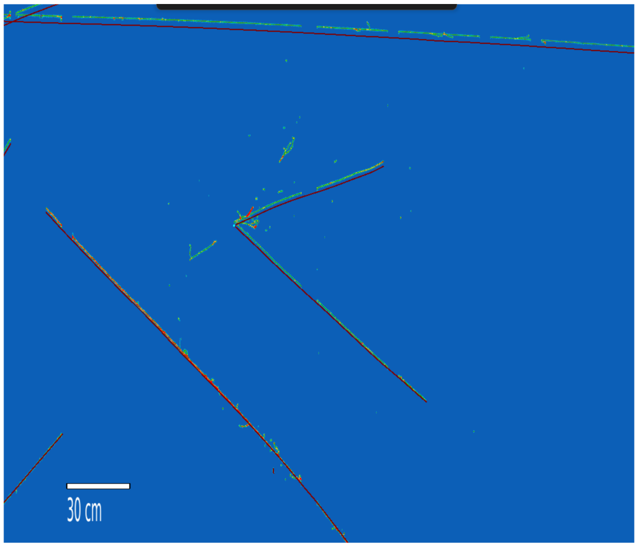


Figure 3.5: *Event display for event 13054 run: 7008.261 in dimuon sample.*

3.2.2 Fiducial Cuts and Pre-selection

The following items summarize the conditions previously discussed:

- Two tracks at least.
- Vertex inside fiducial volume, defined as :

$$21.50 \text{ cm} \leq x \leq 234.85 \text{ cm},$$

$$-95.00 \text{ cm} \leq y \leq 95.00 \text{ cm},$$

$$21.50 \text{ cm} \leq z \leq 966.80 \text{ cm}.$$

Figure 3.6 shows these values in a dashed line compared to the solid line corresponding to the real edges of the detector.

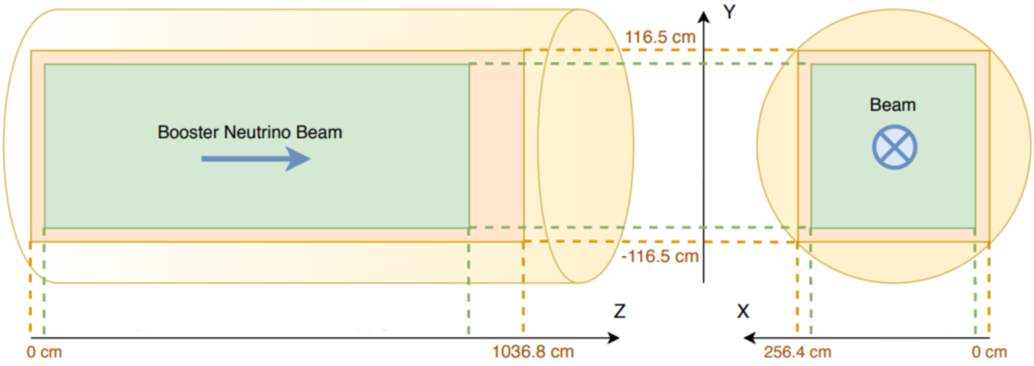


Figure 3.6: *Fiducial volume, picture edited from [27].*

- Start and end of any track should be fully contained in the fiducial volume.
- Start of track at least 4 cm within the vertex.
- Angle between the tracks < 160 deg.

After requiring this conditions almost 90% of the event in the FullMC sample are removed when applying the two tracks requirement, from those events almost 50% were removed by fiducial conditions and additionally 10% were removed with the vertex and angle condition. Table 3.2 shows in detail the number of events removed by each condition for all the samples.

Condition	**Dimuon**	**FullMC**	**Open data**	**Off-beam**	**Dirt**
Initial number of events	15826	1026120	176429	820416	104572
Two tracks minimum	3758	768873	158209	779560	98642
Vertex fully contained	3791	91912	7854	21497	3847
Tracks fully contained	3841	91772	6656	15262	1585
Tracks starts near vertex	2414	18359	913	1285	152
Angle of tracks < 160 deg	35	2247	101	16	7

Table 3.2: *Number of events filtered by the selection conditions.*

Table 3.3 presents the resulting samples. These samples are going to be organized and used in Section 3.3.2 to train a BDT able to classify dimuons.

Sample	**Number of events**
Dimuon sample	1987
FullMC sample	52957
Open data set run1 sample	2696
offbeam sample	2796
Dirt MC run 1 sample	339

Table 3.3: *Number of events in samples after applying selection criteria.*

3.3 BDT analysis

3.3.1 Parameters

A boosted decision tree (BDT) is a classifier with a descending tree-shaped structure (see Figure 3.7) that splits a test sample into background or signal depending a repeated binary decisions that are taken on one single variable at a time until a stop criterion is fulfilled. In general, these criteria are related to the purity of the amount of events that are in each node (see 3.2),

$$p = \frac{S}{S + B}, \tag{3.2}$$

where S means signal and B means background. When the value of p is close to 0 the fraction of events in the node are tagged as pure background. On the other hand, when p is close to 1, the events are tagged as pure signal. If this value is not close enough to this criteria, a subsequent condition could be applied with other variables, and the process is repeated.

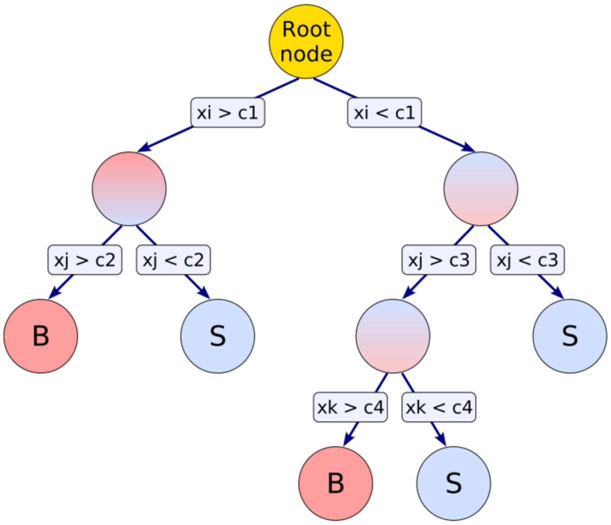


Figure 3.7: *Diagram showing the structure of a BDT with nodes of decision and their corresponding variables used in each step, taken from [28].*

The phase space defined by the events in the original sample is then split into many regions that are eventually classified as signal or background. Depending on the number of training events that end up in the final leaf (signal or background), the leaf is considered to be signal-like or background-like [28].

To do that, the BDT should be trained with samples that were previously known as background and signal. In this study, the samples used for this training are specified in Table 3.5. The multi-variable toolkit (TMVA) of root was used for the BDT analysis, more precisely the TMVAclassification.C example (see section 2.6 of [28]) was used as template, modified and adapted for the needs of this study. The parameters used in the tuning of the BDT are given in Table 3.4.

Parameter	Value or Option
BoostType	AdaBoost
Boost Value	0.25
NTrees	700
MinNodeSize	2.5 %
nCuts	20
MaxDepth	3

Table 3.4: *Parameters of configuration for the BDT.*

AdaBoost is a boosting algorithm that trains classifiers. It assigns weights to misclassified events and their corresponding tree during the training stage to make its contribution smaller to the final classifier [28].

The value chosen for that boost is 0.25, modifying the value by default of 1.0. This value controls the boost performed by AdaBoost. Reducing this value can improve the performance of the classifier by assigning smaller weights (compared to the default value) to the initially not precise trees. In each generation of trees, an initial assumption is made in pursuit of a value that minimizes a particular function (could be, for example, the purity, p

in Equation 3.2). Modifying the boost value makes the steps necessary in this search to be smaller, making that calculation to be more smooth with the cost of having longer computing times.

The number of trees (NTrees) was set to a value of 700, which balances the computational efficiency with the performance of the classifier. Several tests were done, showing that increasing this number did not make a large difference and made the classification slower. The default value is 800.

For the minimum node size (MinNodeSize), the value of 2.5% was chosen. The default value in TMVA classification library is 5%. It was modified because the dimuon signal training sample has less than 2000 events. In the cases where one of the training samples is small, it is a good idea to reduce this value [28]. It means that when a node is filled with that number of events (from the total number in the sample), more splitting is needed. A new node is generated until the number of events present is less than 2.5%, in which case the node becomes a terminal node or a leaf in the BDT.

The number of cuts (nCuts) was kept by default to 20. This means that for each variable, 20 possible thresholds would be explored and one would be chosen for splitting each node. Making a large increment to this value could lead to overfitting. The maximum depth (MaxDepth) controls the complexity of the BDT by defining the levels (number of descending nodes) possible in each tree. For this study, it was kept by default to 3.

3.3.2 Samples used for training and testing the BDT

After applying the pre-selection (see Section 3.2) the resulting samples were organized for the training and testing of a BDT as follows:

BDT Step	Sample	Number of events
Signal training	Dimuon sample	1987
Background training	1/2 full MC + 1/2 off-beam + 1/2 dirt sample	28045
Signal testing	Open data set run1 sample	2696
Background testing	1/2 full MC + 1/2 off-beam + 1/2 dirt sample	28045

Table 3.5: *Samples used for training and testing the BDT.*

As shown on Table 3.5 the dimuon sample was used for the training of the signal given that this sample is conformed only with dimuon events (plus the cosmic overlay) which have all the kinematics and information that is pursued.

On the other hand, half of the full MC neutrino interaction simulation sample was used as background training, given that all the events that conform to this sample are not dimuons (checked event by event) and have the characteristics of events that would be commonly seen in the detector with the contrary phenomenology of dimuons.

3.3.3 Variables chosen for Training and Testing

After applying the fiducial and pre-selection conditions (see Chapter 3.2), seven were the variables chosen for the BDT analysis. Initially, this number was larger, but it was found that having too many variables could require a large amount of events available for the training process.

Plots are normalized to 1 (shape only is shown). The histograms show the distribution for dimuon signal (training), fullMC plus offbeam (training and testing), and open data set run 1 (testing) samples. These variables show discrimination that can help the training of the BDT. The BDT should be able to learn special characteristics that distinguish dimuon-like signals from background events.

As seen in Section 3.1 the length of the tracks is vital to differentiate protons from muons, Figures 3.8 and 3.9 show the behavior of these variables. As expected, the length of the second track tends to be longer in the dimuon signal sample, which is expected given that this sample is a MC that is required to have at least two muons in the final state, and muons leave a long track in the detector.

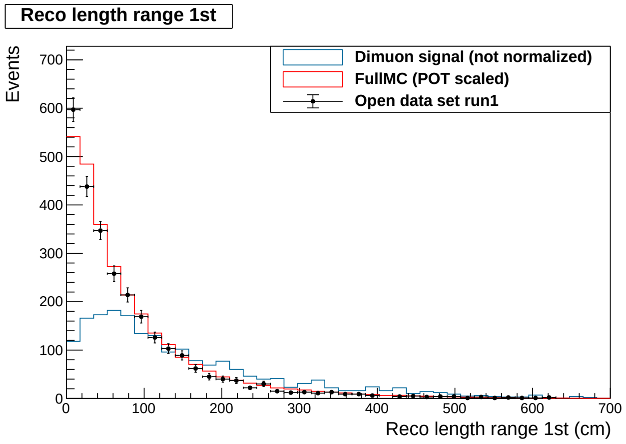


Figure 3.8: *Length of the longest track.*

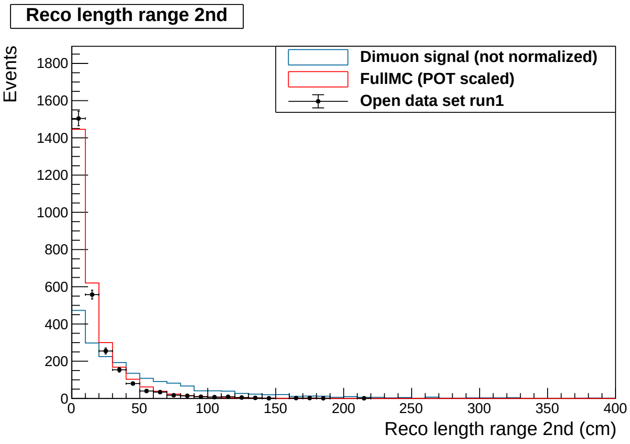


Figure 3.9: *Length of the second longest track.*

The calorimetry energy is the summation of all energy of all hits on each plane. This variable specifically corresponds to the maximum value of energy deposited in any plane. For muons, the calorimeter energy will correlate almost totally with range because muons experience a moderate energy loss while through the detector. This will be less so for pions because they experience strong interaction, therefore helping to distinguish them.

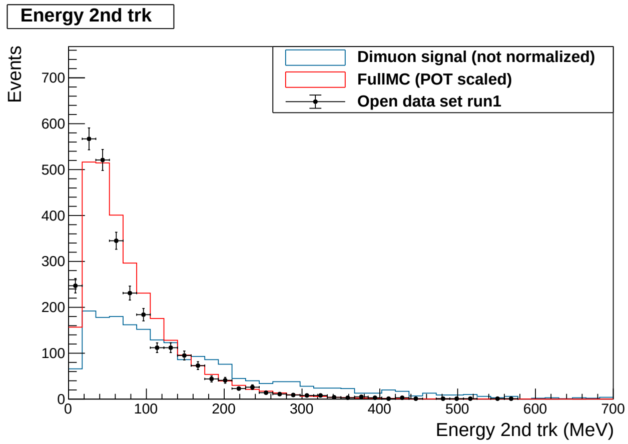


Figure 3.10: *Calorimetry energy of the longest track.*

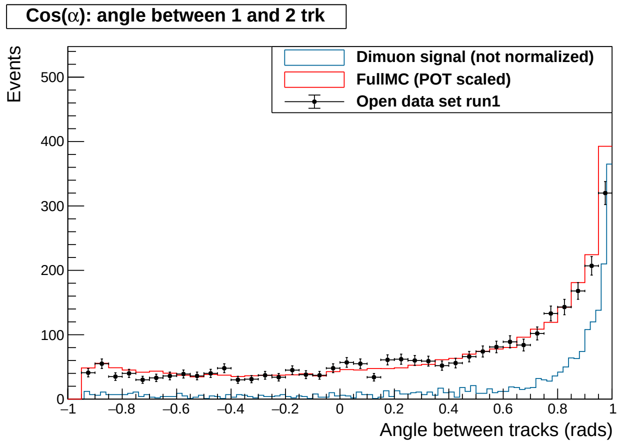


Figure 3.11: *Calorimetry energy of the second longest track.*

The angle between the two longest tracks was calculated using the expression:

$$\cos(\alpha) = \cos(\theta_{yz,1})\cos(\theta_{yz,2}) + \sin(\theta_{yz,1})\sin(\theta_{yz,2})\cos(\phi_{yx,1} - \phi_{yx,2}), \tag{3.3}$$

where the index 1 corresponds to the longest track and 2 to the second longest track. A cut was performed to require $cos(\alpha)$ > -0.94 or α < 160 degrees, which ensures that back-to-back events (highly related to misidentified "broken" tracks) are filtered.

The number of daughters that the tracks have was also chosen for this study. It is more likely that the second muon is a μ^+, which always (see Section 1.5.1) will produce Michel electrons. A π^+ can also produce Michel in the case that it decays, but often it does not.

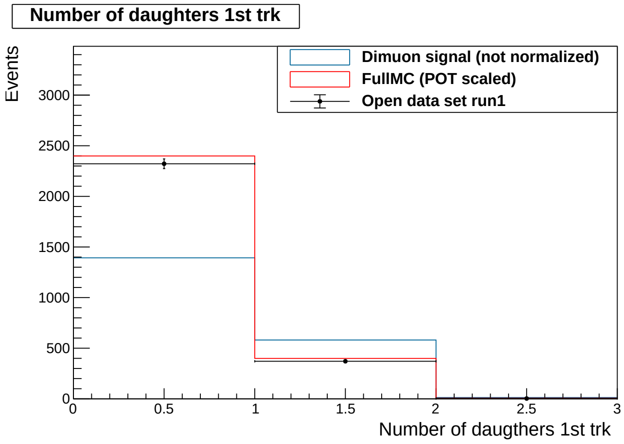


Figure 3.12: *Number of daughters of the longest track.*

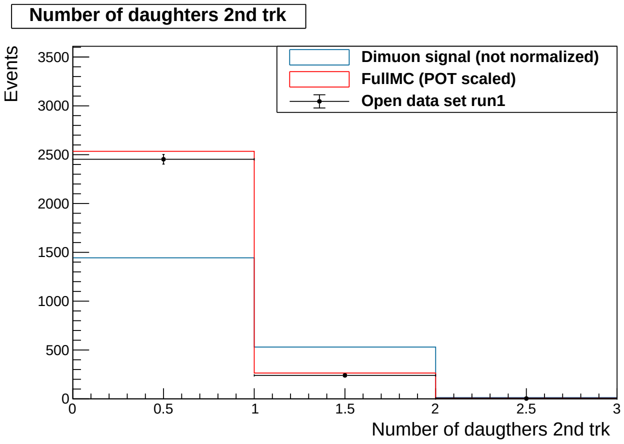


Figure 3.13: *Number of daughters of the second longest track.*

Given that muons present a pronounced Bragg peak, it is expected that the fraction of dE/dx in the last third of the track over the two first thirds is going to be larger than for the case of protons, pions, or other particles. The reconstruction assigns one plane as the "best plane" in each event, which means that was the plane that better reconstructed the value of dE/dx. That value was chosen in all the calculations.

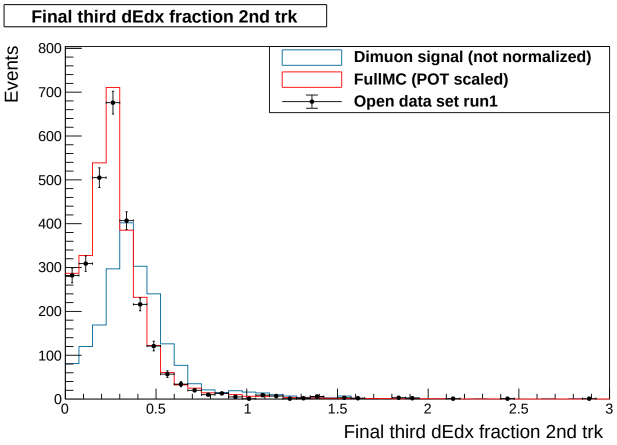


Figure 3.14: *Fraction of dE/dX of the second longest track.*

3.3.4 BDT training

With these 15 variables and the parameters shown in section 3.3.1, the training of the BDT was performed using the whole dimuon signal sample (see Table 2.2) as training signal and half of the fullMC sample as training background. It achieved good discrimination between the dimuon signal and ν interaction MC for μB events.

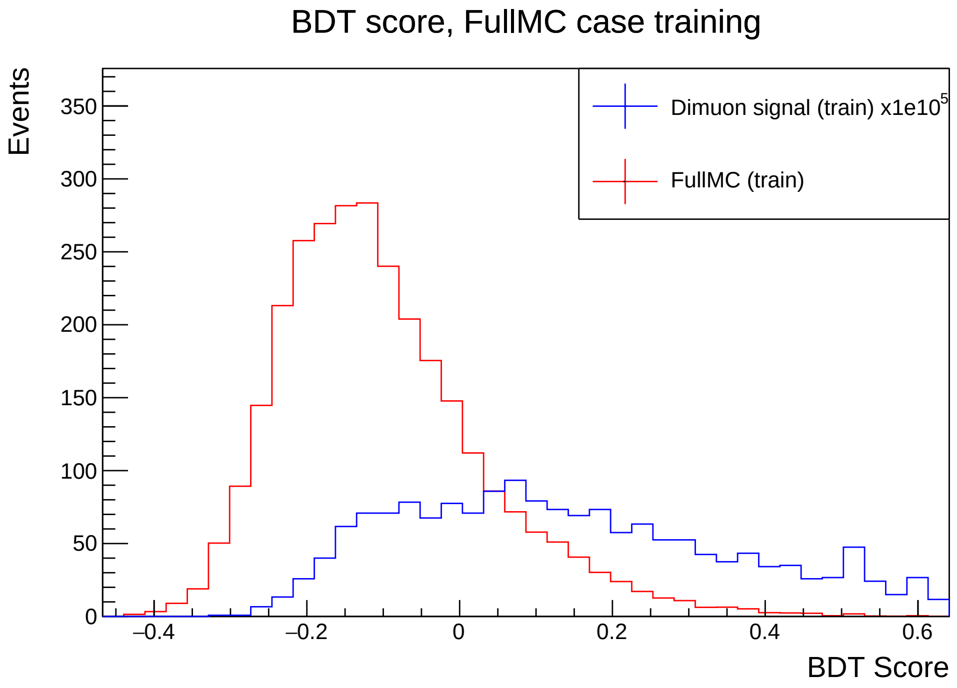


Figure 3.15: *BDT trained using the dimuon and fullMC samples.*

There is a normalization applied in Figure 3.15 to make both distributions look the same size and be able to compare the shapes. Without this, the dimuon distribution (~2K events) would be too small to be able to see in the same frame with fullMC (~28K events for training) distribution. Section 3.3.6 will explain the normalization factor needed to compare dimuon events with fullMC events.

3.3.5 BDT testing

Given that Figure 3.15 shows good discrimination, the next step is to test the classification of the BDT. To do that, the open data set run 1 sample (see 2.2) was used as testing signal and the remaining half of the fullMC sample as testing background. Figure 3.16 presents the test and training overlapped. The test distributions are shown in solid point markers with their corresponding statistical error. The solid histograms correspond to the training (see Figure 3.15).

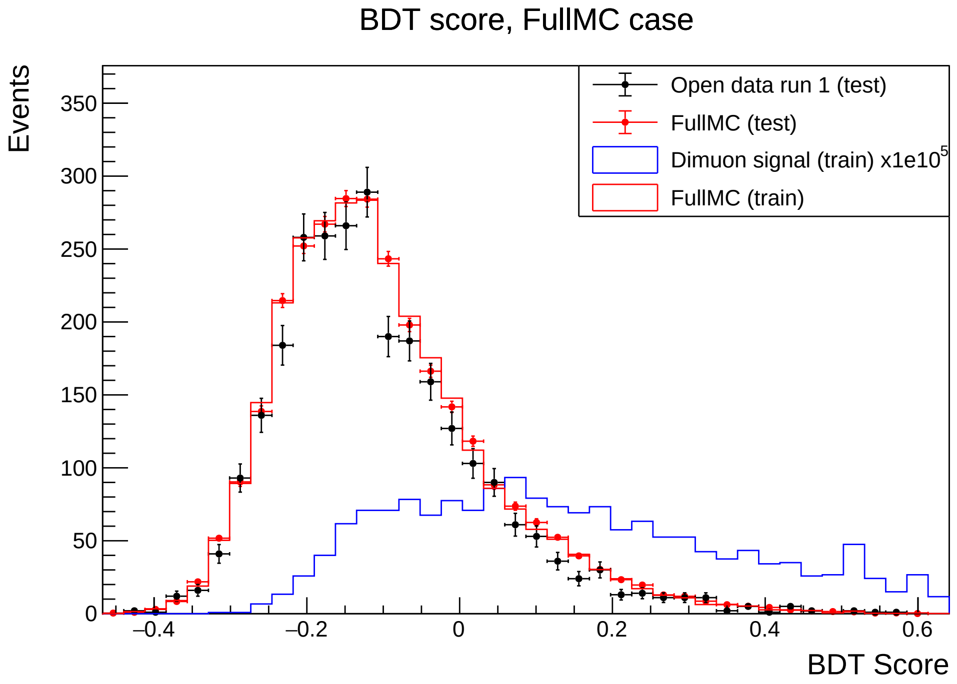


Figure 3.16: *BDT tested with the open data set run1 used as testing signal and half of fullMC as testing background.*

Similarly to Figure 3.15, Figure 3.16 has a normalization to make all distributions look the same size and be able to compare the shapes. Without this, the data distribution (~2.6K events) would be too small to be able to see in the same frame with fullMC (~28K events for training and testing) distributions. Section 3.3.6 will explain the normalization factor needed to compare data events with fullMC events.

Dimuon signal is very rare, that is why in Figure 3.16 there is not discrimination between the BDT score distribution for the open data set and the fullMC. It is possible to make an approximated calculation of the number of dimuons expected in the open data set :

$$10 \text{ dimuons present in fullMC} \times \frac{4.566 \times 10^{19}}{1.28344 \times 10^{21}} = 0.35 \approx 1, \tag{3.4}$$

where the fraction is the normalization of the data sample with respect to the fullMC. The values are the POT for the data set in the numerator and the POT for the fullMC in the denominator. Section 4.2 will explain in more detail the POT values for the different samples.

Then just one event would be expected in the open data set run 1. That explains why the testing BDT distributions are almost overlapping. This also indicates that the fullMC sample is self-consistent because it can effectively describe the behavior of the real data events sample.

3.3.6 Signal efficiency

To determine how efficiently the analysis has been in classifying the dimuon signal, it is necessary to perform a test with a different dimuon testing sample. This sample has 25% the number of dimuon events used for the training. If the BDT is working correctly, there should not be big discrepancies between the testing and training samples. The testing fullMC is the same that in Section 3.3.5.

A small adjustment is applied to the efficiency calculation. There are 78 events in the dimuon sample that were generated (during the GENIE step) outside the fiducial volume of the detector. Those events are removed in the efficiency calculation (see Equation 3.5). That produces a 0.5% difference in the efficiency.

$$\text{Total number of events dimuon sample} = 15826 - 78 = 15748. \tag{3.5}$$

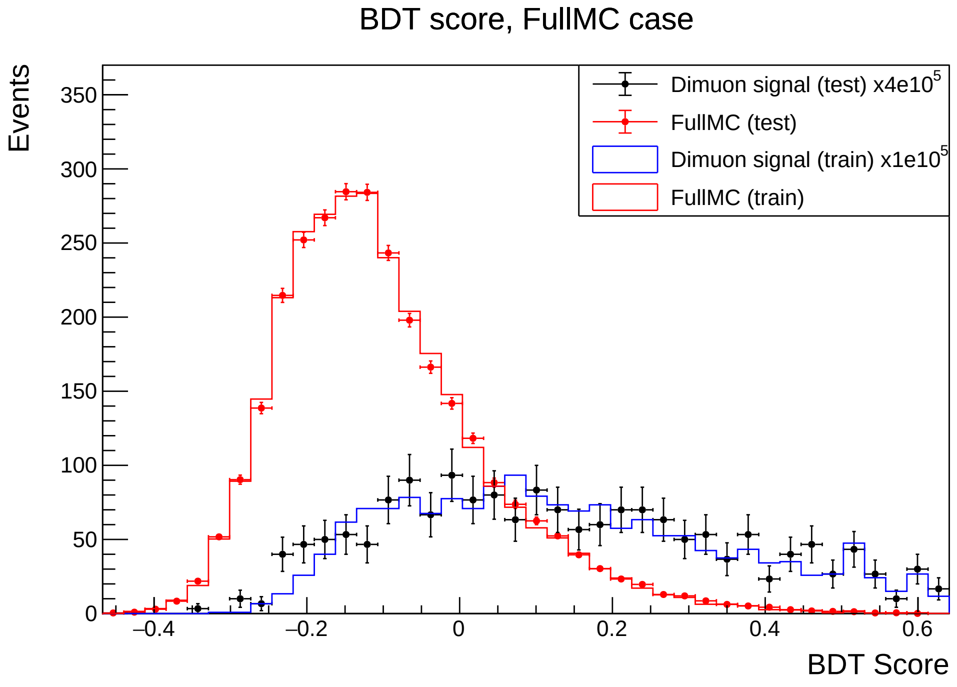


Figure 3.17: *BDT tested using a different dimuon sample.*

Chapter 6 will show a quantitative test to prove that the testing and training distributions show good agreement. This is enough to prove that the efficiency obtained from both distributions is going to be closely related. The plot does not start at 1 because it takes into account the pre-selection efficiency. The efficiency value is going to be important for the signal estimations (see Figure 3.17).

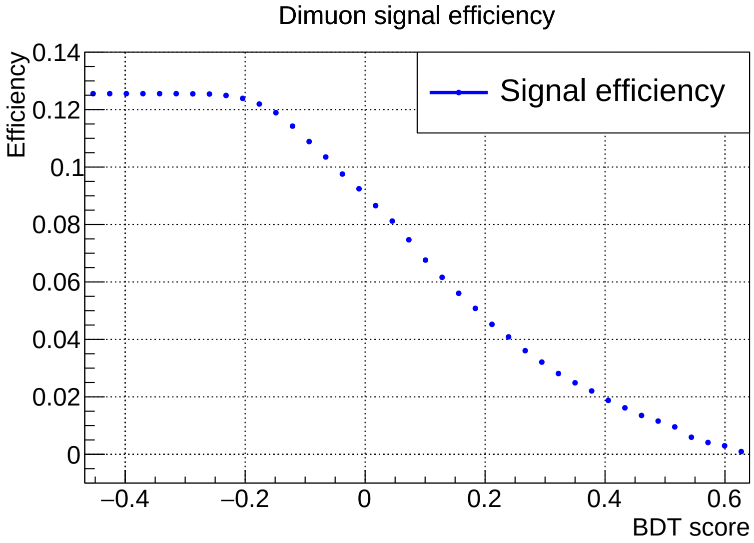


Figure 3.18: *Dimuon signal efficiency.*

Once the efficiency is obtained, it is usual to calculate the purity for different BDT cuts. For this study, the purity is going to be practically zero, due to the background is very large compared to the dimuon signal. This definition is different from the BDT purity shown in Equation 3.2, which is an internal variable used by the classifier as part of its performance analysis. The signal purity would be:

$$\text{Purity} = \frac{\text{Events passed cut dimuon sample}}{\text{Events passed cut dimuon sample} + \text{Events passed cut fullMC sample}}, \tag{3.6}$$

where the number of events in the dimuon sample must be normalized using the POT correspondent to the testing fullMC and dimuon samples (samples used during the training). The normalization factor is:

$$\mathrm{N}_{\text{dimuons}} = \frac{\text{POT 1/2 fullMC sample}}{\text{POT dimuon sample}} = \frac{0.5 \times 1.28344 \times 10^{21}}{5.480298 \times 10^{24}} = 0.00012. \tag{3.7}$$

Equation 3.7 shows why comparing the dimuon events with the fullMC events in the purity equation is going to result in values close to zero. Then, different criteria should be used to determine where to apply BDT cuts that might define the BDT selection region.

3.3.7 Proportions

The test background sample was confirmed by fullMC, off-beam, and dirt events. The normalization factor to compare these background events with open data set events is the ratio (see equation 3.8) between the data and fullMC POTs. For the fullMC POT, a factor of 0.5 was used because just half of the whole fullMC sample was used for testing:

$$\mathrm{N}_{\mathrm{fullMC}} = \frac{\text{POT data sample}}{\text{POT 1/2 fullMC sample}}, \tag{3.8}$$

where the POT for the open data sample is known in μB. The POT value for the fullMC sample was calculated by counting the sub-runs in the MC.

$$\mathrm{N}_{\mathrm{fullMC}} = \frac{\text{POT data sample}}{\text{POT 1/2 fullMC sample}} = \frac{4.566 \times 10^{19}}{0.5 \times 1.28344 \times 10^{21}} = 0.0712. \tag{3.9}$$

For the normalization with respect to the off-beam (cosmic background) sample, the number of spills should be considered [24]. For the off-beam sample, a factor of 0.5 was used because just half of the whole sample was used for testing in combination with the fullMC:

$$\mathrm{N}_{\mathrm{offbeam}} = \frac{\text{Spills data sample}}{\text{1/2 Spills off beam sample}} = \frac{10127594}{0.5 \times 28190365} = 0.7185. \tag{3.10}$$

Additionally, the normalization with respect to the dirt events was calculated.

$$\mathrm{N}_{\mathrm{dirt}} = \frac{\text{POT data sample}}{\text{POT 1/2 dirt sample}} = \frac{4.566 \times 10^{19}}{0.5 \times 3.23362 \times 10^{20}} = 0.2824. \tag{3.11}$$

Keeping the proportions, the combined normalization is:

$$\mathrm{N} = \frac{\text{Events fullMC}}{\text{Totalevents}} \times 0.07058 + \frac{\text{Events offbeam}}{\text{Totalevents}} \times 0.7185 + \frac{\text{Events dirt}}{\text{Totalevents}} \times 0.2824 = 0.1045. \tag{3.12}$$

3.3.8 Background estimation and open data as function of BDT score

To make a dimuon background estimation it is needed to count how many events were selected by the BDT in the testing background sample. This test sample is the addition of fullMC, off-beam, and dirt events. Then, this number is going to be defined as:

$$B_T = B_{MC} + B_{Ext} + B_{Dirt}. \tag{3.13}$$

The estimation of the background is obtained by normalizing the total background:

$$B_e = B_T \times N = (B_{MC} + B_{Ext} + B_{Dirt}) \times N. \tag{3.14}$$

The background estimation and the open data set (see Figure 3.15) could be obtained as a function of the BDT score :

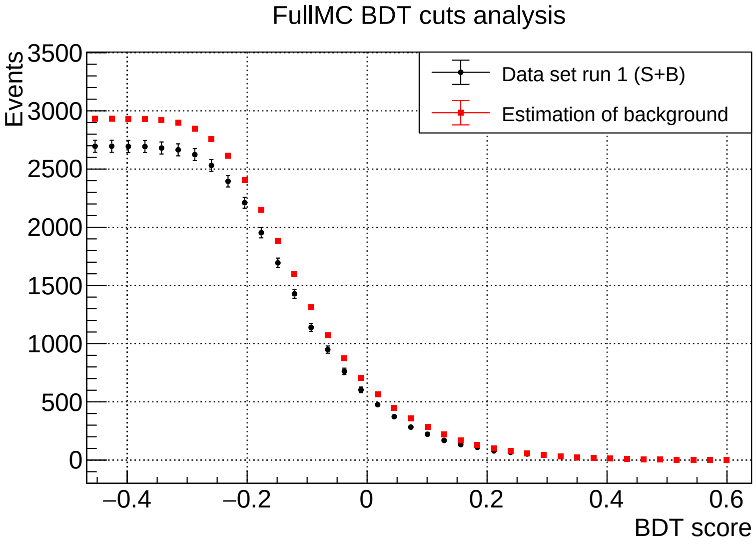


Figure 3.19: *Background estimation and open data set as a function of BDT score.*

These functions show good agreement with each other. There is no considerable excess or deficit of data. A logarithmic scale was applied to the plot to see more clearly the high BDT region.

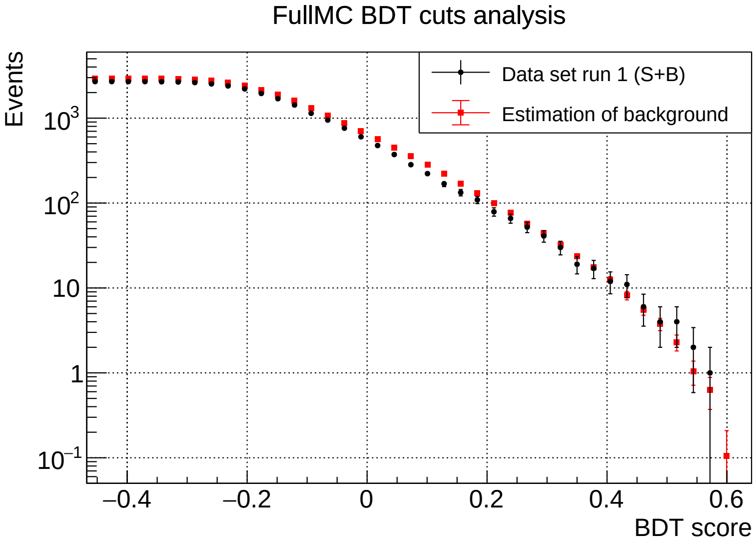


Figure 3.20: *Background estimation and open data set as a function of BDT score.*

The ratio plot shown in Figure 3.21 confirms that the proportions of estimated background and data are kept during most of the BDT score values (the ratio plot being a flat line). The last part shows some small variations because of the low statistics and considerable fluctuations because of that.

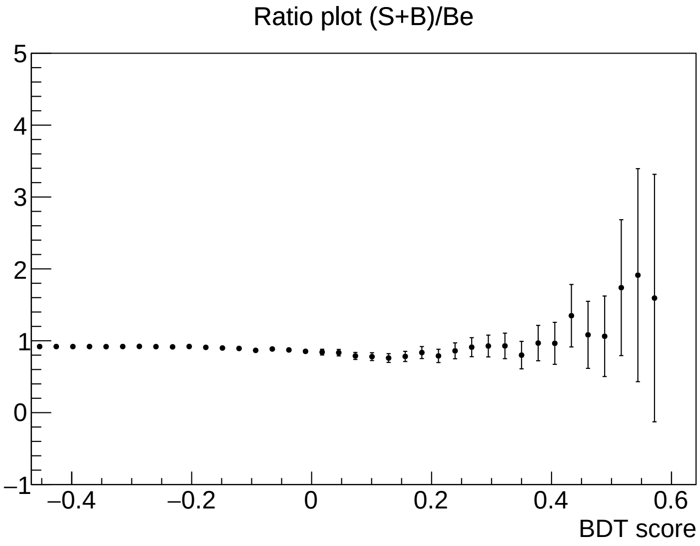


Figure 3.21: *Ratio plot for open data set vs. background estimation as a function of BDT score.*

3.4 Results

3.4.1 BDT cut

Two criteria are going to be used to determine the BDT cut value. The first criterion is going to explore a BDT cut that allows more background to pass and, at the same time, more efficiency to be obtained. This requires the estimation of the background to be fixed to a value as close as possible to 16. This choice is arbitrary, but convenient. A two sigma statistical excess would simply be $2 \times \sqrt{16} = 8$ events.

The second case requires minimizing the estimated background. This is that the BDT cut is such that the estimation of the background is equal to 1. Doing the calculations, the BDT score cut values that ensure these requirements are:

$$BDTcut_1 = 0.37, \tag{3.15}$$

$$BDTcut_2 = 0.53. \tag{3.16}$$

These cuts are indicated in the next plot. $BDTcut_1$ is shown by the first arrow from the left. $BDTcut_2$ is shown by the second arrow from the left.

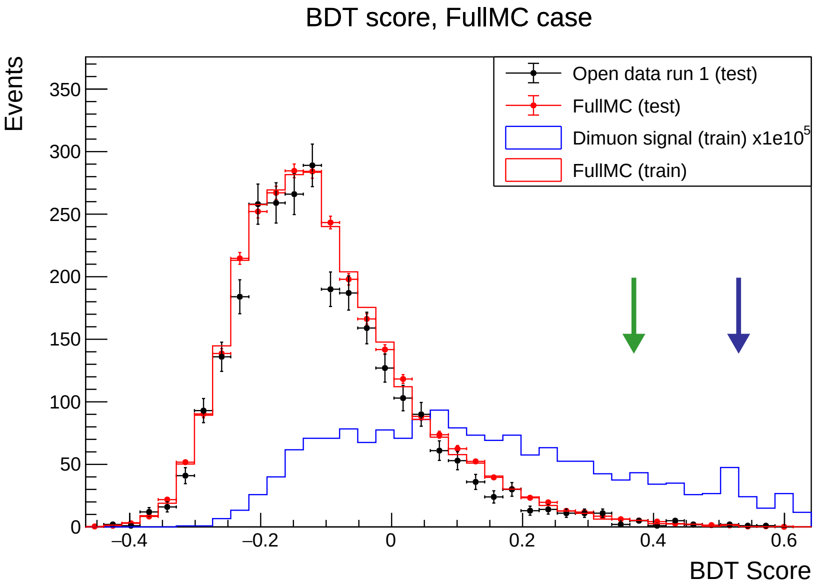


Figure 3.22: *BDT cut scores.*

The behavior of the variables used for the BDT analysis after applying the first cut is:

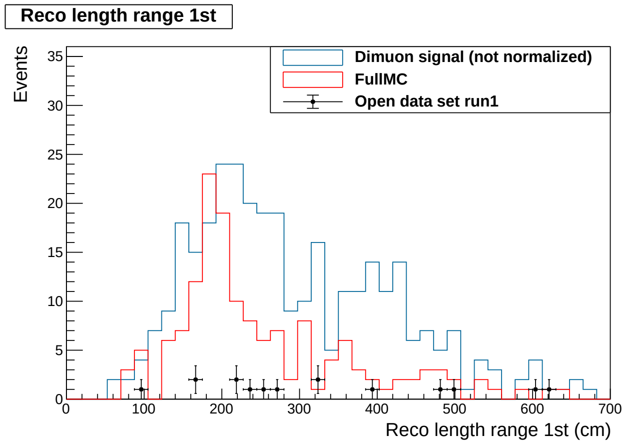


Figure 3.23: *Length of the first track after the BDT cut.*

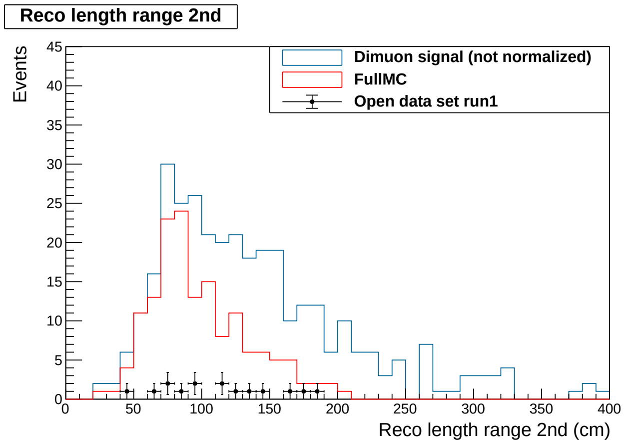


Figure 3.24: *Length of the second track after the BDT cut.*

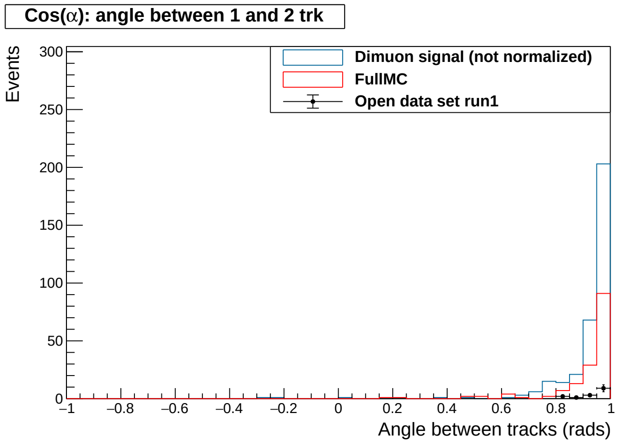


Figure 3.25: *Angle between tracks after the BDT cut.*

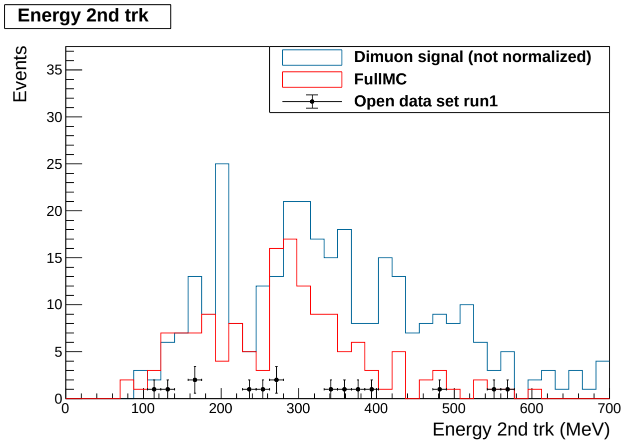


Figure 3.26: *Calorimetry energy of the second track after the BDT cut.*

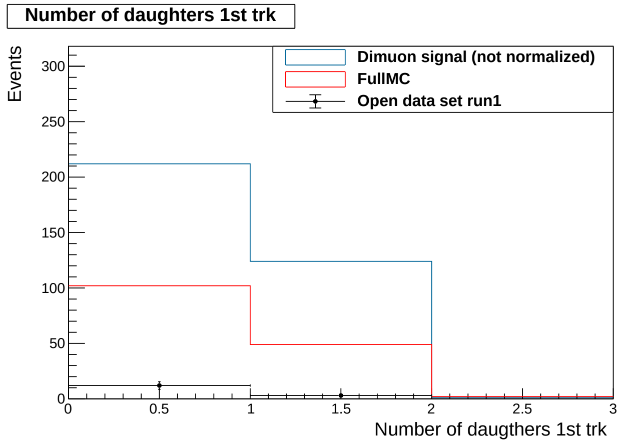


Figure 3.27: *Number of daughters of the first track after the BDT cut.*

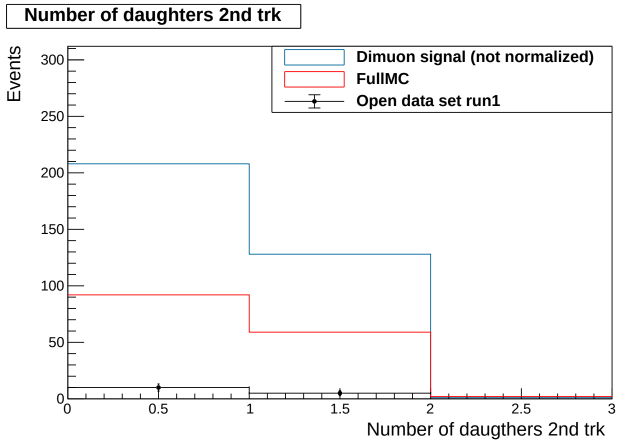


Figure 3.28: *Number of daughters of the second track after the BDT cut.*

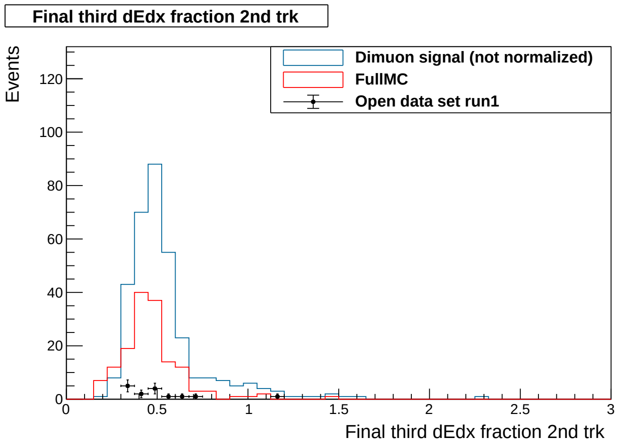


Figure 3.29: *Fraction of dE/dX of the second track after the BDT cut.*

The behavior of the variables used for the BDT analysis after applying the second cut is:

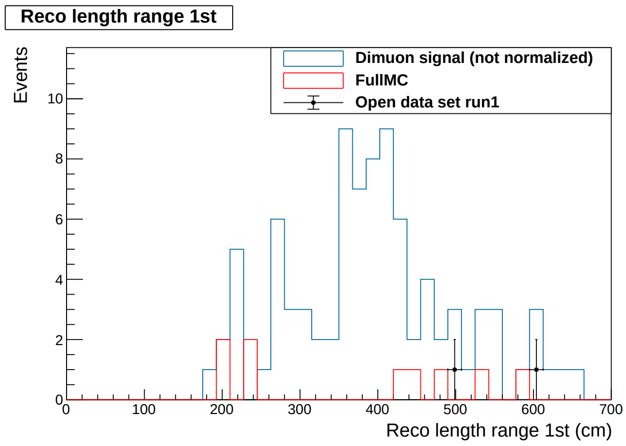


Figure 3.30: *Length of the first track after the BDT cut.*

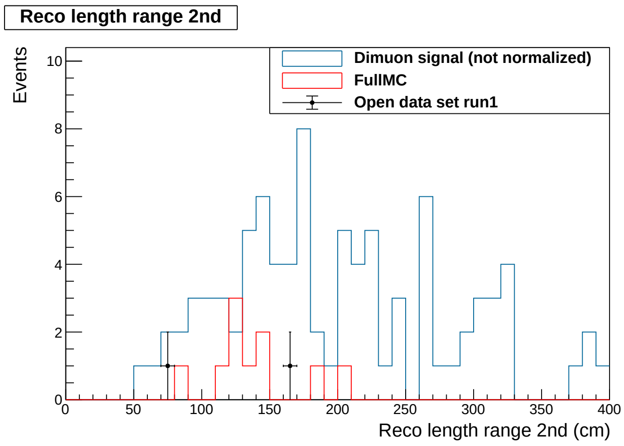


Figure 3.31: *Length of the second track after the BDT cut.*

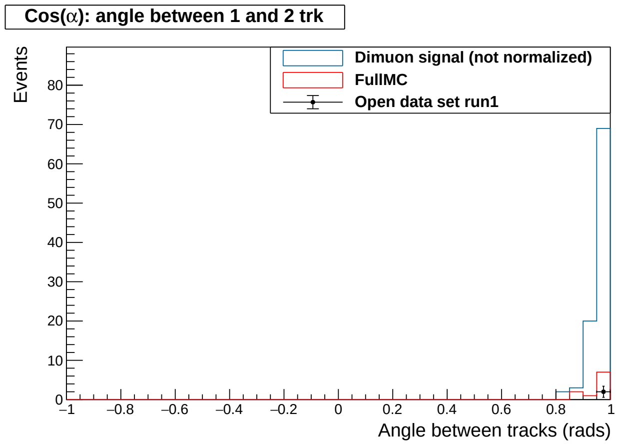


Figure 3.32: *Angle between tracks after the BDT cut.*

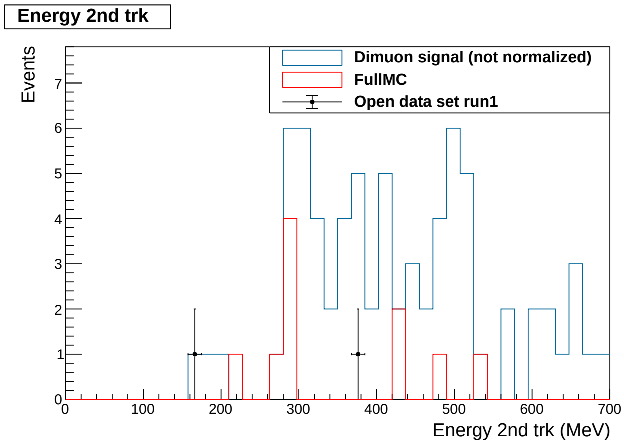


Figure 3.33: *Calorimetry energy of the second track after the BDT cut.*

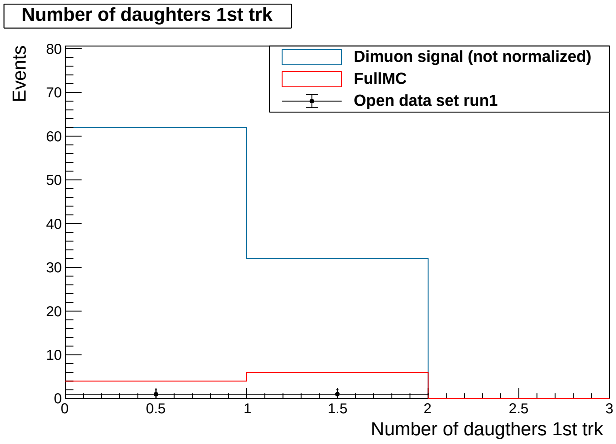


Figure 3.34: *Number of daughters of the first track after the BDT cut.*

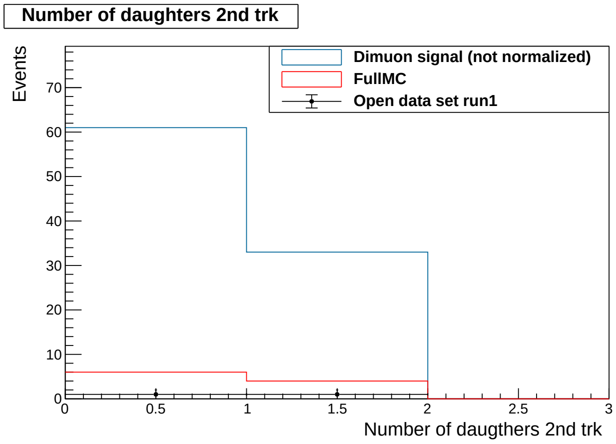


Figure 3.35: *Number of daughters of the second track after the BDT cut.*

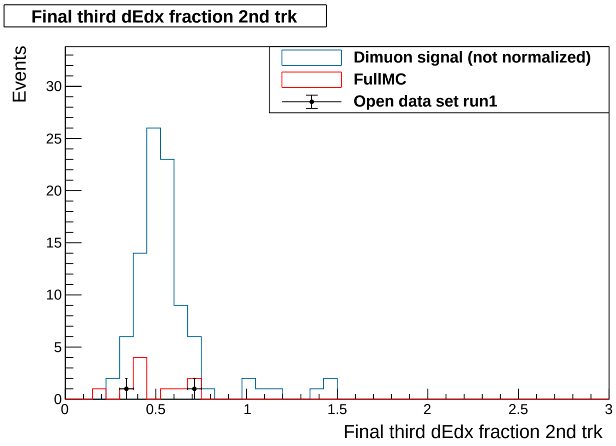
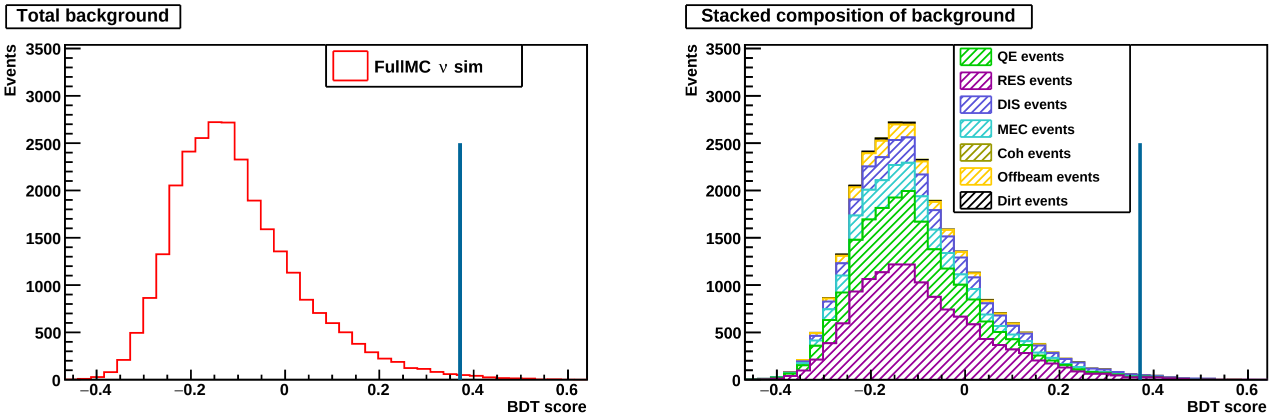


Figure 3.36: *Fraction of dE/dX of the second track after the BDT cut.*

The BDT was successful in learning characteristics from the dimuon-enhanced sample. After applying the BDT analysis, the variables seem to be more similar to the dimuon distribution, which indicates a good performance for the selection. The BDT was able to pick up events with two long tracks (see Figures 3.23 and 3.24), with a small energy loss for most of the track until reaching the end of the track where the loss was more evident (see Figures 3.29 and 3.36 respectively).

3.4.2 Background composition anti-cut region for $BDTcut_1$

The anti-cut region is defined by values smaller than the BDT score cuts (see Section 3.4.1). The background composition in this region for the $BDTcut_1$ is:


Total background

Stacked composition of background

Figure 3.37: *Total BDT testing background distribution (left). Background composition for the different neutrino interaction modes (right), solid line indicates the BDT cut value.*

Table 3.6 shows the number of events for each neutrino interaction mode (see Section 1.6) present in the BDT testing background distribution in the anti-cut region for $BDTcut_1$.

ν **interaction mode**	**Number of events**
RES	13009
QE	7026
DIS	3124
MEC	3095
Off-beam	1397
Dirt	169
Coherent	71

Table 3.6: *Events in the BDT anti-cut region for* $BDTcut_1$ *in the testing background sample.*

As expected in the anti-cut region, the RES and QE modes dominate (see Figure 3.38), given that the BNB mean neutrino energy is ~ 800 MeV.

3.4.3 Background composition BDT selection region for $BDTcut_1$

Figure 3.39 presents a zoom-in to the dimuon selection region, defined by values of BDT score bigger than $BDTcut_1$, indicated by the solid blue line.

In this region, the total number of events in the testing background sample (1/2 fullMC) that passed this cut was 154. Table 3.7 shows the composition for the modes that are present after the cut.

ν interaction mode	**Number of events**
RES	76
DIS	47
MEC	14
QE	8
Coherent	8
Offbeam	1
Dirt	0

Table 3.7: *Events composition for events that passed the BDT cut in the testing background sample.*

The fact that the resonance interaction mode is dominant could be explain by the neutrino cross-section (see Figure 3.38) given that to get dimuons the energy required is high. For example, for values bigger than 1.5 GeV (see Figure 2.10 related to the incoming neutrino energy in the dimuon sample), the RES curve dominates, followed by the DIS mode. Additionally, there were a few events from coherent pion production (see Section 1.6.5).

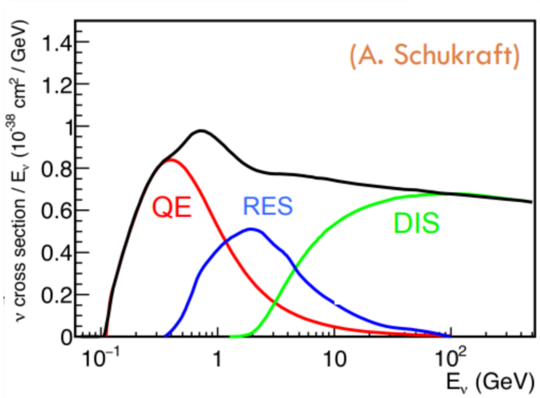


Figure 3.38: *Neutrino cross-section for different ν interaction modes, taken from A. Schukraff, G. Zeller [29].*

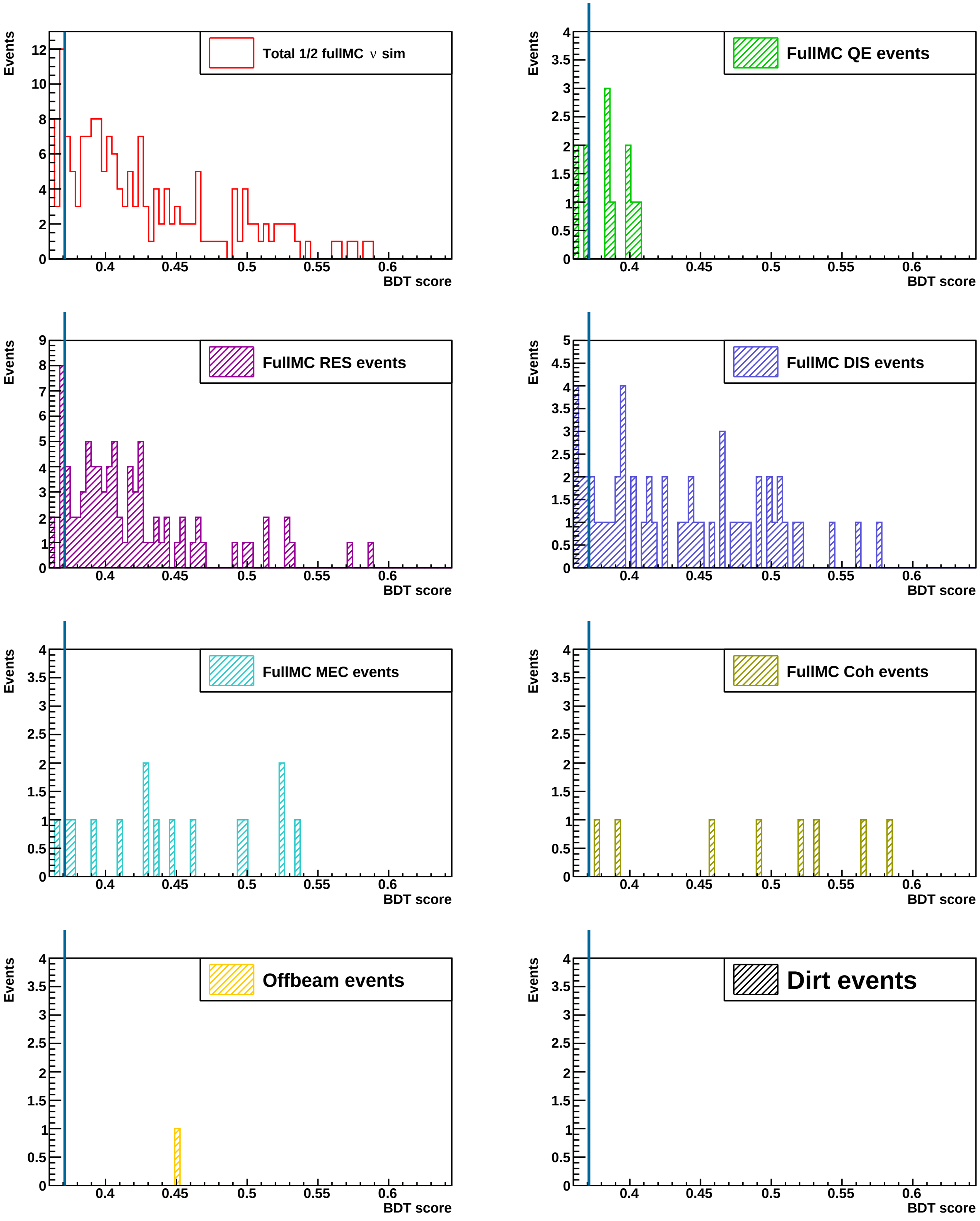


Figure 3.39: *Background composition for the different neutrino interaction modes in the BDT cut region, solid line indicates the BDT cut value.*

3.4.4 Background composition anti-cut region for $BDTcut_2$

The background composition in this region for the $BDTcut_2$ is:

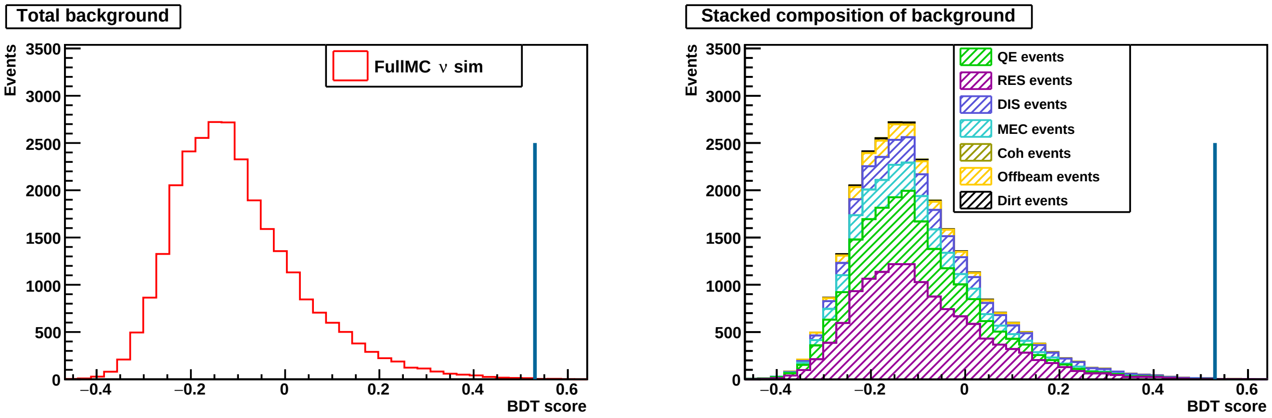


Figure 3.40: *Total BDT testing background distribution (left). Background composition for the different neutrino interaction modes (right), solid line indicates the BDT cut value.*

Table 3.8 shows the number of events for each neutrino interaction mode present in the BDT testing background distribution in the anti-cut region for $BDTcut_2$.

ν **interaction mode**	**Number of events**
RES	13082
QE	7034
DIS	3168
MEC	3108
Off-beam	1398
Dirt	169
Coherent	76

Table 3.8: *Events in the BDT anti-cut region for $BDTcut_2$ in the testing background sample.*

3.4.5 Background composition BDT selection region for $BDTcut_2$

Figure 3.41 presents a zoom-in to the dimuon selection region, defined by values of BDT score bigger than $BDTcut_2$, indicated by the solid blue line.

In this region, the total number of events in the testing background sample (1/2 fullMC) that passed this cut was 10. Table 3.9 shows the composition for the modes that are present after the cut.

ν interaction mode	**Number of events**
RES	3
DIS	3
Coherent	3
MEC	1

Table 3.9: *Events composition for events that passed the BDT cut in the testing background sample.*

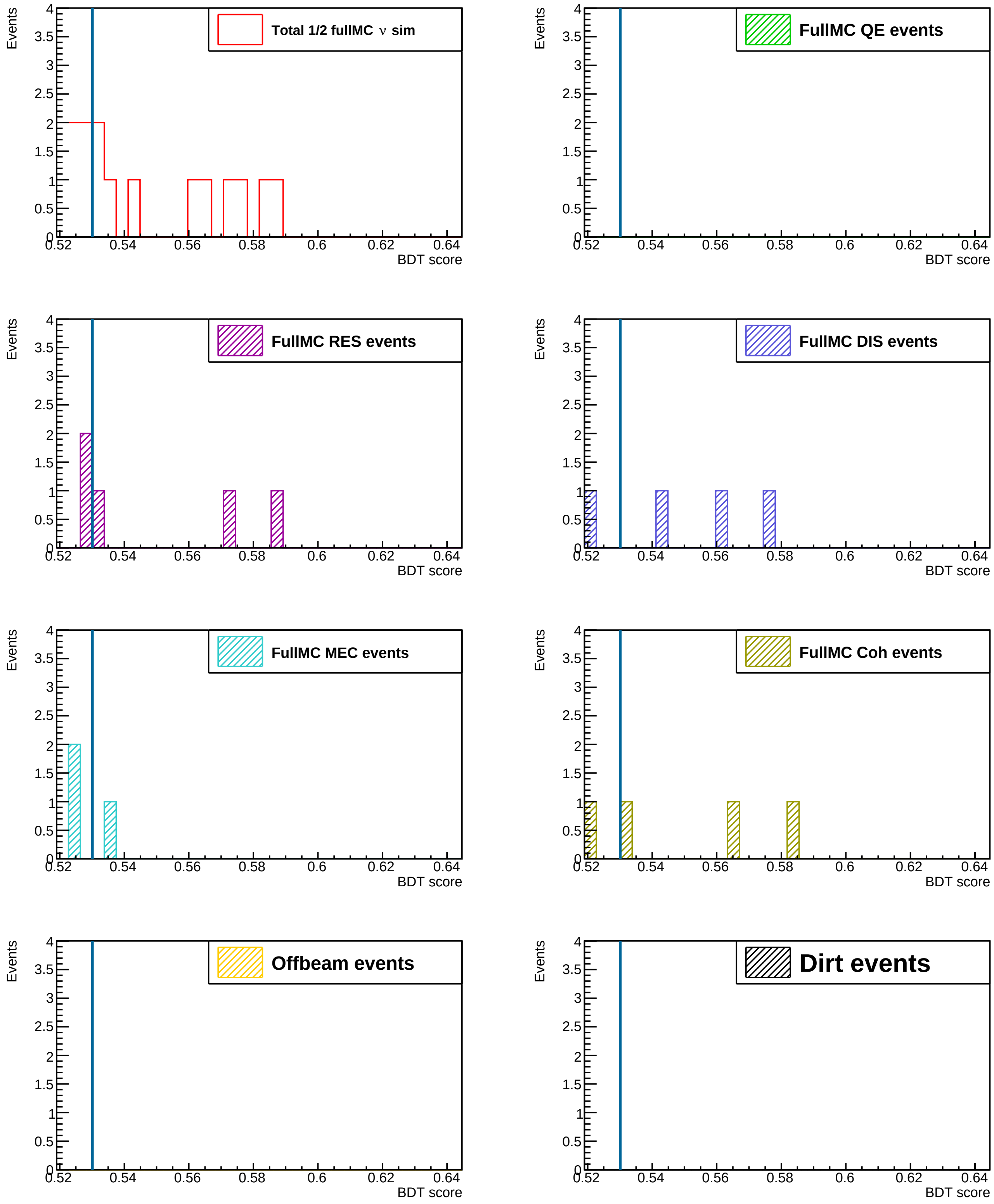


Figure 3.41: *Background composition for the different neutrino interaction modes in the BDT cut region, solid line indicates the BDT cut value.*

3.4.6 Signal estimation

For first BDT cut

Once the estimation of the background is done, it is possible to calculate the estimated detected dimuon signal present in the open data set run 1. Tables 3.10 and 3.11 show the results from the BDT selection per each BDT cut. The open data set run 1 was generated with a POT of 4.5e19. The background sample always has the addition of the off-beam cosmic rays and dirt sample events, even if it is not explicitly added in the name (mainly the MC part is mentioned in the name). Only statistical errors are shown in Tables 3.10 and 3.11.

The normalized background estimation was obtained using the normalization discussed in section 3.3.8 more exactly equation 3.12). The meaning for each column is defined as follows:

- Background sample: identify the sample chosen for training and testing the background.
- $S + B$: number of events that passed the BDT score cut in the open data set. Where S are the signal candidates and B are the background candidates.
- B_T: number of events that passed the BDT score cut in the testing background sample, half fullMC. This number needs to be renormalized later to get the actual estimation of the background, but it is presented to explain the intermediate step.
- B_e : is the estimated background. It is obtained from the normalization of B_T using the equation 3.12.
- $(S + B) - B_e$: indicates the detected dimuon signal.
- S_d: is the number of events that passed the BDT score cut in the training signal, dimuon MC sample
- ϵ: is the efficiency of the dimuon signal sample. It was explained in section 3.3.6.

- $\frac{(S+B)-B_e}{\epsilon}$: is the estimated detected dimuon signal. It considers the efficiency of the dimuon signal during the BDT analysis.

This study defines the estimation of the detected dimuon signal following the expression:

$$Estimation_{\mu\mu} = \frac{(S+B)-B_e}{\epsilon}. \tag{3.17}$$

The results for the BDT selection region for the $BDTcut_1$ are:

Background	$S+B$	B_T	B_e	$(S+B)-B_e$	S_d	ϵ	$\frac{(S+B)-B_e}{\epsilon}$
FullMC	15 ± 3.9	154 ± 12.4	16.1 ± 1.3	-1.1 ± 5.2	337	0.021 ± 0.001	-52 ± 240

Table 3.10: *Result for fullMC CV sample using the open data set run 1 for the* $BDTcut_1$.

The statistical error has been added and properly propagated through the calculations. The estimation of the background and the data are statistically consistent.

For second BDT cut

The results for the BDT selection region for the $BDTcut_2$ are:

Background	$S+B$	B_T	B_e	$(S+B)-B_e$	S_d	ϵ	$\frac{(S+B)-B_e}{\epsilon}$
FullMC	2 ± 1.4	10 ± 3.2	1.1 ± 0.3	0.9 ± 1.7	94	0.006 ± 0.001	158 ± 309

Table 3.11: *Result for fullMC CV sample using the open data set run 1 for the* $BDTcut_2$.

The statistical error has been added and properly propagated through the calculations. The estimation of the background and the data are statistically consistent.

Chapter 4

Systematic uncertainties

To study the different sources of systematic uncertainties, it is necessary to identify the sources of uncertainty present in the μB experiment:

- Detector modeling
- Neutrino beam or flux
- Interaction models related to Geant4
- Cross sections models used in the GENIE generator

For the detector modeling case, a set of ten different samples is going to be created using a patched version of GENIE, which will modify the original parameters and conditions of the detector modeling used for the CV sample. The parameters are varied one parameter at a time, which is why this kind of sample is going to be called **Unisim**. Once again the updated gLEE samples were used for this case. Section 4.2 will explain in more detail this set of detector variations.

For the next three sources of uncertainty, a **Multisim** analysis was performed. It is called like that because a set of parameters are varied simultaneously around its central value from an independent Gaussian distribution to create new “universes” where the CV conditions for flux, cross section and interaction models behave in a different way.

The following Table shows a summary of how these variations were performed for each source of uncertainty:

Source of uncertainties	**Method**
Detector modeling	Unisim
Neutrino beam or flux	Multisim
Interaction models (Geant4)	Multisim
Cross section models (GENIE)	Multisim

Table 4.1: *Sources of uncertainties and simulation method.*

4.1 Unisim and Multisim

Unisim

A unisim is a variation made to a single parameter for a given source of uncertainty. In the case of detector variations, that variable could be, for example, the wire response in the x-axis, a modification of the space charge effect (SCE) map, etc.

Then the difference between the observations in this unisim variation and the CV sample gives the uncertainty, and is possible to quantify that value using the covariance matrix [27]:

$$E_{ab}^{unisim} = \sum_{u=1}^{U} \left(n_a^{CV} - n_a^u\right)\left(n_b^{CV} - n_b^u\right), \tag{4.1}$$

where $n_{a,b}^u$ is the measurement count of events the unisim variation “u”, $n_{a,b}^{CV}$ is the total count in the CV sample and U is the number of detector variation samples used.

Multisim

A multisim is a variation made to a set of parameters at the same time. Each parameter variation could be defined as a unisim individually, the difference is that a large number of universes is going to be produce for each of this unisims. To avoid generating a large number of different samples with these variations, a re-weighting method was used. This allows getting hundreds of weight vectors for their corresponding hundreds of universes. The Equation 4.1 converts for the multisim case to:

$$E_{ab}^{multisim} = \frac{1}{N}\sum_{m=1}^{M}\left(\sum_{k=1}^{N} W_{ka}^{CV} - \sum_{k=1}^{N} W_{ka}^{m}\right)\left(\sum_{k=1}^{N} W_{kb}^{CV} - \sum_{k=1}^{N} W_{kb}^{m}\right), \tag{4.2}$$

where W_{ka}^{m} is the k-th weight for the a-th observation in the multisim variation “m” and W_{ka}^{CV} is the k-th weight for the a-th observation in the CV sample. M is the number of multisims available, and N is the number of universes. A summation must be done in the k index to get all the contributions from the correspondent universes.

The multisim method can be used for flux and cross-section uncertainties because the same simulated events, with different weights, can be propagated through the same detector simulation model. By contrast, detector-related uncertainties require the creation of different physics detector models for the simulated events, which can not be related to one another by reweighting. Reweighting methods are much more efficient when available.

4.2 Detector variations

GENIE unisims for detector variations

The CV sample is one MC neutrino interaction simulation for μB. It was chosen to be the main MC, and with respect to this sample and tuning, most of the analyses are performed. For producing an analysis of the systematic uncertainties related to the detector properties, it is necessary to create different MC samples with different properties than the CV sample. A set of ten detector variation samples were used. Each sample corresponds to one universe where the single parameter was changed with a usual variation of the order $\pm 1\sigma$.

The strategy is to reproduce the same analysis shown in Section 3.3 with each detector variation sample replacing the CV sample. A re-training is performed in each universe, given that the background training sample changes depending on the MC used. Because of this re-training, the BDT cut is going to change as well see Table 4.4. Table 4.2 shows the SAM definitions used to produce the samples. The updated gLEE samples were used for the 10 cases.

SAM def	**Kind**
prodgenie_bnb_nu_overlay_DetVar_CV_reco2_v08_00_00_38_run1_reco2_reco2	Active Volume modification
prodgenie_bnb_nu_overlay_DetVar_LYAttenuation_v08_00_00_38_run1_reco2_reco2	Light Yield
prodgenie_bnb_nu_overlay_DetVar_LYDown_v08_00_00_37_run1_reco2_reco2	Light Yield
prodgenie_bnb_nu_overlay_DetVar_LYRayleigh_v08_00_00_37_run1_reco2_reco2	Light Yield
prodgenie_bnb_nu_overlay_reco2_detvar_WireModThetaXZ_run1_reco2	Wire modification
prodgenie_bnb_nu_overlay_reco2_detvar_WireModThetaYZ_run1_reco2	Wire modification
prodgenie_bnb_nu_overlay_v08_00_00_51_pandora_reco2_detvar_WireModX_run1_reco2	Wire modification
prodgenie_bnb_nu_overlay_reco2_detvar_WireModYZ_run1_reco2	Wire modification
prodgenie_bnb_nu_uboone_overlay_detvar_SCE_run1_reco2	Alternative SC map
high_stats_bnb_nu_detvar_run1_recomb2_run1_reco2	Recombination

Table 4.2: *Sam definitions for the detector variation samples.*

Cases

There are different cases of detector variations:

SCE and Recombination: Changes in the MCC9 simulation of the TPC active volume.

- Recombination: in the simulation of the Ar ionization, changes the values of the α and β parameters used in the modified box model [27], see Figure 4.1.
- Space Charge Effect (SCE): uses a different model to simulate the electric field and drift current of electrons inside the TPC, see Figure 4.2.

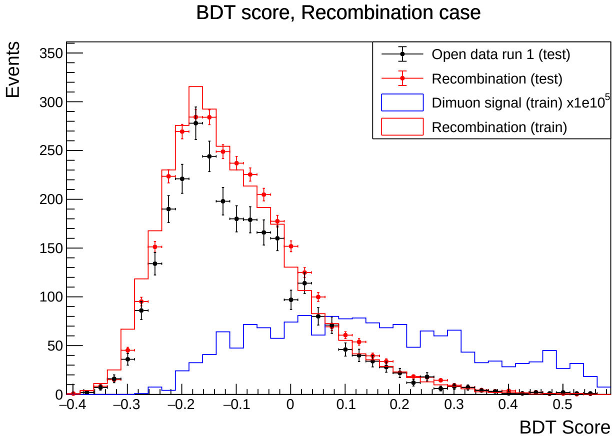


Figure 4.1: *Recombination sample.*

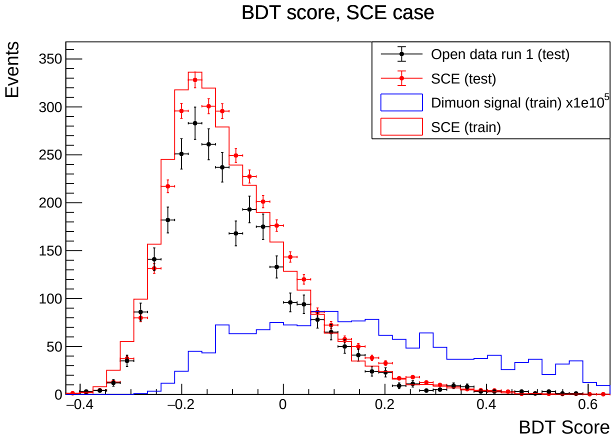


Figure 4.2: *SCE sample.*

Active volume modification: Different description of the active volume in the TPC:

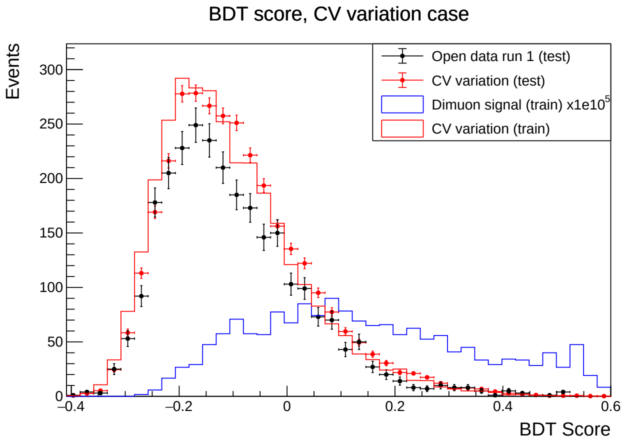


Figure 4.3: *CV variation sample.*

Light yield: is related to the response of the photomultiplier tubes to the light yield. Three sub-cases are studied:

- Light yield down: reduces the whole light yield of energy deposition by 25%, see Figure 4.4.
- Light yield attenuation: reduces the whole light yield depending on the location inside the detector, see Figure 4.5.
- Light yield Rayleigh: modifies the Rayleigh scattering length in the CV by 50% [27], see Figure 4.6.

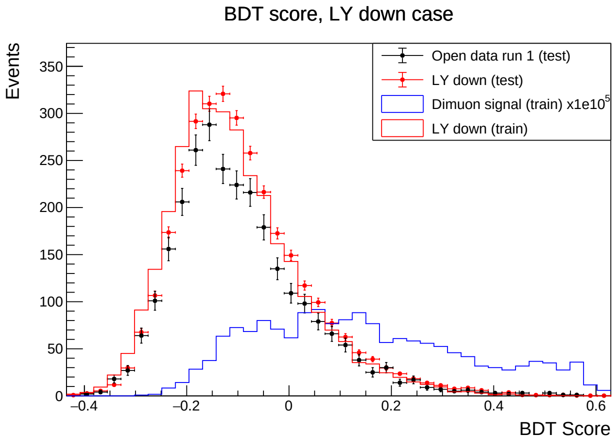


Figure 4.4: *Light Yield down sample.*

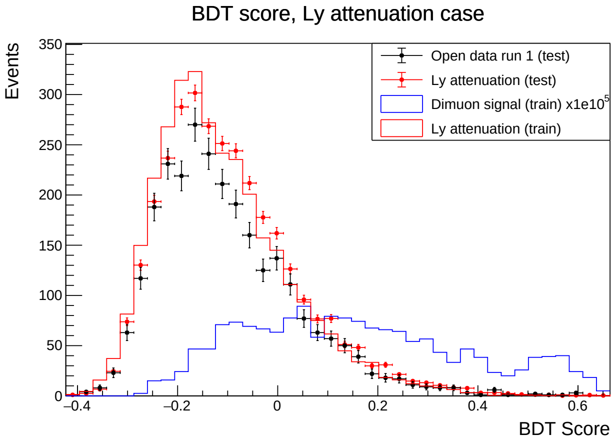


Figure 4.5: *Light Yield attenuation sample.*

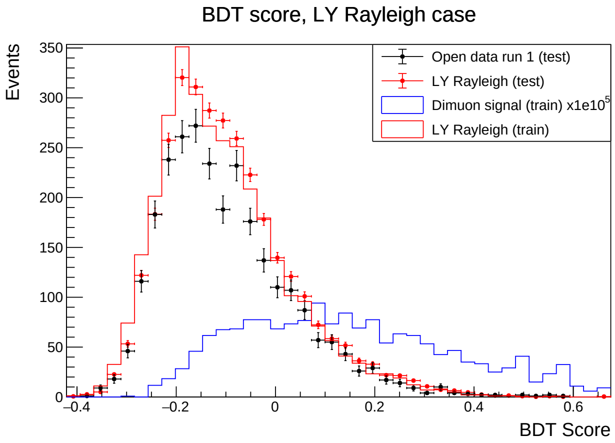
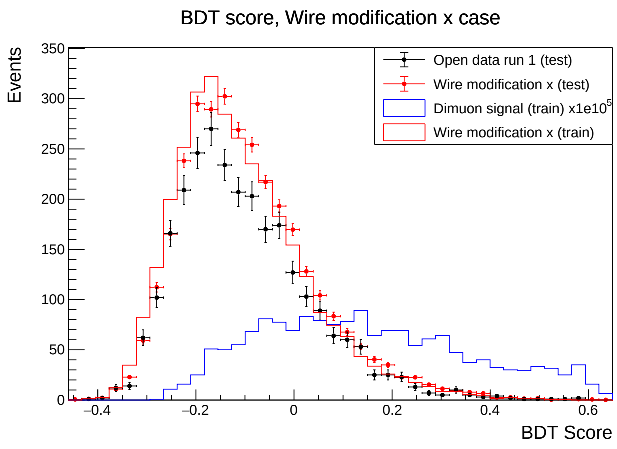


Figure 4.6: *Light Yield Rayleigh sample.*

Wire modification: Transformations as a function of the energy deposition and direction of propagation of a track are applied to the Time Projection Chamber (TPC) electric signal waveforms based on data/MC ratios observations [27], see Figures 4.7, 4.8, 4.9, and 4.10.



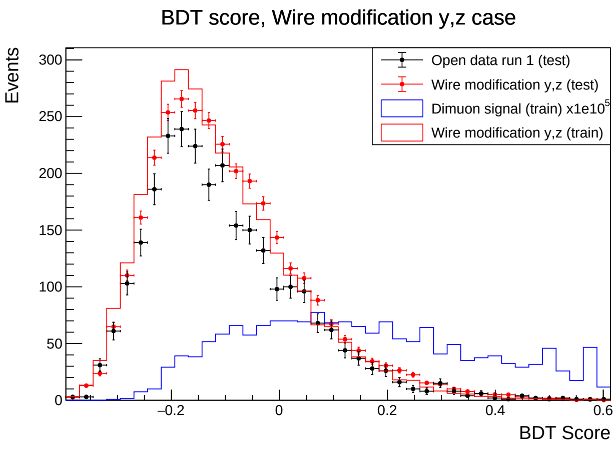
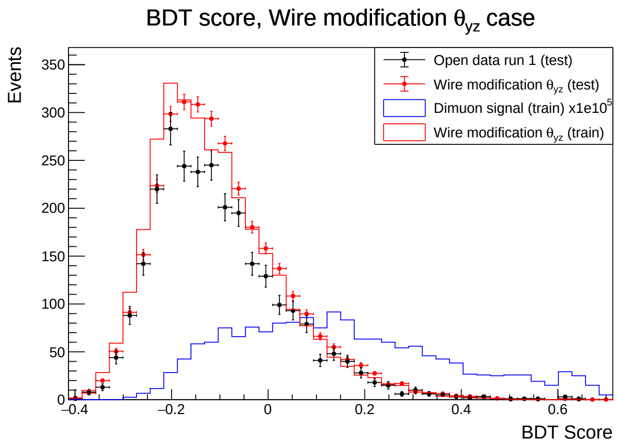


Figure 4.7: *Wire modification x axis sample.*

Figure 4.8: *Wire modification y, z axis sample.*



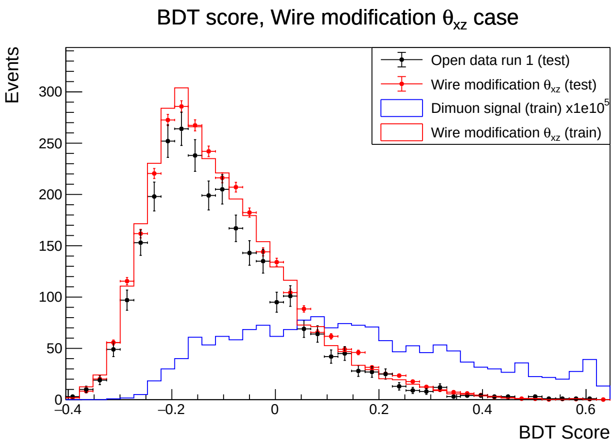


Figure 4.9: *Wire modification θ_{yz} sample.*

Figure 4.10: *Wire modification θ_{xz} sample.*

Detector variation POTs

Next Table summarize all the POT used to create the CV and detector variation samples, these values are important because were used to compare the dimuon events (see Section 3.3.6) and to compare the data events (see Section 3.3.7) with these MC simulation samples, more specifically when using normalization factors.

CV and detector variation sample	**POT**
FullMC	1.28344×10^{21}
Recombination	5.99993×10^{20}
SCE	5.94273×10^{20}
CV variation	6.11241×10^{20}
Light yield down	6.02007×10^{20}
Light yield attenuation	6.06413×10^{20}
Light yield Rayleigh	6.11665×10^{20}
Wire modification x	6.05409×10^{20}
Wire modification y,z	5.97524×10^{20}
Wire modification θ_{yz}	6.00586×10^{20}
Wire modification θ_{xz}	1.19800×10^{21}

Table 4.3: *CV and detector variations samples with its correspondent POT.*

4.3 Background estimation

As seen in section 3.3.8 it is possible to count the number of events in the testing background sample that passed the BDT cut. Then, given that there are different universes, depending on the detector variation of interest, this value is going to change. That is because the training changes depending on the universe used. In other words, each detector variation universe has its own training and testing. Two BDT cuts are applied similarly to Section 3.4.1 for the CV sample. The requirements are the same, the first criteria is going to explore a BDT cut that allows more background to pass and at the same time more efficiency to be obtained. This requires the estimation of the background to be fixed to a value as close as possible to 16. The second case requires minimizing the estimated background. This is that the BDT cut is such that the estimation of the background is equal to 1. Tables 4.4 and 4.5 show the BDT cuts per each detector variation. BDT cut values for the first case:

Background sample	**Optimized BDT score cut**
FullMC	0.53
Recombination	0.49
SCE	0.57
CV variation	0.53
Light yield down	0.55
Light yield attenuation	0.62
Light yield Rayleigh	0.55
Wire modification x	0.58
Wire modification y,z	0.56
Wire modification θ_{yz}	0.62
Wire modification θ_{xz}	0.58

Table 4.4: *BDT score cut values for CV and detector variation samples for first case.*

BDT cut values for the second case:

Background sample	**Optimized BDT score cut**
FullMC	0.37
Recombination	0.34
SCE	0.37
CV variation	0.36
Light yield down	0.36
Light yield attenuation	0.38
Light yield Rayleigh	0.36
Wire modification x	0.37
Wire modification y,z	0.40
Wire modification θ_{yz}	0.37
Wire modification θ_{xz}	0.37

Table 4.5: *BDT score cut values for CV and detector variation samples for the second case.*

4.4 Estimated signal

4.4.1 Estimated signal for open data set run 1 for $BDTcut_1$

As discussed in section 3.4.6 using the expression 3.17, it is possible to determine the estimated signal in the open data set, in this case, run 1. Table 4.6 shows the events that passed the BDT cut in the open data set, MC sample, the estimation of the background, the efficiency of the dimuon signal (see 3.3.6), and the last column with the estimated dimuon signal present in the open data set run 1 for $BDTcut_1$.

Background sample	$S+B$	B_T	B_e	$(S+B)-B_e$	ϵ	$\frac{(S+B)-B_e}{\epsilon}$
FullMC	15 ± 3.9	154 ± 12.4	16.1 ± 1.3	-1.1 ± 5.2	0.021 ± 0.001	-52 ± 240
Recombination	16 ± 4.0	79 ± 8.9	16.1 ± 1.8	-0.1 ± 5.8	0.022 ± 0.001	-5 ± 269
SCE	16 ± 4.0	79 ± 8.9	16.2 ± 1.8	-0.2 ± 5.8	0.022 ± 0.001	-8 ± 266
CV variation	16 ± 4.0	80 ± 8.9	16.0 ± 1.8	-0.0 ± 5.8	0.022 ± 0.001	-2 ± 268
Light yield down	17 ± 4.1	80 ± 8.9	16.2 ± 1.8	0.8 ± 5.9	0.021 ± 0.001	36 ± 288
Light yield attenuation	16 ± 4.0	80 ± 8.9	16.2 ± 1.8	-0.2 ± 5.8	0.022 ± 0.001	-8 ± 265
Light yield Rayleigh	16 ± 4.0	80 ± 8.9	16.0 ± 1.8	-0.0 ± 5.8	0.022 ± 0.001	-1 ± 261
Wire modification x axis	15 ± 3.9	80 ± 8.9	16.2 ± 1.8	-1.2 ± 5.7	0.021 ± 0.001	-55 ± 266
Wire modification y,z axis	14 ± 3.7	79 ± 8.9	16.2 ± 1.8	-2.2 ± 5.6	0.020 ± 0.001	-109 ± 274
Wire modification θ_{yz}	15 ± 3.9	79 ± 8.9	16.1 ± 1.8	-1.1 ± 5.7	0.021 ± 0.001	-52 ± 268
Wire modification θ_{xz}	19 ± 4.4	147 ± 12.1	16.0 ± 1.3	3.0 ± 5.7	0.022 ± 0.001	138 ± 272

Table 4.6: *Estimated signal for CV and detector variations for* $BDTcut_1$.

Systematic uncertainty for detector variations

Table 4.6 shows the estimation of the background in each detector variation universe. To calculate the systematic uncertainty related to that measurement, it is necessary to compare with respect to the CV estimation. The uncertainties for the $BDTcut_1$ case:

Detector variation	**Uncertainty for background estimation**
Recombination	0.00
SCE	0.08
CV variation	0.06
Light yield down	0.14
Light yield attenuation	0.07
Light yield Rayleigh	0.08
Wire modification X axis	0.06
Wire modification Y,Z axis	0.06
Wire modification θ_{yz}	0.02
Wire modification θ_{xz}	0.07
Total	**0.23**

Table 4.7: *Detector variation systematic uncertainties for* $BDTcut_1$.

4.4.2 Estimated signal for open data set run 1 for $BDTcut_2$

Table 4.8 shows the events that passed the BDT cut in the open data set, MC sample, the estimation of the background, the efficiency of the dimuon signal (see 3.3.6) and the last column with the estimated dimuon signal present in the open data set run 1 for $BDTcut_2$.

Background sample	$S+B$	B_T	B_e	$(S+B)-B_e$	ϵ	$\frac{(S+B)-B_e}{\epsilon}$
FullMC	2 ± 1.4	10 ± 3.2	1.1 ± 0.3	0.9 ± 1.7	0.006 ± 0.001	158 ± 309
Recombination	4 ± 2.0	5 ± 2.2	1.0 ± 0.5	3.0 ± 2.5	0.006 ± 0.001	518 ± 486
SCE	1 ± 1.0	5 ± 2.2	1.0 ± 0.5	-0.0 ± 1.5	0.004 ± 0.001	-6 ± 384
CV variation	0 ± 0.0	5 ± 2.2	1.0 ± 0.4	-1.0 ± 0.4	0.005 ± 0.001	-189 ± 62
Light yield down	1 ± 1.0	5 ± 2.2	1.0 ± 0.5	-0.0 ± 1.5	0.004 ± 0.001	-4 ± 338
Light yield attenuation	0 ± 0.0	6 ± 2.4	1.2 ± 0.5	-1.2 ± 0.5	0.001 ± 0.000	-1129 ± 178
Light yield Rayleigh	2 ± 1.4	5 ± 2.2	1.0 ± 0.4	1.0 ± 1.9	0.006 ± 0.001	180 ± 355
Wire modification x axis	0 ± 0.0	5 ± 2.2	1.0 ± 0.5	-1.0 ± 0.5	0.003 ± 0.000	-320 ± 95
Wire modification y,z axis	2 ± 1.4	5 ± 2.2	1.0 ± 0.5	1.0 ± 1.9	0.005 ± 0.001	209 ± 426
Wire modification θ_{yz}	1 ± 1.0	5 ± 2.2	1.0 ± 0.5	-0.0 ± 1.5	0.004 ± 0.001	-5 ± 383
Wire modification θ_{xz}	2 ± 1.4	10 ± 3.2	1.1 ± 0.3	0.9 ± 1.8	0.005 ± 0.001	169 ± 347

Table 4.8: *Estimated signal for CV and detector variations for* $BDTcut_2$.

Systematic uncertainty for detector variations

Table 4.8 shows the estimation of the background in each detector variation universe. To calculate the systematic uncertainty related to that measurement, it is necessary to compare with respect to the CV estimation. The uncertainties for the $BDTcut_2$ case:

Detector variation	**Uncertainty for background estimation**
Recombination	0.03
SCE	0.02
CV variation	0.04
Light yield down	0.03
Light yield attenuation	0.17
Light yield Rayleigh	0.04
Wire modification X axis	0.04
Wire modification Y,Z axis	0.02
Wire modification θ_{yz}	0.03
Wire modification θ_{xz}	0.05
Total	**0.20**

Table 4.9: *Detector variation systematic uncertainties for* $BDTcut_2$.

4.5 Flux variations

4.5.1 BNB

For this study, the Booster Neutrino Beam at Fermilab was used. Different situations are taken into account to create sets of flux variations related to the BNB conditions when the p + Be, Al collisions happen.

These cases are considered unisims by themselves. The **skin effect** is related to how depth the currents reach into the horn conductor in the beam. The **nucleon interaction** for cases like pion-nucleus and nucleon-nucleus interactions happening in the targets. And the **hadron production** measurements for charged pions or kaons [27]. See Table 4.10 for a summary of all the unisims explored in this study.

It was needed to re-weight the fullMC CV sample. A set of 1000 weights were obtained related to 1000 universes for each unisim case (see Table 4.11). Re-process and create new samples from scratch would cost a lot of computing resources, to avoid that, is needed to weight the CV sample events using a μB specific patched version of GENIE. For each event in the central-value sample, there is a weight W^u:

$$W^u = \frac{P^u}{P^{CV}}, \tag{4.3}$$

where u refers to the alternative universe, P^u the probability of generating that event in u and P^{CV} the equivalent probability in the central-value sample [27].

The neutrino interaction MC that was initially used had a tree with the map corresponding to the reweighting vectors, but it was empty. Then, the module SinglePhoton.cc should be modified to be able to save the map weights, which include the vectors for the re-weighting for each event. Also, it is necessary to modify the .fcl, which calls this module run_SinglePhoton.fcl, and define the boolean variable as true.

4.5.2 Systematic uncertainties from neutrino beam or flux

To calculate the uncertainties for each flux unisim, what was done was to count how many events were in the testing BDT background sample (in the correspondent universe) and compare that with the counting in the CV sample. Figure 4.11 shows the distribution for the first unisim case, corresponding to the skin effect flux unisim. In solid blue is presented the CV counting, even though is not possible to see clearly due to overlap, this value is 153 (that value can be confirmed in the number of entries of the histogram on the top right, given that the CV is the first distribution plotted) and in gray are the 1000 universes count on top of the CV.

As an example of this calculation, the unisim for the skin effect is presented for the first BDT cut case:

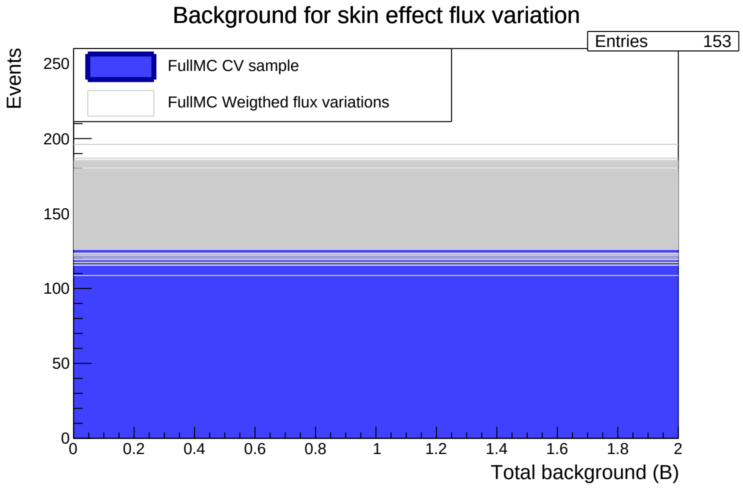


Figure 4.11: *Skin effect unisim for CV and weighted universes.*

Now, using the equation 4.2, every unisim universe distribution is subtracted to the CV histogram, and each contribution is squared and stored in a 1x1 covariance matrix (Figure 4.12). The square root of the "diagonal" of this matrix gives the uncertainty per each unisim. In other words, adding the contribution of each universe in quadrature gives the value for the uncertainty of the estimation of the background for each unisim.

The off-beam and dirt events are not included in this calculation, given that there are no weights for these cases. Then it is needed to propagate this error to get the systematic uncertainty related to each flux unisim for the estimated detected dimuon signal (see Table 4.11).

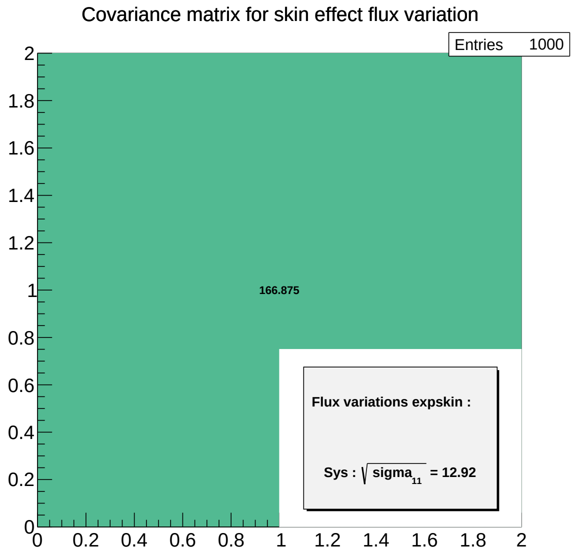


Figure 4.12: *Covariance matrix for the skin effect unisim.*

4.6 Systematic uncertainties from interaction models (Geant4)

After the initial neutrino interaction happens, resulting particles may re-interact with other Ar atom in the surroundings. In the re-weighting process shown in section A.4.5 the toolkit Geant4 is in charge of model and evaluate the systematic uncertainties related to this. The Geant4 re-weighting toolkit varies the total cross-section modeling for pions-nucleus and proton-nucleus [27]. Similarly to the flux case, the variations made to the re-interaction models include weights for 1000 universes.

The reinteractions section in Table 4.11 shows the results of the uncertainty in the estimated dimuon signal for the case of π^- and proton (p); the other cases were obtained to be zero or neglected.

4.7 Systematic uncertainties from cross-section models (GENIE)

A set of parameters is varied simultaneously around the CV sample using a normal distribution of 1σ. This is a multi-sim re-weighting process used to evaluate the uncertainties related to the GENIE model cross sections. Unlike all the other systematic uncertainties assessed in this study (which use 1000 universes), the GENIE case used 600 universes, the recommended number by different analyses.

4.8 Summary of neutrino flux, GENIE, and Geant4 as source of uncertainties.

For the case of detector variations, the uncertainty is calculated taking the square root of the difference between the estimated signal for the CV (first element of last column of Tables 4.6 and 4.8) and the same estimation for each detector variation universe. The total is obtained by doing an addition in quadrature. For the flux, re-interactions (Geant4), and GENIE case, the uncertainty was gotten with the counting explained in section A.4.5 and the 1x1 matrix calculation (Figure 4.12) for each unisim. Table 4.10 presents a list of all the variations for the neutrino beam uncertainty case and their explanations.

Flux unisim	**Description**
expskin_FluxUnisim	Skin effect, depth electric currents penetrate conductor
horncurrent_FluxUnisim	Horn current in magnetic focusing horns
nucleoninexsec_FluxUnisim	Nucleon total inelastic cross section on Beryllium target
nucleonqexsec_FluxUnisim	Nucleon total quasi-elastic cross-section on Beryllium target
nucleontotxsec_FluxUnisim	Nucleon total cross-section on Beryllium target
pioninexsec_FluxUnisim	Pion total inelastic cross section on Beryllium target
pionqexsec_FluxUnisim	Pion total quasi-elastic cross-section on Beryllium target
piontotxsec_FluxUnisim	Pion total cross-section on Beryllium target
piminus_PrimaryHadronSWCentralSplineVariation	Primary hadron Sanford Wang central spline variation
piplus_PrimaryHadronSWCentralSplineVariation	Primary hadron Sanford Wang central spline variation
kminus_PrimaryHadronNormalization	Primary hadron normalization
kplus_PrimaryHadronFeynmanScaling	Primary hadron Feynman Scaling
kzero_PrimaryHadronSanfordWang	Primary hadron Sanford Wang

Table 4.10: *Variations for estimation of systematic uncertainties related to the neutrino beam or flux [30].*

4.8.1 For $BDTcut_1$

The next Table presents a summary of all the systematic uncertainties in the background estimation for each unisim for the flux, GENIE, and Geant4 cases, for $BDTcut_1$.

Variation	**Unisim**	**Background estimation uncertainty**
Flux	expskin_FluxUnisim	1.35
	horncurrent_FluxUnisim	0.14
	nucleoninexsec_FluxUnisim	0.12
	nucleonqexsec_FluxUnisim	0.41
	nucleontotxsec_FluxUnisim	0.09
	pioninexsec_FluxUnisim	0.17
	pionqexsec_FluxUnisim	0.11
	piontotxsec_FluxUnisim	0.13
	piminus_PrimaryHadronSWCentralSplineVariation	0.00
	piplus_PrimaryHadronSWCentralSplineVariation	0.91
	kminus_PrimaryHadronNormalization	0.00
	kplus_PrimaryHadronFeynmanScaling	0.28
	kzero_PrimaryHadronSanfordWang	0.00
	Total	**1.73**
Reinteractions	reinteractions_piminus_Geant4	0.16
	reinteractions_piplus_Geant4	0.56
	reinteractions_proton_Geant4	0.43
	Total	**0.72**
GENIE	**All_UBGENIE**	**2.83**
	TOTAL	**3.40**

Table 4.11: *Systematic uncertainties for flux, GENIE and Geant4, for* $BDTcut_1$.

4.8.2 For $BDTcut_2$

The next Table presents a summary of all the systematic uncertainties in the background estimation for each unisim for the flux, GENIE, and Geant4 cases, for $BDTcut_2$.

Variation	**Unisim**	**Background estimation uncertainty**
Flux	expskin_FluxUnisim	0.09
	horncurrent_FluxUnisim	0.01
	nucleoninexsec_FluxUnisim	0.01
	nucleonqexsec_FluxUnisim	0.03
	nucleontotxsec_FluxUnisim	0.00
	pioninexsec_FluxUnisim	0.01
	pionqexsec_FluxUnisim	0.01
	piontotxsec_FluxUnisim	0.01
	piminus_PrimaryHadronSWCentralSplineVariation	0.00
	piplus_PrimaryHadronSWCentralSplineVariation	0.10
	kminus_PrimaryHadronNormalization	0.00
	kplus_PrimaryHadronFeynmanScaling	0.03
	kzero_PrimaryHadronSanfordWang	0.00
	Total	**0.14**
Reinteractions	reinteractions_piminus_Geant4	0.04
	reinteractions_piplus_Geant4	0.14
	reinteractions_proton_Geant4	0.04
	Total	**0.15**
GENIE	**All_UBGENIE**	**0.20**
	TOTAL	**0.29**

Table 4.12: *Systematic uncertainties for flux, GENIE and Geant4, for* $BDTcut_2$.

This concludes that the biggest systematic effects are the GENIE variations. The total systematic uncertainty for the estimation of the dimuon signal is ± 3.4 for $BDTcut_1$ and ± 0.4 for $BDTcut_2$. These values are including the total contribution from the detector variation shown in Tables 4.7 and 4.9.

Estimated background with statistical and systematic uncertainties

Additionally to the systematic uncertainties, the next Tables present the statistical uncertainties (from Section 3.4.6) for the estimation of the background in the dimuon production. The measurement in the open data set run 1 is shown in the column $S + B$. In the background estimation (B_e) column, the first uncertainty corresponds to the statistical uncertainty, and the second one corresponds to the total systematic uncertainty. The measurements in the data are less than 2 σs from the estimation of the background, which concludes that no dimuon signal was observed in the open data set run 1.

For the first case for $BDTcut_1$:

Background sample	$S + B$	B_e
FullMC	15	16.1 ± 1.3 ± 3.4

Table 4.13: *Summary background estimation with statistical and systematic uncertainties, for* $BDTcut_1$.

For the second case for $BDTcut_2$:

Background sample	$S + B$	B_e
FullMC	2	1.1 ± 0.3 ± 0.4

Table 4.14: *Summary background estimation with statistical and systematic uncertainties, for* $BDTcut_2$.

Chapter 5

Estimation of dimuon signal in data set run 1

Additionally to the open data set, access to data set run 1 was required. The strategy is to run the analysis with this whole sample and study the results, comparing them with the results shown in Chapters 3 and 4. This data set is around four times bigger than the open data set run 1.

The characteristics of the data set run 1 are shown in Table 5.1. The sample was produced from the reconstruction step to get the final n-tuple.

Data set run 1	
Original number of events	638507
Number of events after pre-selection	9708
POT	1.67×10^{20}
Spills	37273255

Table 5.1: *Data set run 1 characteristics.*

5.1 Normalization

To calculate the normalization factor to compare the data set corresponding to run 1 with the fullMC, it is needed to calculate the normalization factor. Similarly to Section 3.3.7, the test background sample was confirmed by fullMC, off-beam, and dirt events. For the fullMC POT, a factor of 0.5 was used because just half of the whole fullMC sample was used for testing :

$$\mathrm{N_{fullMC}} = \frac{\text{POT data sample}}{\text{POT 1/2 fullMC sample}}, \tag{5.1}$$

where the POT for the data set run 1 sample is known in μB. The POT value for the fullMC sample was calculated by counting the sub-runs in the MC.

$$\mathrm{N_{fullMC}} = \frac{\text{POT data sample}}{\text{POT 1/2 fullMC sample}} = \frac{1.67 \times 10^{20}}{0.5 \times 1.28344 \times 10^{21}} = 0.2602. \tag{5.2}$$

For the normalization with respect to the off-beam (cosmic background) sample, the number of spills should be considered [24]. For the off-beam sample a factor of 0.5 was used because just half of the whole sample was used for testing in combination with the fullMC :

$$\mathrm{N_{offbeam}} = \frac{\text{Spills data sample}}{\text{1/2 Spills off beam sample}} = \frac{37273255}{0.5 \times 28190365} = 2.6444. \tag{5.3}$$

Additionally, the normalization with respect to the dirt events was calculated.

$$\mathrm{N_{dirt}} = \frac{\text{POT data sample}}{\text{POT 1/2 dirt sample}} = \frac{1.67 \times 10^{20}}{0.5 \times 3.23362 \times 10^{20}} = 1.0328. \tag{5.4}$$

Keeping the proportions, the combined normalization is:

$$\mathrm{N} = \frac{\text{Events fullMC}}{\text{Totalevents}} \times 0.2602 + \frac{\text{Events offbeam}}{\text{Totalevents}} \times 2.6444 + \frac{\text{Events dirt}}{\text{Totalevents}} \times 1.0328 = 0.38. \tag{5.5}$$

The estimation of the background is going to change because the normalization is different when the sample POT changes. The calculation is the same as in Section 3.3.8.

5.2 BDT test

The data set run 1 sample (see Table 5.1) was used as a testing signal sample with the CV BDT shown in Figure 3.15. Figure 5.1 presents testing and training overlapped. The test distributions are shown in solid point markers with their corresponding statistical error. The solid histograms correspond to the training.

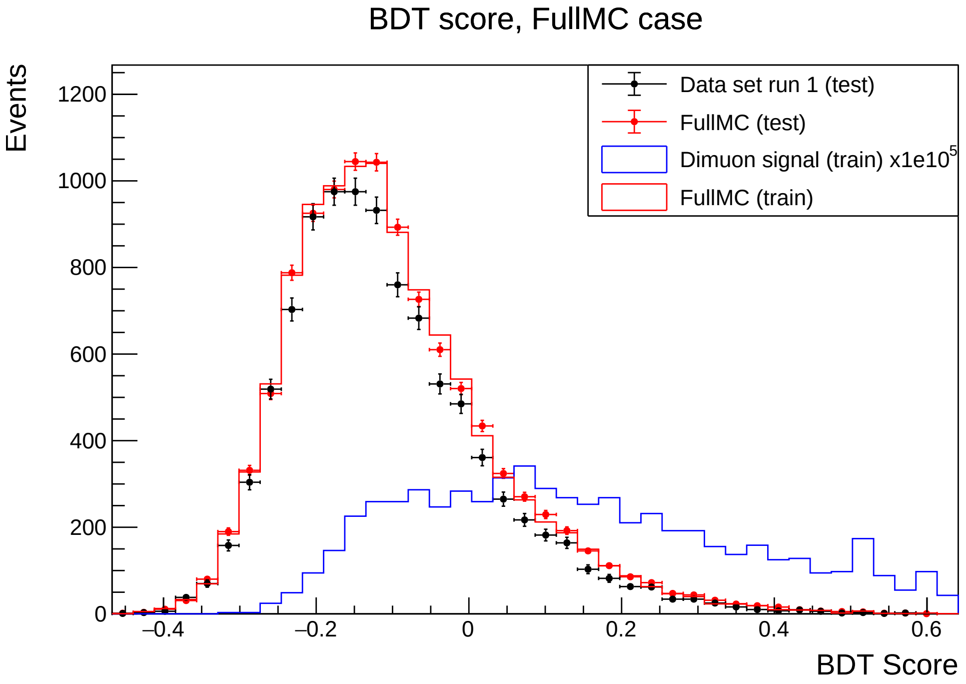


Figure 5.1: *CV BDT tested with the whole data set run 1.*

The estimation of the background is obtained by normalizing the total background. The normalization obtained in Equation 5.5 is used in Equation 3.14. Figure 5.2 shows the data measurement and the estimation of the background as a function of the BDT score.

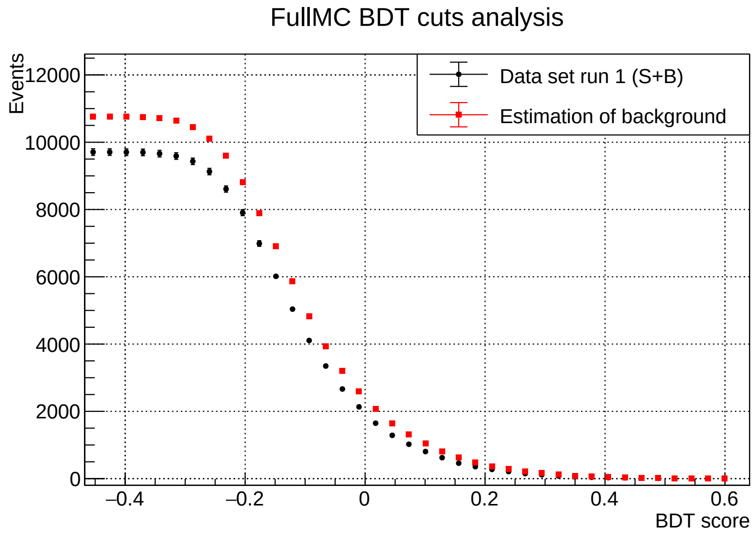


Figure 5.2: *Background estimation and data set run 1 as a function of BDT score.*

There is agreement between the two distributions which indicates no evidence of dimuon observation. A logarithmic scale was applied to the plot to see the high BDT region more clearly.

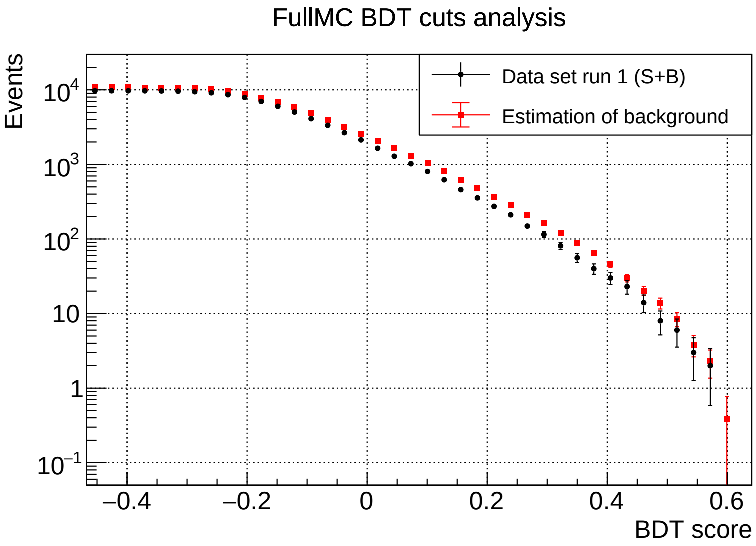


Figure 5.3: *Background estimation and open data set as a function of BDT score.*

The ratio plot shown in Figure 5.4 confirms that the proportions of estimated background and data measurement are kept constant during the whole BDT score values (the ratio plot being a flat line).

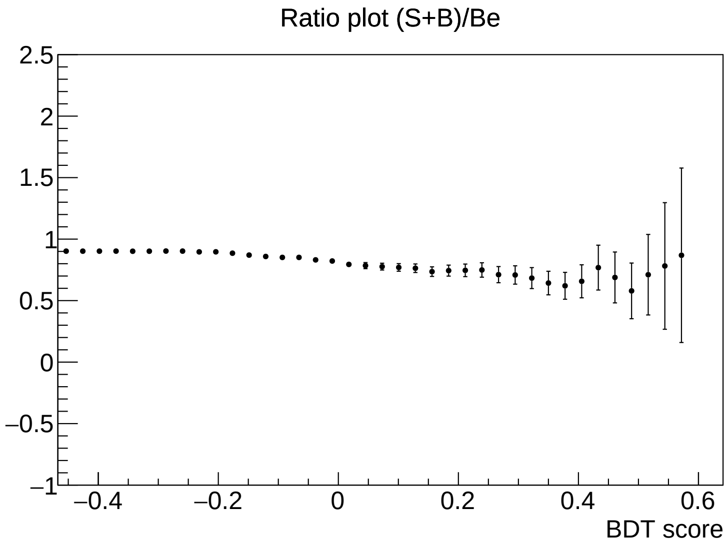


Figure 5.4: *Ratio plot for data set run 1 vs. background estimation as a function of BDT score.*

5.3 Results for $BDTcut_1$

The BDT score cuts are the same as the ones used for the open data set (see Equations 3.15 and 3.16). Given that the normalization changes, the estimation of the background is going to change. The strategy is to obtain a comparison between the estimation of the background and the measurement in the data run 1. The results for the CV MC and all the detector variations (see Chapter 4) for both BDT cuts are shown in the next Sections.

5.3.1 Signal estimation

Background sample	$S+B$	B_T	B_e	$(S+B)-B_e$	ϵ	$\frac{(S+B)-B_e}{\epsilon}$
FullMC	37 ± 6.1	154 ± 12.4	59.1 ± 4.8	-22.1 ± 10.8	0.021 ± 0.001	-1038 ± 445
Recombination	45 ± 6.7	79 ± 8.9	59.0 ± 6.6	-14.0 ± 13.3	0.022 ± 0.001	-648 ± 577
SCE	40 ± 6.3	79 ± 8.9	59.3 ± 6.7	-19.3 ± 13.0	0.022 ± 0.001	-883 ± 540
CV variation	39 ± 6.2	80 ± 8.9	58.8 ± 6.6	-19.8 ± 12.8	0.022 ± 0.001	-916 ± 536
Light yield down	43 ± 6.6	80 ± 8.9	59.5 ± 6.7	-16.5 ± 13.2	0.021 ± 0.001	-795 ± 585
Light yield attenuation	43 ± 6.6	80 ± 8.9	59.3 ± 6.6	-16.3 ± 13.2	0.022 ± 0.001	-744 ± 557
Light yield Rayleigh	42 ± 6.5	80 ± 8.9	58.7 ± 6.6	-16.7 ± 13.0	0.022 ± 0.001	-754 ± 542
Wire modification x axis	39 ± 6.2	80 ± 8.9	59.2 ± 6.6	-20.2 ± 12.9	0.021 ± 0.001	-961 ± 551
Wire modification y,z axis	34 ± 5.8	79 ± 8.9	59.2 ± 6.7	-25.2 ± 12.5	0.020 ± 0.001	-1276 ± 550
Wire modification θ_{yz}	40 ± 6.3	79 ± 8.9	58.9 ± 6.6	-18.9 ± 13.0	0.021 ± 0.001	-903 ± 561
Wire modification θ_{xz}	46 ± 6.8	147 ± 12.1	58.8 ± 4.8	-12.8 ± 11.6	0.022 ± 0.001	-592 ± 503

Table 5.2: *Estimated signal for CV and detector variations for $BDTcut_1$, using the data set run 1.*

5.3.2 Summary systematic uncertainty for detector variations

Detector variation	**Uncertainty for background estimation**
Recombination	0.09
SCE	0.21
CV variation	0.30
Light yield down	0.43
Light yield attenuation	0.17
Light yield Rayleigh	0.37
Wire modification X axis	0.12
Wire modification Y, Z axis	0.14
Wire modification θ_{yz}	0.15
Wire modification θ_{xz}	0.33
Total	**0.81**

Table 5.3: *Detector variation systematic uncertainties for* $BDTcut_1$ *for data set run 1.*

5.3.3 Summary of neutrino flux, GENIE, and Geant4 as source of uncertainties.

The next Table presents a summary of all the systematic uncertainties in the background estimation for each unisim for the flux, GENIE, and Geant4 cases, for $BDTcut_1$.

Variation	**Unisim**	**Background estimation uncertainty**
Flux	expskin_FluxUnisim	5.17
	horncurrent_FluxUnisim	0.54
	nucleoninexsec_FluxUnisim	0.45
	nucleonqexsec_FluxUnisim	1.55
	nucleontotxsec_FluxUnisim	0.35
	pioninexsec_FluxUnisim	0.66
	pionqexsec_FluxUnisim	0.44
	piontotxsec_FluxUnisim	0.50
	piminus_PrimaryHadronSWCentralSplineVariation	0.00
	piplus_PrimaryHadronSWCentralSplineVariation	3.49
	kminus_PrimaryHadronNormalization	0.00
	kplus_PrimaryHadronFeynmanScaling	1.06
	kzero_PrimaryHadronSanfordWang	0.00
	Total	**6.63**
Reinteractions	reinteractions_piminus_Geant4	0.62
	reinteractions_piplus_Geant4	2.15
	reinteractions_proton_Geant4	1.66
	Total	**2.77**
GENIE	**All_UBGENIE**	**10.83**
	TOTAL	**12.99**

Table 5.4: *Systematic uncertainties for flux, GENIE and Geant4, for* $BDTcut_1$ *using data set run 1.*

5.3.4 Estimated background with statistical and systematic uncertainties

Additionally to the systematic uncertainties, the next Tables present the statistical uncertainties (from Table 5.2) for the estimation of the background in the dimuon production. The measurement in the data set run 1 is shown in the column $S + B$.

In the background estimation (B_e) column, the first uncertainty corresponds to the statistical uncertainty, and the second one corresponds to the total systematic uncertainty. The measurements in the data are less than 2 σs from the estimation of the background, which concludes that no dimuon signal was observed in the whole data set run 1.

Background sample	$S + B$	B_e	B_e with total error
FullMC	37	59.1 ± 4.8 ± 13.0	59.1 ± 13.9

Table 5.5: *Summary background estimation with statistical and systematic uncertainties, for* $BDTcut_1$ *using data set run 1.*

With this measurements it is possible to obtain an upper limit for the dimuon detection. The concept is that the experiment could be repeated many times and the results compared. To do that, a Gaussian distribution is used to simulate the estimation of the background (given that the error was estimated). Then, the Poisson distribution could be used to calculate the discrete number of events obtained for the background and the signal (S+B). Doing that simulation a 1000 times resulted in an upper limit:

$$\text{Observed dimuon} = 4.9, \tag{5.6}$$

for the observed dimuon signal, with a 95% confidence level. The efficiency correction for this value (using the efficiency shown in Table 5.2), results:

$$\text{Produced dimuon} = 233, \tag{5.7}$$

which is the estimated production of dimuon signal in the data set run 1. This result allows to obtain the parameter X in Equation 1.1:

$$X = \frac{\sigma(\mu\mu)}{\sigma_{SM}(\mu\mu)} = \frac{\frac{233}{638507}}{\frac{10}{1026120}} = 37, \tag{5.8}$$

where on the numerator appears the dimuon production relative to the total number of events in the data set run 1 and in the denominator is the dimuon production relative to the total number of events in the fullMC.

The upper limit for the production of dimuon signal that this study estimates is 37 times the one the theoretical framework (GENIE) addresses. If there are BSM channels of dimuon production, these are going to produce a maximum of 37 times the number of dimuons expected by GENIE.

5.4 Results for $BDTcut_2$

5.4.1 Signal estimation

Background sample	$S+B$	B_T	B_e	$(S+B)-B_e$	ϵ	$\frac{(S+B)-B_e}{\epsilon}$
FullMC	3 ± 1.7	10 ± 3.2	3.8 ± 1.2	-0.8 ± 2.9	0.006 ± 0.001	-141 ± 480
Recombination	4 ± 2.0	5 ± 2.2	3.7 ± 1.7	0.3 ± 3.7	0.006 ± 0.001	46 ± 643
SCE	1 ± 1.0	5 ± 2.2	3.8 ± 1.7	-2.8 ± 2.7	0.004 ± 0.001	-726 ± 607
CV variation	0 ± 0.0	5 ± 2.2	3.7 ± 1.6	-3.7 ± 1.6	0.005 ± 0.001	-692 ± 229
Light yield down	2 ± 1.4	5 ± 2.2	3.7 ± 1.7	-1.7 ± 3.1	0.004 ± 0.001	-400 ± 665
Light yield attenuation	0 ± 0.0	6 ± 2.4	4.4 ± 1.8	-4.4 ± 1.8	0.001 ± 0.000	-4138 ± 653
Light yield Rayleigh	3 ± 1.7	5 ± 2.2	3.7 ± 1.6	-0.7 ± 3.4	0.006 ± 0.001	-121 ± 593
Wire modification x axis	0 ± 0.0	5 ± 2.2	3.7 ± 1.7	-3.7 ± 1.7	0.003 ± 0.000	-1172 ± 349
Wire modification y,z axis	2 ± 1.4	5 ± 2.2	3.7 ± 1.7	-1.7 ± 3.1	0.005 ± 0.001	-374 ± 615
Wire modification θ_{yz}	1 ± 1.0	5 ± 2.2	3.7 ± 1.7	-2.7 ± 2.7	0.004 ± 0.001	-720 ± 605
Wire modification θ_{xz}	2 ± 1.4	10 ± 3.2	4.0 ± 1.3	-2.0 ± 2.7	0.005 ± 0.001	-372 ± 455

Table 5.6: *Estimated signal for CV and detector variations for $BDTcut_2$, using the data set run 1.*

5.4.2 Summary systematic uncertainty for detector variations

Detector variation	**Uncertainty for background estimation**
Recombination	0.10
SCE	0.08
CV variation	0.16
Light yield down	0.12
Light yield attenuation	0.61
Light yield Rayleigh	0.17
Wire modification X axis	0.14
Wire modification Y, Z axis	0.09
Wire modification θ_{yz}	0.11
Wire modification θ_{xz}	0.16
Total	**0.72**

Table 5.7: *Detector variation systematic uncertainties for* $BDTcut_2$ *for data set run 1.*

5.4.3 Summary of neutrino flux, GENIE, and Geant4 as source of uncertainties.

The next Table presents a summary of all the systematic uncertainties in the background estimation for each unisim for the flux, GENIE, and Geant4 cases for $BDTcut_2$.

Variation	**Unisim**	**Background estimation uncertainty**
Flux	expskin_FluxUnisim	0.35
	horncurrent_FluxUnisim	0.03
	nucleoninexsec_FluxUnisim	0.03
	nucleonqexsec_FluxUnisim	0.12
	nucleontotxsec_FluxUnisim	0.02
	pioninexsec_FluxUnisim	0.05
	pionqexsec_FluxUnisim	0.03
	piontotxsec_FluxUnisim	0.03
	piminus_PrimaryHadronSWCentralSplineVariation	0.00
	piplus_PrimaryHadronSWCentralSplineVariation	0.37
	kminus_PrimaryHadronNormalization	0.00
	kplus_PrimaryHadronFeynmanScaling	0.12
	kzero_PrimaryHadronSanfordWang	0.00
	Total	**0.54**
Reinteractions	reinteractions_piminus_Geant4	0.16
	reinteractions_piplus_Geant4	0.54
	reinteractions_proton_Geant4	0.15
	Total	**0.58**
GENIE	**All_UBGENIE**	**0.76**
	TOTAL	**1.10**

Table 5.8: *Systematic uncertainties for flux, GENIE and Geant4, for $BDTcut_2$ using data set run 1.*

5.4.4 Estimated background with statistical and systematic uncertainties

Additionally to the systematic uncertainties, the next Tables present the statistical uncertainties (from Table 5.6) for the estimation of the background in the dimuon production. The measurement in the data set run 1 is shown in the column $S+B$.

In the background estimation (B_e) column, the first uncertainty corresponds to the statistical uncertainty, and the second one corresponds to the total systematic uncertainty. The measurements in the data are less than 1 σ from the estimation of the background, which concludes that no dimuon signal was observed in the whole data set run 1.

Background sample	$S+B$	B_e	B_e with total error
FullMC	3	3.8 ± 1.2 ± 1.3	3.8 ± 1.8

Table 5.9: *Summary background estimation with statistical and systematic uncertainties, for $BDTcut_2$ using data set run 1.*

Similarly to Section 5.3.4, an estimation on the upper limit of dimuon detection can be obtained. This value for the $BDTcut_2$ case is:

$$\text{Observed dimuon} = 5.1, \tag{5.9}$$

for the observed dimuon signal, with a 95% confidence level. The efficiency correction for this value (using the efficiency shown in Table 5.2), results:

$$\text{Produced dimuon} = 850, \tag{5.10}$$

which is the estimated production of dimuon signal in the data set run 1. This result allows to obtain the parameter X in Equation 1.1:

$$X = \frac{\sigma(\mu\mu)}{\sigma_{SM}(\mu\mu)} = \frac{\frac{850}{638507}}{\frac{10}{1026120}} = 137. \tag{5.11}$$

The upper limit for the production of dimuon signal that this study estimates is 137 times the one the theoretical framework (GENIE) addresses. If there are BSM channels of dimuon production, these are going to produce a maximum of 137 times the number of dimuons expected by GENIE.

This value differs from the one shown in Section 5.3.4 because the efficiencies are different for both cases, being this last result more precise.

Chapter 6

KS test and sideband control plots

6.1 KS test

One way to check the performance of the BDT classification is to apply the Kolmogorov-Smirnov (KS) test. The test compares the shapes of the testing and training samples (per signal and background separately) and returns a test value. For this analysis, the KS test for the CV case is:

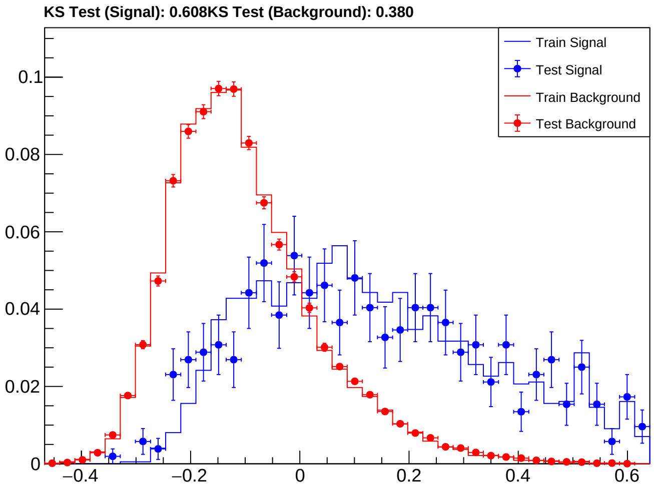


Figure 6.1: *Using 75% of the dimuon signal sample for training and 25% for testing.*

The numerical result shown in Figure 6.1 shows that it is not possible to discard the null hypothesis, in this case, that the distributions are identical. This means that there is enough agreement between the samples, which suggests that there is no over-training.

6.1.1 No signal side band

In section 3.3.5 Figure 3.16 shows that the BDT score region where dimuon signal is expected is around ~-0.3 and ~0.6 (maximum BDT score value), which allows a no-signal side band study. Plotting the behavior of the training and testing variables in the ~-0.44 (minimum BDT score value) and ~0.0 BDT score region would choose events where dimuons are not expected to appear. There must be an agreement between the testing samples, given that the open data set and the fullMC are considered to be scored in the same way by the BDT.

In Figure 6.2, the vertical line indicates the limit of the non-signal sideband to the left of this line to the minimum value of the BDT score in this region.

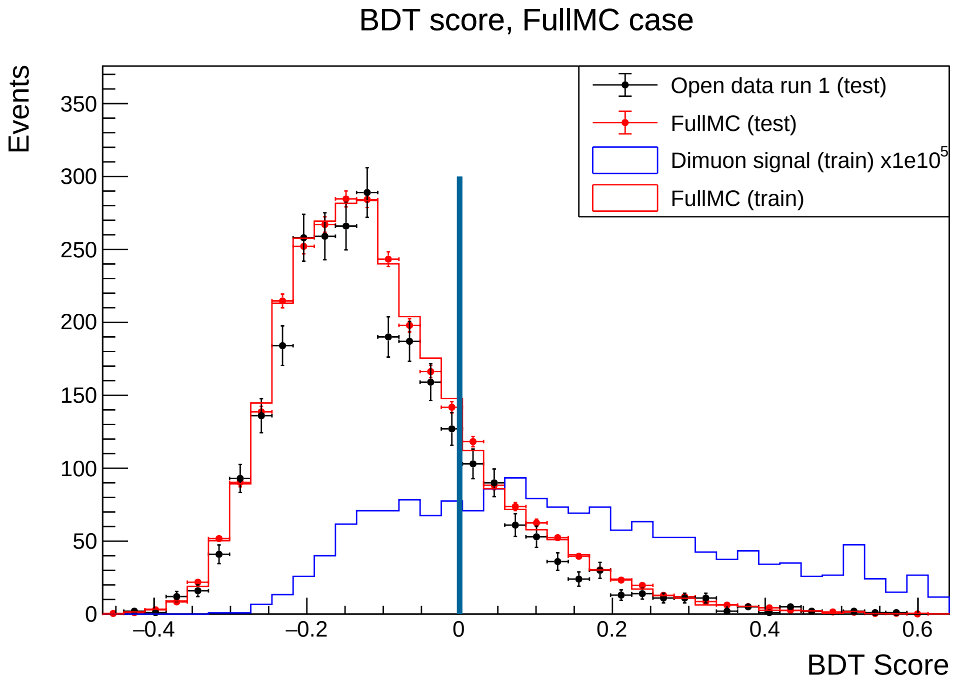


Figure 6.2: *The non-signal region is from the minimum value to the line that corresponds to a 0.0 score.*

Overall, a good agreement is appreciated. The χ^2/ndf test is presented in each comparison. The 7 variables in section 3.3.3 are shown as follows.

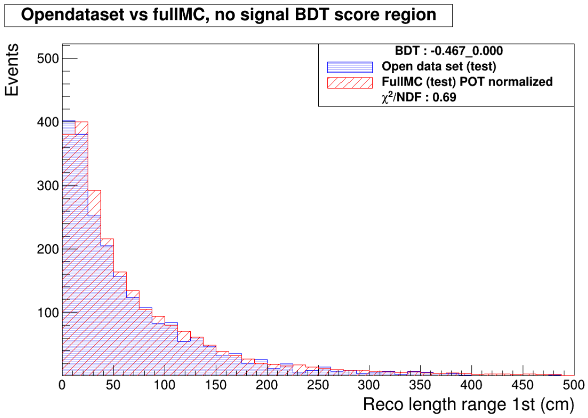


Figure 6.3: *Length first track in non-signal region.*

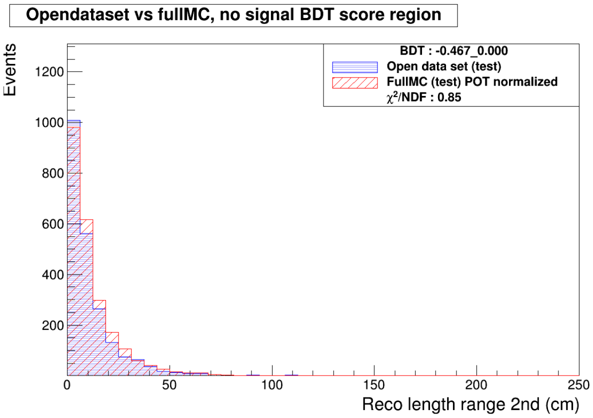


Figure 6.4: *Length second track in non-signal region.*

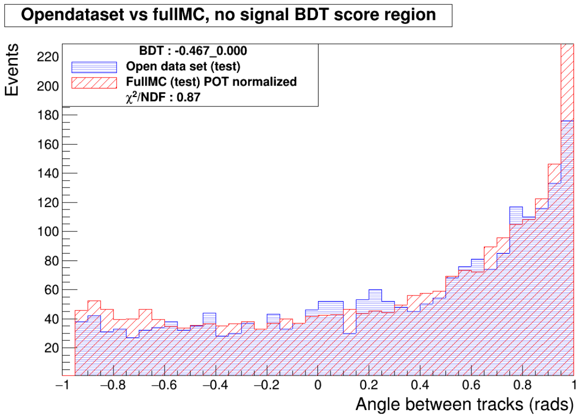


Figure 6.5: *Angle between the two longest tracks in non-signal region.*

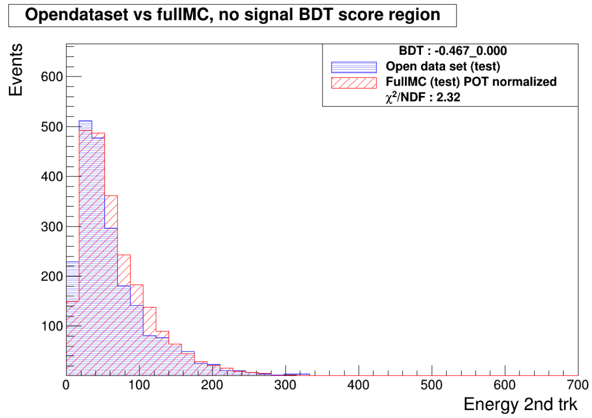


Figure 6.6: *Calorimetry energy for second track in non-signal region.*

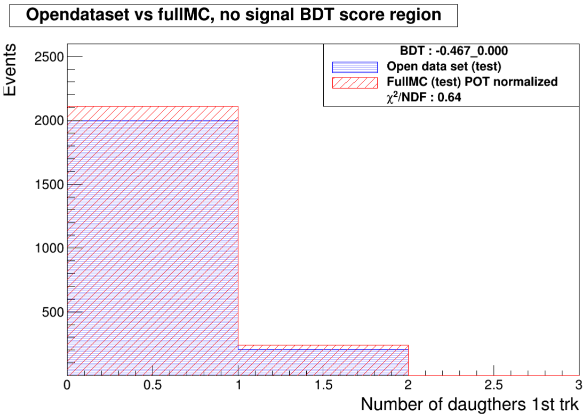


Figure 6.7: *Number of daughters for first track in non-signal region.*

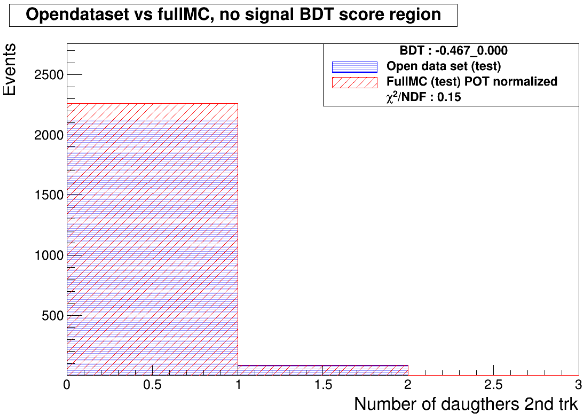


Figure 6.8: *Number of daughters for second track in non-signal region.*

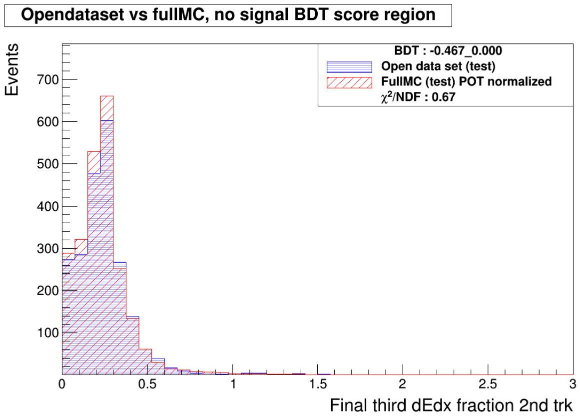


Figure 6.9: *Fraction dE/dX for second track in non signal region.*

6.2 Simulated data or Fake data analysis

Doing a fake data analysis is important because allows a test of the classification performance of the BDT. In order to do that a set of 10 samples were created, from the combination of two initial samples. The first sample is a fraction of events from the fullMC BDT testing sample (events that were excluded from the remaining fullMC for testing) and a completely different dimuon signal sample with 333 dimuon events (it is important for this sample to be different to avoid over-training or coincidences).

Then external support was provided from the K-State high energy physics research group, to guarantee that the author is not biased and does not have knowledge of the amount of dimuon events present in the final fake data samples. The proportions were varied to compare different scenarios.

The events were chosen after the pre-selection (these numbers, called original number of dimuons, are shown in the last columns of Tables 6.1 and 6.2). The seventh column shows the true number of dimuon events in the samples. This value is the original value of dimuons present in the mix divided by the pre-selection efficiency. This calculation must be done given that the comparison is done between estimations based on the signal efficiency. The estimation of the signal from the BDT analysis is shown in the sixth column with its corresponding statistical error. The 10 cases were studied for both BDT cuts used during this study (see Section 3.4.1).

For $BDTcut_1$:

Background sample	$S+B$	B_T	B_e	$(S+B)-B_e$	ϵ	$\frac{(S+B)-B_e}{\epsilon}$	Truth value	Value
Fake data 1	100 ± 10.0	104 ± 10.2	38.5 ± 4.0	61.5 ± 14.0	0.021 ± 0.001	2886 ± 834	2388	301
Fake data 2	105 ± 10.2	104 ± 10.2	38.5 ± 4.0	66.5 ± 14.3	0.021 ± 0.001	3121 ± 860	2595	327
Fake data 3	100 ± 10.0	104 ± 10.2	38.5 ± 4.0	61.5 ± 14.0	0.021 ± 0.001	2886 ± 834	2365	298
Fake data 4	103 ± 10.1	104 ± 10.2	38.5 ± 4.0	64.5 ± 14.2	0.021 ± 0.001	3027 ± 850	2492	314
Fake data 5	96 ± 9.8	104 ± 10.2	38.5 ± 4.0	57.5 ± 13.8	0.021 ± 0.001	2698 ± 813	2214	279
Fake data 6	48 ± 6.9	104 ± 10.2	38.5 ± 4.0	9.5 ± 10.9	0.021 ± 0.001	445 ± 541	0	0
Fake data 7	64 ± 8.0	104 ± 10.2	38.5 ± 4.0	25.5 ± 12.0	0.021 ± 0.001	1196 ± 637	675	85
Fake data 8	72 ± 8.5	104 ± 10.2	38.5 ± 4.0	33.5 ± 12.5	0.021 ± 0.001	1571 ± 683	1254	158
Fake data 9	92 ± 9.6	104 ± 10.2	38.5 ± 4.0	53.5 ± 13.6	0.021 ± 0.001	2510 ± 792	2056	259
Fake data 10	66 ± 8.1	104 ± 10.2	10.3 ± 4.0	55.7 ± 12.1	0.021 ± 0.001	2616 ± 730	2397	302

Table 6.1: *Fake data results for different cases, for* $BDTcut_1$.

For $BDTcut_2$:

Background sample	$S+B$	B_T	B_e	$(S+B)-B_e$	ϵ	$\frac{(S+B)-B_e}{\epsilon}$	Truth value	Value
Fake data 1	17 ± 4.1	7 ± 2.6	2.6 ± 1.0	14.4 ± 5.2	0.006 ± 0.001	2402 ± 1141	2388	301
Fake data 2	19 ± 4.4	7 ± 2.6	2.6 ± 1.0	16.4 ± 5.4	0.006 ± 0.001	2735 ± 1219	2595	327
Fake data 3	17 ± 4.1	7 ± 2.6	2.6 ± 1.0	14.4 ± 5.2	0.006 ± 0.001	2402 ± 1141	2365	298
Fake data 4	18 ± 4.2	7 ± 2.6	2.6 ± 1.0	15.4 ± 5.3	0.006 ± 0.001	2568 ± 1180	2492	314
Fake data 5	17 ± 4.1	7 ± 2.6	2.6 ± 1.0	14.4 ± 5.2	0.006 ± 0.001	2402 ± 1141	2214	279
Fake data 6	2 ± 1.4	7 ± 2.6	2.6 ± 1.0	-0.6 ± 2.5	0.006 ± 0.001	-98 ± 398	0	0
Fake data 7	6 ± 2.4	7 ± 2.6	2.6 ± 1.0	3.4 ± 3.5	0.006 ± 0.001	568 ± 648	675	85
Fake data 8	11 ± 3.3	7 ± 2.6	2.6 ± 1.0	8.4 ± 4.4	0.006 ± 0.001	1402 ± 890	1254	158
Fake data 9	16 ± 4.0	7 ± 2.6	2.6 ± 1.0	13.4 ± 5.0	0.006 ± 0.001	2235 ± 1101	2056	259
Fake data 10	15 ± 3.9	7 ± 2.6	0.7 ± 1.0	14.3 ± 4.9	0.006 ± 0.001	2385 ± 1097	2397	302

Table 6.2: *Fake data results for different cases, for* $BDTcut_2$.

Plotting the BDT estimation with the truth value of the dimuon signal present in the fake data samples, for $BDTcut_1$:

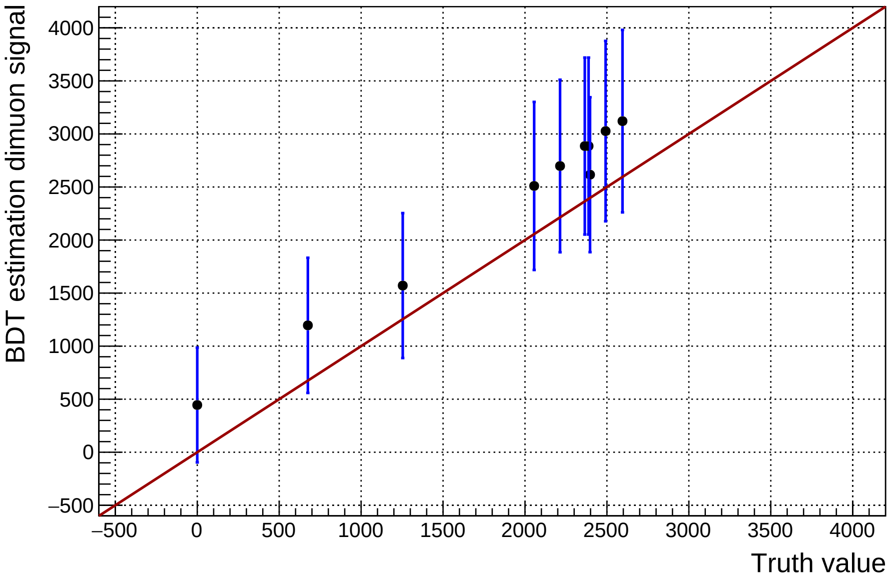


Figure 6.10: *BDT estimation dimuon signal vs. truth value dimuon signal, for* $BDTcut_1$.

The BDT estimation with the truth value of the dimuon signal present in the fake data samples, for $BDTcut_2$:

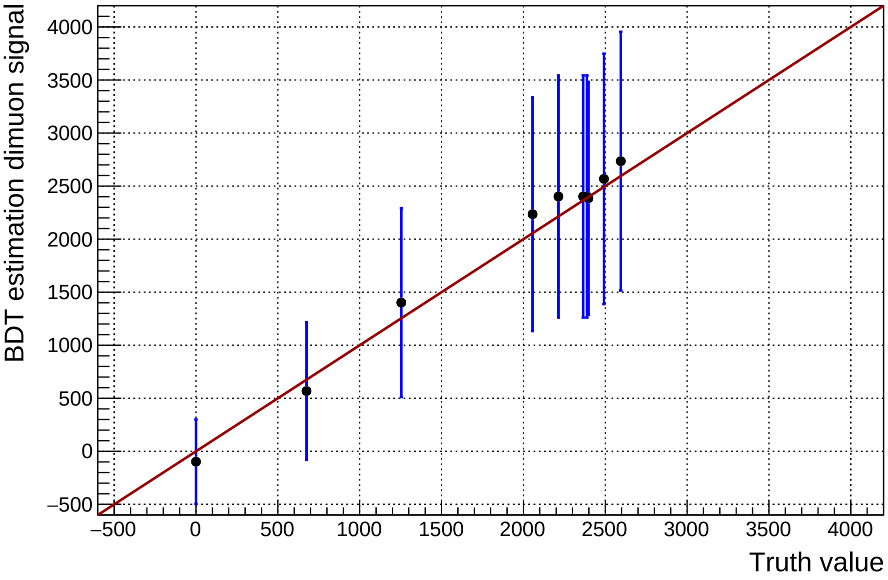


Figure 6.11: *BDT estimation dimuon signal vs. truth value dimuon signal, for* $BDTcut_2$.

From Figures 3.39 and 3.41, it is possible to conclude that the BDT is showing a good performance identifying the estimated dimuon signal. The solid line indicates a slope 1 linear function. The values are in agreement with the line, which again confirms that the estimation concords with the original truth about events included in the fake data samples.

It is interesting to see that increasing the BDT cut (like in the case for $BDTcut_2$) tends to take the point closer to the line while slightly losing precision (longer error bars). This is expected because increasing the cut reduces efficiency.

Chapter 7

Conclusions

Figure 3.16 shows the performance of the BDT trained for this study. Good discrimination during the training stage between the dimuon signal and the background is observed. The BDT score distribution of the neutrino interaction MC shows good modeling and agreement with respect to the open data set (see Figure 3.16) and the data set run 1 (see Figure 5.1).

Section 3.4.1 shows the behavior of the distributions used in this analysis after the BDT score cut. The BDT seems to have successfully learned characteristics of the dimuon signal and selected events with dimuon attributes. Figure 3.19 shows that the estimation of the background in the MC is consistent with the data events selected by the BDT, which indicates that no evidence of excess or deficit of dimuon-like events was observed.

Tables 4.13 and 4.14 for the open data set and Tables 5.5 and 5.9 for the data set run 1, show the background estimation of dimuon signal and the measurement in the data for both BDT score cuts. The measurements in the data are less than 2 σs from the estimation of the background, which concludes again that no dimuon signal was observed in the data samples.

Equations 5.7 and 5.10 present the 95% CL upper limit estimation of the produced dimuon signal in the data set run 1. This measurement allows to obtain how many times (X) maximum it is possible to observe dimuons (see Equations 1.1, 5.8 and 5.11).

Good agreement between data and fullMC in the non-signal region of the BDT score leads to confidence in the simulation. Fake data studies show self-consistency. The KS test performed in Section 6.1 indicates low chance of over-fitting.

Bibliography

[1] R. Acciarri, C. Adams, and et. al. A proposal for a three detector short-baseline neutrino oscillation program in the fermilab booster neutrino beam, 2015. URL `https://arxiv.org/abs/1503.01520`.

[2] B. Abi, R. Acciarri, and et. al. Deep underground neutrino experiment (DUNE), far detector technical design report, volume I: Introduction to DUNE, 2020. URL `https://arxiv.org/abs/2002.02967`.

[3] Costas Andreopoulos, Christopher Barry, Steve Dytman, Hugh Gallagher, Tomasz Golan, Robert Hatcher, Gabriel Perdue, and Julia Yarba. The GENIE neutrino monte carlo generator: Physics and user manual, 2015. URL `https://arxiv.org/abs/1510.05494`.

[4] C. Arguelles, I. Esteban, and et. al. MicroBooNE and the electron neutrino interpretation of the MiniBooNE low energy excess. *Physical Review Letters*, 128 (24), Jun 2022. ISSN 1079-7114. doi: 10.1103/physrevlett.128.241802. URL `http://dx.doi.org/10.1103/PhysRevLett.128.241802`.

[5] R. Acciarri et al. Design and construction of the MicroBooNE Detector. *JINST*, 12 (02):P02017, 2017. doi: 10.1088/1748-0221/12/02/P02017.

[6] C Adams, M Alrashed, and et. al. Rejecting cosmic background for exclusive charged current quasi elastic neutrino interaction studies with liquid argon TPCs; a case study with the MicroBooNE detector. *The European Physical Journal C*, 79, 2019. doi: 10.1140/epjc/s10052-019-7184-7. URL `https://doi.org/10.1140/epjc/s10052-019-7184-7`.

[7] Hanyu Wei. Low-energy excess and new physics searches with MicroBooNE. URL: https://indico.kps.or.kr/event/30/contributions/854/attachments/146/315/Hanyu%20Wei.pdf, May 2022.

[8] Carlo Rubbia. The liquid-argon time projection chamber: a new concept for neutrino detectors. Technical report, CERN, Geneva, 1977. URL https://cds.cern.ch/record/117852.

[9] P. Abratenko, R. An, Anthony, and et al. Novel approach for evaluating detector-related uncertainties in a LArTPC using MicroBooNE data. *The European Physical Journal C*, 82(5), May 2022. ISSN 1434-6052. doi: 10.1140/epjc/s10052-022-10270-8. URL http://dx.doi.org/10.1140/epjc/s10052-022-10270-8.

[10] I. Stancu. Technical design report for the 8 gev beam. doi: 10.2172/1212167. URL https://www.osti.gov/biblio/1212167.

[11] A. A. Aguilar-Arevalo, C. E. Anderson, and et. al. Neutrino flux prediction at MiniBooNE. *Phys. Rev. D*, 79:072002, Apr 2009. doi: 10.1103/PhysRevD.79.072002. URL https://link.aps.org/doi/10.1103/PhysRevD.79.072002.

[12] Mark Thomson. *Modern Particle Physics*. Cambridge University Press, 2013.

[13] G. Aad, T. Abajyan, and et. al. Observation of a new particle in the search for the Standard Model Higgs boson with the ATLAS detector at the LHC. *Physics Letters B*, 716(1):1–29, 2012. ISSN 0370-2693. doi: https://doi.org/10.1016/j.physletb.2012.08.020. URL https://www.sciencedirect.com/science/article/pii/S037026931200857X.

[14] Tom W. B. Kibble. The standard model of particle physics, 2014. URL https://arxiv.org/abs/1412.4094.

[15] C. L. Cowan, F. Reines, F. B. Harrison, H. W. Kruse, and A. D. McGuire. Detection of the Free Neutrino: a confirmation. *Science*, 124(3212):103–104, 1956. doi: 10.1126/

science.124.3212.103. URL https://www.science.org/doi/abs/10.1126/science.124.3212.103.

[16] Markus Bischofberger. *Quasi elastic charm production in neutrino-nucleon scattering.* Doctoral dissertation, SWISS FEDERAL INSTITUTE OF TECHNOLOGY ZURICH, 2005.

[17] The MicroBooNE Collaboration. Neutrino interaction model and uncertainties for MicroBooNE analyses. ISSN MicroBooNE Public Note 1074.

[18] Kevin McFarland. Neutrino Interactions, 2008. URL https://arxiv.org/abs/0804.3899.

[19] Hariom Sogarwal and Prashant Shukla. Coherent pion production in neutrino (anti-neutrino)-nucleus interaction. *Nuclear Physics A*, 1027:122494, November 2022. ISSN 0375-9474. doi: 10.1016/j.nuclphysa.2022.122494. URL http://dx.doi.org/10.1016/j.nuclphysa.2022.122494.

[20] Particle Data Group. Review of Particle Physics. *Progress of Theoretical and Experimental Physics*, 2020(8):083C01, 08 2020. ISSN 2050-3911. doi: 10.1093/ptep/ptaa104. URL https://doi.org/10.1093/ptep/ptaa104.

[21] P. Abratenko, D. Andrade Aldana, and et al. First measurement of quasielastic Λ baryon production in muon antineutrino interactions in the MicroBooNE detector. *Phys. Rev. Lett.*, 130:231802, Jun 2023. doi: 10.1103/PhysRevLett.130.231802. URL https://link.aps.org/doi/10.1103/PhysRevLett.130.231802.

[22] MicroBooNE collaboration. First measurement of η production in neutrino interactions on argon with MicroBooNE. *arXiv*, 2305.16249, 2023.

[23] P. Abratenko, J. Anthony, and et al. Search for long-lived heavy neutral leptons and Higgs portal scalars decaying in the MicroBooNE detector. *Physical Review D*, 106 (9), November 2022. ISSN 2470-0029. doi: 10.1103/physrevd.106.092006. URL http://dx.doi.org/10.1103/PhysRevD.106.092006.

[24] M. Ross-Lonergan, G. Ge, and et al. [single-photon gLEE internal technote] the MicroBooNE single photon LEE search v3.0. ISSN MicroBooNE Document 27768-v3.

[25] R Acciarri, C Adams, and et. al. A study of electron recombination using highly ionizing particles in the ArgoNeuT Liquid Argon TPC. *Journal of Instrumentation*, 8(08): P08005–P08005, August 2013. ISSN 1748-0221. doi: 10.1088/1748-0221/8/08/p08005. URL http://dx.doi.org/10.1088/1748-0221/8/08/P08005.

[26] P. Abratenko, R. An, Anthony, and et al. Calorimetric classification of track-like signatures in liquid argon tpcs using MicroBoone data. *Journal of High Energy Physics*, 2021(12), December 2021. ISSN 1029-8479. doi: 10.1007/jhep12(2021)153. URL http://dx.doi.org/10.1007/JHEP12(2021)153.

[27] MicroBooNE collaboration. Measurement of double-differential cross sections for mesonless charged-current muon neutrino interactions on argon with final-state protons using the MicroBooNE detector. *arXiv*, 2403.19574, 2024.

[28] A. Hoecker, P. Speckmayer, and et. al. TMVA - Toolkit for multivariate data analysis, 2009. URL https://arxiv.org/abs/physics/0703039.

[29] S. Zeller and A. Schukraff. Final state nucleons for neutrino-nucleus interactions workshop at jlab, 2015. URL https://jlab.org/conferences/funfact/talks/thursday/zeller.pdf.

[30] Wouter Van De Pontseele. *SEARCH FOR ELECTRON NEUTRINO ANOMALIES WITH THE MICROBOONE DETECTOR.* Doctoral dissertation, University of Oxford, 2020. URL https://lss.fnal.gov/archive/thesis/2000/fermilab-thesis-2020-11.pdf.

[31] Rene Brun, Fons Rademakers, and et. al. root-project/root: v6.18/02, June 2020. URL https://doi.org/10.5281/zenodo.3895860.

[32] E. L. Snider and G. Petrillo. LArSoft: Toolkit for Simulation, Reconstruction and

Analysis of Liquid Argon TPC Neutrino Detectors. *J. Phys. Conf. Ser.*, 898(4):042057, 2017. doi: 10.1088/1742-6596/898/4/042057.

Appendix A

MC and dimuon sample production

A.1 Setting up uboonecode

The Liquid Argon Software (LArSoft) collaboration develops a large amount of physics software for all kind of Liquid Argon Time Projection Chamber (LArTPC) experiments. For this study LArSoft was mainly used for the following stages of MC simulation:

- GENIE, generation of mass, charge, momentum.
- Geant4, simulate the conditions and geometry of the detector.
- Detsim, implements the simulation and the design of the detector.
- Pandora, algorithm of reconstruction.

The first step is to set up the μB server with the uboonecode frame desired, for this study the version of uboonecode (see [32]) used to produce the dimuon sample was:

v08_00_00_69 -q e17:prof

Which corresponds to the LArSoft version:

v08_05_00_20

The Art version:

v3_01_02

And the GENIE version:

v3_00_04_ub4

Initially all the samples are correspondent to the run 1, which in the case of the dimuon sample means that the overlay file chosen is correspondent to run 1, section A.4 shows exactly the name of the SAM definition file.

A.2 Dimuon signal MC sample

A sample with events of dimuons was produced from scratch. This sample was thought to be the signal sample for the training stage of a BDT (see Section 3.3). A GENIE filter was used for the generation stage. The main lines in the code are:

- physics.filters.finalstatefilter.PDG: [13,-13]
- physics.filters.finalstatefilter.PDGCount: [1,1]
- physics.filters.finalstatefilter.PDGCountExclusivity: [true,true]

The first item indicates the particle data group [20] code associated to muons. It requires the event to have two muons with opposite sign, the second item requires to have at least one of each and the third item says that is a exclusive signal, so it is possible to have more than one of each particle, in the fullMC sample there were two evens with 3 muons, $2\mu^-$ and $1\mu^+$ coming from a η production.

These lines besides other standard parameters are included in a source file which will be submitted to the Fermilab grid to ensure a fast production, given that produce just 10 events in the server took 8 hours (test done as a first approach initially) after filtering ~1 million events.

A.3 Filter submitted to FermiGrid

The FermiGrid client version used was:
larbatch v01_51_13 with the option: **jobsub_client v_lite**.
To submit the GENIE stage a .xml file is needed. It will include the workflow needed to get the final rootfile with the ntuple with the variables that are going to be studied.

A.4 Overlay workflow

During the trigger, there could be many cosmic rays that happen to overlay the events in the real data sets. Then, it is needed to include that information in the MC samples. In this case, the overlay was done to the dimuon MC sample. Again a .fcl has to do the connection in between the generation stage and the rest of the stages in the grid, in this case it was used:

standard_overlay_gen_SimInfoMixer_prod.fcl.

This .fcl is key because it mediates between the GENIE filter and the rest of the workflow that will produce the overlay.
And for the overlay an input SAM definition should be chosen, in this case as a recommendation from the gLEE group the samdef used was:

prod_extunbiased_swizzle_inclusive_v3_goodruns_mcc9_run1_high_lifetime_overlay_nc_pi0.

This contains the cosmic rays that will be overlaid to the dimuon events. More precisely, this file contains 2712 Pandora reco2 files. This value is important because is going to be the maximum amount of jobs that were submitted to the grid simultaneously.

A.4.1 Geant4

Geant4 is a toolkit that simulates the interaction of particles with matter among a wide variety of other applications. It is the first stage in the workflow and is going to include in the sample all the interactions that happen between the GENIE simulated particles with the detector while these propagate through it. This task is in charge of the .fcl:

wirecell_g4_uboone.fcl.

A.4.2 Detector simulation

The detsim stage simulates the geometry of the detector and uses the Geant4 information to simulate how the detector is going to respond to the neutrino interaction that is happening in its volume. This is the .fcl control file:

wirecell_detsim_overlay_uboone.fcl.

A.4.3 Overlay

The next stages are related to the overlay processes to be added to the interactions. Doing these steps is important because they allow a later comparison between this MC and actual real data. It is an usual workflow for the overlay of MC samples. It was provided by the production group:

standard_overlay_uboone.fcl,
reco_uboone_mcc9_8_driver_overlay_stage1a.fcl,
reco_uboone_mcc9_8_driver_overlay_stage1b.fcl,
standard_larcv_uboone_mctruth_prod.fcl,
reco_uboone_mcc9_8_driver_overlay_stage1c.fcl,
wirecell_detsim_optical_overlay_uboone.fcl,
standard_overlay_optical_uboone.fcl,

reco_uboone_mcc9_8_driver_overlay_optical.fcl,

standard_larcv_uboone_mc2d_optical_prod.fcl,
reco_uboone_mcc9_8_driver_overlay_stage2.fcl.

A.4.4 Ntuple creation

The last step is to create the ntuple which is going to have all the variables that are going to be studied in this analysis.

The module used is the SinglePhoton analyzer module, which was created by the gLEE group in μB collaboration, is the last step in the workflow submitted to the grid:
run_SinglePhoton_Overlay.fcl

A.4.5 Workflow for re-weighting the CV sample

The SAM definition used to create the fullMC sample is:

prodgenie_bnb_nu_uboone_overlay_mcc9.1_v08_00_00_26_filter_run1_reco2_reco2,
and it was used as input definition for the workflow to get the re-weighted sample:

run_eventweight_microboone_sep24.fcl,
run_eventweight_microboone_sep24_extragenieall_1.fcl,
run_eventweight_microboone_sep24_extragenieall_2.fcl,
run_eventweight_microboone_sep24_extragenieall_3.fcl,
run_eventweight_microboone_sep24_extragenieall_4.fcl,
run_eventweight_microboone_sep24_extragenieall_5.fcl,
run_SinglePhoton_Overlay.fcl.

This will save the tree “ true_eventweight_tree” with the map “mcweight” branch saved on it. Then the new fullMC sample now contains the vectors with 1000 entries corresponding each universe for each unisim. The fullMC new was created with the version of uboonecode v08_00_00_70 and has size of 97Gb.

Appendix B

Tools

B.1 Root

Root is a scientific framework widely used in physics to develop applications for analysis of large scale amounts of data. The toolkit has more than 1.5 million of lines of code and has standout characteristics like the possibility for the user to create their own classes and work the structure of data desired [31].

The version used for this study is ROOT 6.26/04 in the author's personal laptop and mainly version 6.12/06 in μB server.

Root was widely used for this study including MC truth analysis, establishing fiducial conditions and pre-selection for the reconstructed data, BDT training and testing, preparation of all the plots shown in this document, systematic uncertainties and simulated/fake data studies.