

Study of NuMI Dirt Monte Carlo Events in the ICARUS T600 Detector

Fermilab Internship Final Report

Marco Caporale

supervisor Doctor Minerba Betancourt

special thanks to Alice Campani and Bruce Howard

October 2023

Abstract

The moving of the Icarus detector from Laboratori Nazionali del Gran Sasso to Fermilab in order to pursue the physics goals of the Short-Baseline Neutrino (SBN) program has introduced the necessity of a more careful study of backgrounds. In this work some developments of the validation of the Monte Carlo backgrounds for the Icarus run 1 data taking are presented, focusing on the study of the so called “Dirt” events from the NuMI beamline. In particular the analysis hints at a possible underestimation of the muon background. Hopefully this issue will be solved at a later stage of the analysis with the inclusion of components that were not considered in the present work.

Contents

1	Introduction	3
2	The SBN Program	4
2.1	ICARUS T600 Detector	4
3	Rock Muon Analysis	5
3.1	Dataset and Applied Cuts	5
3.1.1	Vertex Position	7
3.1.2	Track End Position	8
3.1.3	Entering Direction of the Track	8
3.1.4	Track Length	8
3.1.5	PMT Flash Time	9
3.2	Summary of Selection Cuts	9
3.3	Inclusion of Systematics	11
3.4	Investigated Variables	11
3.4.1	Track Vertex Position	11
3.4.2	Energy Loss vs Residual Range	12
3.4.3	Track Length	12
3.4.4	Angular Distributions	12
4	Conclusions	20
	References	21
A	Setup and Macro Example	22
A.1	Login	22
A.2	Environment Setup	22
A.3	Macro Example	23
A.3.1	Macro Execution and ROOT	24
A.3.2	Systematics Effects	24
B	CAFana Cuts	25
B.1	Vertex Position	25
B.2	Track End Position	26
B.3	Track Y Direction	26
B.4	Track Length	26
B.5	PMT Flash Times	27
C	List of Systematics	27
C.1	GENIE MultiSigma	27
C.2	GENIE Multisim	27
C.3	GEANT4	28
D	Useful Paths	28
D.1	Code	28
D.2	Datasets	28

1 Introduction

The joint analysis of the data collected at LEP points towards a constraint of only 3 types of light flavoured neutrinos [4]. While on one hand the introduction of neutrino oscillation showed by the SNO [2] and Super-kamiokande [3] experiments helped solving the main neutrino anomalies present in the Homestake and Gallex experiments, on the other hand some anomalies remain as of today still unsolved [1]. One possible solutions to these anomalies could be the existence of a new particle, namely an unflavoured neutrino.

Such neutrino would extend the Pontecorvo mixing from (1),

$$\begin{pmatrix} \nu_e \\ \nu_\mu \\ \nu_\tau \end{pmatrix} = V_{PMNS} \begin{pmatrix} \nu_1 \\ \nu_2 \\ \nu_3 \end{pmatrix} = \begin{pmatrix} U_{e1} & U_{e2} & U_{e3} \\ U_{\mu1} & U_{\mu2} & U_{\mu3} \\ U_{\tau1} & U_{\tau2} & U_{\tau3} \end{pmatrix} \begin{pmatrix} \nu_1 \\ \nu_2 \\ \nu_3 \end{pmatrix}, \quad (1)$$

to (2),

$$\begin{aligned} \begin{pmatrix} \nu_e \\ \nu_\mu \\ \nu_\tau \\ \nu_s \end{pmatrix} &= V'_{PMNS} \begin{pmatrix} \nu_1 \\ \nu_2 \\ \nu_3 \\ \nu_4 \end{pmatrix} = \\ &= \begin{pmatrix} U_{e1} & U_{e2} & U_{e3} & U_{e4} \\ U_{\mu1} & U_{\mu2} & U_{\mu3} & U_{\mu4} \\ U_{\tau1} & U_{\tau2} & U_{\tau3} & U_{\tau4} \\ U_{s1} & U_{s2} & U_{s3} & U_{s4} \end{pmatrix} \begin{pmatrix} \nu_1 \\ \nu_2 \\ \nu_3 \\ \nu_4 \end{pmatrix}, \end{aligned} \quad (2)$$

where $\nu_1, \nu_2, \nu_3, \nu_4$ are mass eigenstates, $\nu_e, \nu_\mu, \nu_\tau, \nu_s$ are respectively the 3 flavoured and the 1 sterile neutrino eigenstates and V is the Pontecorvo-Maki-Nakagawa-Sakata matrix. This model is also known as the 3+1 neutrino model. While flavoured neutrinos are not allowed to interact strongly or electromagnetically, sterile neutrinos are also not allowed to interact weakly. For this reason their existence could only be probed via oscillation experiments observing the indirect effect on the behaviour of flavoured neutrinos.

In the spontaneous symmetry breaking formalism the most general lagrangian mass term for the lepton sector we may write is (3). Note that $\tilde{H} = i\sigma^2 H^*$, where H is the Higgs doublet.

$$\mathcal{L}_{mass} = -Y_{ij}^e \bar{L}^i H e_R^j - Y_{ij}^\nu \bar{L}^i \tilde{H} \nu_R^j - iM_{ij} (\nu_R^T \sigma^2)^i \nu_R^j + h.c. \quad (3)$$

In particular leading to the seesaw lagrangian for the ν mass terms of the form shown in (4).

$$\mathcal{L}_\nu = -Y_{ij} \bar{L}_i \tilde{H} \nu_j - \frac{M_{ij}}{2} \bar{N}_i^c N_j + h.c. \quad (4)$$

Upon spontaneous symmetry breaking we can retrieve the neutrino mass matrix for the light neutrinos, which receives most of its contribution from flavourful neutrinos. The sterile components are instead mostly relegated to higher masses. It is nonetheless possible to move to a minimal extension of this model which allows for sterile eV masses as shown in [8].

The current research focus, driven by the hints of the current observed anomalies, is in fact aimed at a possible sterile neutrino such that m_4 is in the eV mass-range, resulting in the condition (5).

$$|m_{41}|^2 \gg |m_{31}|^2, |m_{21}|^2 \quad (5)$$

As long as (5) holds the oscillation probability from one flavour α to another one β may be approximated to (6).

$$P_{\alpha\beta} \simeq \delta_{\alpha\beta} - 4|U_{\alpha\beta}|^2 (\delta_{\alpha\beta} - |U_{\alpha\beta}|^2) \sin^2\left(\frac{\Delta m_{41}^2 L}{4E}\right) \quad (6)$$

Under these hypotheses it would be possible to restrict ourselves to the observation of the P_{ee} , $P_{\mu\mu}$, $P_{\mu e}$ channels, which could lead to the observation of the sterile neutrino presence. In particular we get

$$P_{ee} \simeq 1 - 4(1 - |U_{e4}|^2)|U_{e4}|^2 \sin^2\left(\frac{1.27\Delta m_{41}^2 L}{E}\right), \quad (7)$$

$$P_{\mu\mu} \simeq 1 - 4(1 - |U_{\mu4}|^2)|U_{\mu4}|^2 \sin^2\left(\frac{1.27\Delta m_{41}^2 L}{E}\right), \quad (8)$$

$$P_{\mu e} \simeq 4|U_{e4}|^2|U_{\mu4}|^2 \sin^2\left(\frac{1.27\Delta m_{41}^2 L}{E}\right). \quad (9)$$

The final factor

$$\frac{1.27\Delta m_{41}^2 L}{E} \quad (10)$$

is crucial because such dependence is not present in other models which try to explain neutrino beamline anomalies without resorting to the introduction of sterile neutrinos. The consistency of experimental data with such dependence may provide useful hints about the existence of sterile neutrinos.

2 The SBN Program

Despite numerous experimental efforts to solve the neutrino anomalies a conclusive evidence supporting one theory or another is still not available.

The goal of the Short Baseline Neutrino program (SBN) is to observe neutrino oscillations with a series of three Liquid Argon Time Projection Chamber (LArTPC) detectors. In particular the three detectors, shown in Figure 1, are situated at ≈ 100 m (SBND), 500 m (MicroBooNE) and 600 m (ICARUS) from the Booster Neutrino Beamline (BNB), which produces neutrinos thanks to a 8 GeV proton beam impinging on a target.

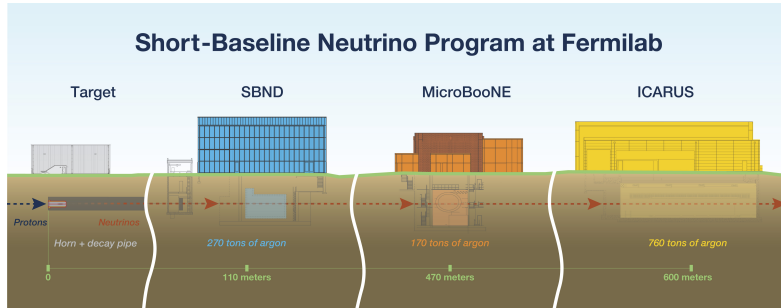


Figure 1: *Schematics of the three SBN experiments with respect to the BNB. From left to right, the BNB target, SBND, MicroBooNE and ICARUS.*

The ICARUS detector also sees 5° off-axis neutrinos produced in the NuMI beamline, which produces higher energy neutrinos compared to those coming from BNB thanks to a 120 GeV proton beam.

2.1 ICARUS T600 Detector

The ICARUS T600 detector is a LArTPC with a total argon mass of approximately 600 tons. Its purpose is the reconstruction at millimeter precision of tracks caused by charged particles passing

through the detector, which is composed of a large cryostat containing 2 chambers filled with ultra-pure liquid argon. The two modules are approximately $3.6*3.9*19.6\text{ m}^3$ in dimension. The passage of charged particles causes a ionisation of the Argon. In order to observe the interaction an electric field induces a drift of the electrons which migrate, repelled by the central cathodes, towards the walls of the two chambers where they reach wire planes which allow for event readout. A pictorial representation of the original detector configuration can be seen in Figure 2.

After the end of the operations at LNGS the detector has been transported to CERN for

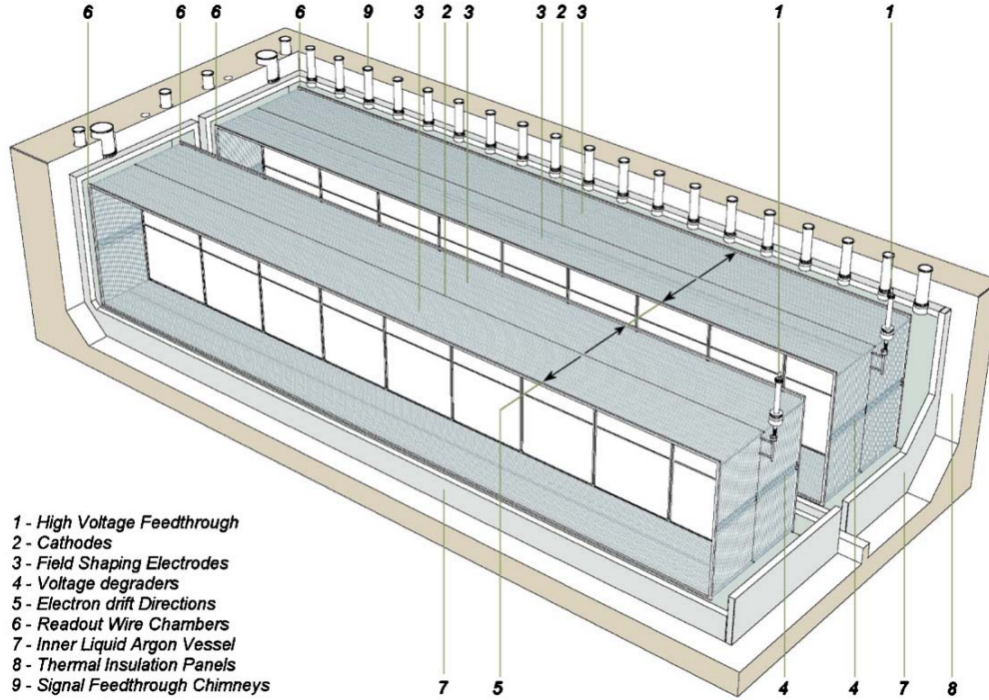


Figure 2: Representation of the ICARUS detector at the Laboratori Nazionali del Gran Sasso [7].

refurbishment and then it has been installed at Fermilab to become part of the SBN program. At LNGS the ICARUS detector was shielded against cosmic radiation by $\approx 3400\text{ m}$ of water equivalent shielding. Such protection is not available at Fermilab and the ICARUS collaboration had to develop a way to cope with the increase of cosmic radiation. In particular, the Cosmic Ray Tagging (CRT) system has been added with respect to what is shown in Figure 2. The CRT covers all the sides of the detector, providing information about muons (or more rarely other charged particles) entering the detector. Pictures of the Top and the Side CRT can be seen in Figure 3.

3 Rock Muon Analysis

3.1 Dataset and Applied Cuts

The aim of this work is the study of backgrounds for the ICARUS run 1 data taking. In particular at the ICARUS detector several background contributions are to be expected, namely

- muons from secondary cosmic showers,



(a) Top CRT.

(b) Side CRT.

Figure 3: (3a) Picture of the Top CRT in 2022 before the concrete overburden installation [5] and (3b) picture of the Side CRT taken during my internship (2023).

- beam neutrinos interacting before reaching the detector leading to a muon entering the argon volume,
- non-beam neutrinos crossing the Earth and interacting leading to a muon crossing the detector. Such neutrinos are also called *dirt neutrinos*.

A pictorial representation of such backgrounds is shown in Figure 4.

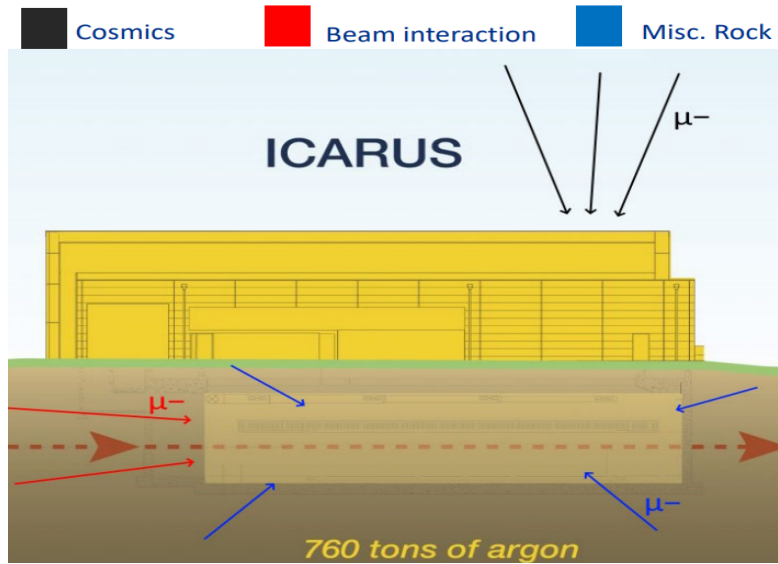


Figure 4: Expected types of background events.

Since in the original Monte Carlo for the ICARUS run1 analysis the contribution of dirt neutrinos was not present it is crucial to factor it in before proceeding further with the analyses. For this comparison three datasets have been created:

- **All MC**, a Monte Carlo containing cosmics and beam interaction muons

- **Rock MC**, a Monte Carlo containing cosmics, beam interaction muons and the dirt contribution
- **Data**, data collected during the ICARUS run 1

In order to focus the attention on the relevant dirt contribution it is necessary to implement a selection that could get rid of most of the background. The importance of the selection becomes clearer by looking at Figure 5, which shows how few the neutrino induced events are with respect to the total number of events in the Monte Carlo samples.

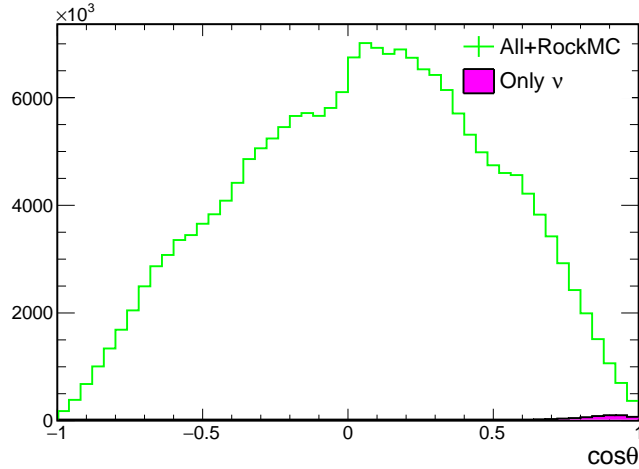


Figure 5: Total number of events compared to the number of neutrino induced events in the Monte Carlo samples. It is clear how a careful selection of the cuts is crucial in order to perform the analysis.

Some cuts have therefore been introduced in order to optimise the selection of events relevant for the analysis. The introduced cuts, thanks to the CAFana framework, may act either at slice or spill level.

At slice level the constraints act on:

- position of the track vertex,
- position of the track end,
- entering direction of the track,
- track length.

The main cut acting at spill level is:

- PMT flash time constraint

3.1.1 Vertex Position

The first requirement introduced acts on the registered vertex position of the track. If we are only looking at events that happen outside of the detector we need the tracks to start outside of the fiducial volume.

The fiducial volume constraint is defined at a 25 cm distance from the TPC boundaries for the x and y axis, while for the z axis the region is chosen as 30 cm from the front and 50 cm from the back of the detector. The constraint is introduced in CAFana code as shown in Section B.1. A representation of the region defined by this requirement is shown in Figure 6.

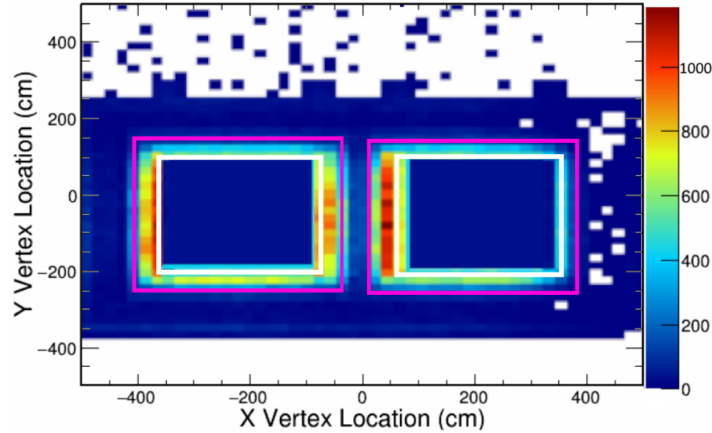


Figure 6: *Track start of a subset of the Rock MC sample with respect to the detector. The pink region highlights the borders of the two TPCs while the white region highlights the fiducial volume.*

3.1.2 Track End Position

Tracks exiting from the detector lead to lack of information about the total energy of the exiting muon. For this reason a full containment of the track is required. The implementation is similar as the one shown in Section 3.1.1, where the accepted borders are loosened to 10 cm. The CAFana implementation is shown in Section B.2.

3.1.3 Entering Direction of the Track

Most of the cosmic background is expected to reach the detector from the sky, as most of such muons are produced as part of secondary cosmic ray showers. One of the most effective reduction of cosmic backgrounds can be operated by cutting with respect to the arriving direction of a track. The effects of the CAFana cut, described in Section B.3, are show in Figure 7.

3.1.4 Track Length

After applying all the previous cuts it can be noticed that a significant portion of non neutrino events still survives the selection. When comparing the track length of the remaining tracks it can be noticed that most of the cosmics concentrate at lower lengths, while neutrino tracks easily reach bigger values. For this reason a constraint on longest track length was introduced during a previous analysis cutting all tracks below 15 cm. The constraint has been tightened during this analysis to 20 cm. Also an even more severe 100 cm boundary has been tried but no significant effects in the results has been observed. A distribution of the longest track length for the Rock MC can be seen in Figure 9. As it can be seen in Section B.4, this cut is implemented similarly to the Track Y Direction cut.

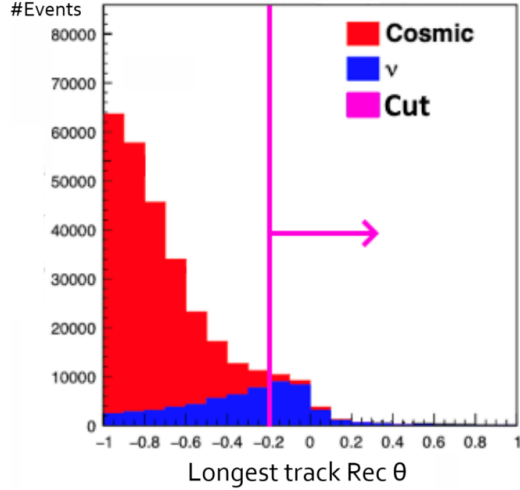


Figure 7: Arriving direction of the tracks for the combined Monte Carlo sample. Note how this cut is very effective in cutting unwanted cosmic background.

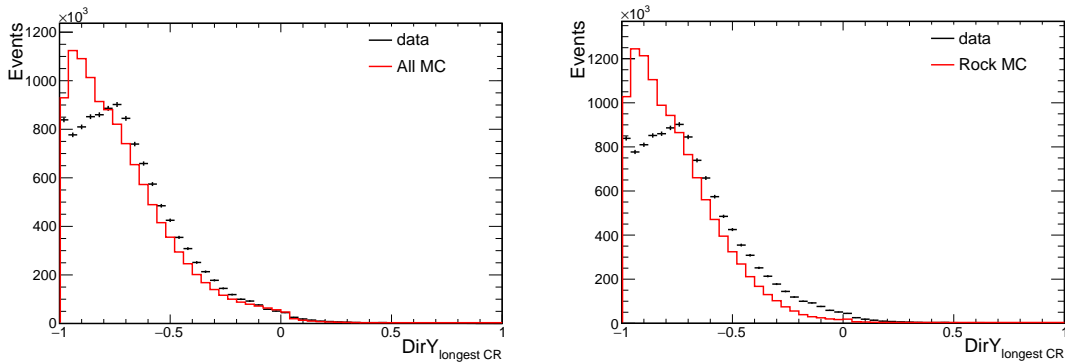


Figure 8: Monte Carlo versus data comparison of the arriving direction of non-fiducial vertex tracks. Histograms are area normalised.

3.1.5 PMT Flash Time

On the other hand at spill level it is sensible to consider that beam neutrinos are expected to reach the detector in time with the NuMI beam. This translates to an observed excess of events between $-0.2\mu\text{s}$ and $9.8\mu\text{s}$ with respect to the spill time, as it can be seen in Figure 10. This constraint is further tightened while implementing the cut leading to an accepted time range spanning between -0.1 and $9.7\mu\text{s}$. The precise definition of the cut is shown in Section B.5.

3.2 Summary of Selection Cuts

Finally the full list of constraints acting on the datasets are given:

- vertices outside of the fiducial volume,
- full containment of track ends (10 cm or more from detector borders),

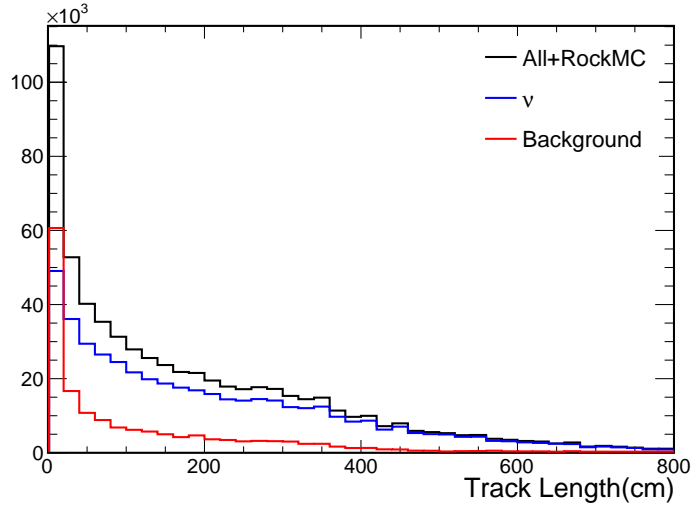


Figure 9: Monte Carlo longest track length after applying entering direction and flash time constraint (see Section 3.1.5). Clearly many cosmics survive the selection below 20 cm.

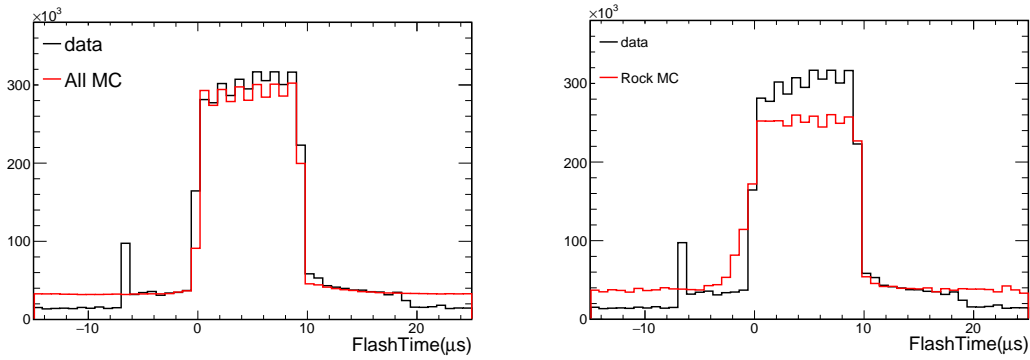


Figure 10: Area normalised histograms of the Monte Carlo Flash times vs data. No cuts are applied to the datasets.

- direction along the vertical axis (DirY) above the -0.2 mark,
- track lengths above 20 cm,
- tracks in time (-0.1→9.7 μs) with respect to the beam spill.

In total the tracks surviving the selection (accounting for a Proton on Target normalisation of $6 * 10^{20}$ protons) are

- 121344 for the Data dataset,
- 54471 for the Rock Monte Carlo dataset,
- 55962 for the All Monte Carlo dataset.

3.3 Inclusion of Systematics

One of the possible issues affecting the mismatch between the the Monte Carlo and the data collected during the ICARUS run1 data taking may be attributed to the simplistic model used to generate the Monte Carlo. In order to overcome this issue the Monte Carlo samples have been regenerated accounting for some of the systematic effects. The major contributions currently under study can be divided in three categories:

- Flux contributions,
- Genie contributions,
- GEANT4 contributions.

For what concerns the GEANT4 contributions they are somewhat at a earlier stage of development and therefore have not yet been fully implemented in this analysis. The flux contributions instead originate from different modellings of the neutrino beam flux. Finally the GENIE contributions use the GENIE software [6] to account for many different neutrino interaction models.

All the systematics get introduced with respect to a weighted non-shifted reference spectrum. The departure of each systematic contribution is then added in quadrature finally leading to an estimate of the total effect of the systematic correction.

This procedure is an oversimplification of the real effects acting on the spectra. A future analysis may address the issue of establishing more precise weights for each correction and possibly looking into correlations between different systematics. Obviously such modification will lead to different contributions compared to those obtained by this analysis, which is mainly aiming at a rough estimate of these effects.

The list of the systematics included in the two Monte Carlo samples is available at Section C. As it will be seen in the plots in the next pages the most dominant contributions are generally coming from GENIE components.

3.4 Investigated Variables

In order to ensure the experimental data aligns with the two Monte Carlo samples, it is essential to compare the distribution of selected variables. In particular in the present work the analysed variables are

- Track vertex position
- Energy loss against residual range
- Track length
- Angular distributions ($\cos \theta_{\text{NuMI}}$, ϕ , $\cos \theta$)

3.4.1 Track Vertex Position

One of the first sanity checks to perform is the comparison of the vertices of the interactions observed in the detector. Despite the low statistics of the Rock Monte Carlo and the experimental a somewhat compatible pattern between the position of the vertices can be seen in Figures 11, 12, 13. In any case a more careful study of the geometry of the detector may be in order in the future.

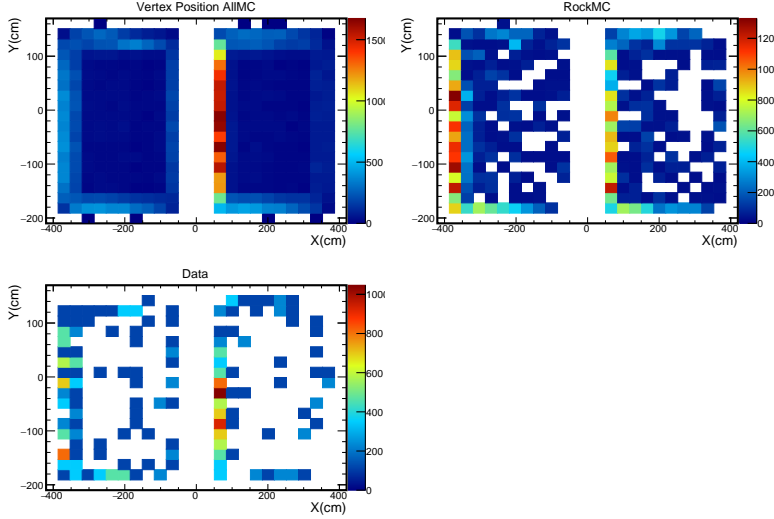


Figure 11: XY plane projection of the vertex position of the tracks surviving the cuts.

3.4.2 Energy Loss vs Residual Range

Another sanity check comes from the comparison of the energy loss and the residual range for the datasets. Unfortunately, due to file storage constraints, only the last 30 cm of the tracks are saved in data files not allowing to fully explore the tails of the histograms past the 30 cm residual range mark. The plots can be seen in Figure 14, while the profiles can be seen in Figure 15.

3.4.3 Track Length

Finally in Figure 16 the distribution of the track length is shown. The relative error components are shown in Figure 17.

3.4.4 Angular Distributions

It is also convenient to look into angular distributions of the tracks. In particular we use ϕ , $\cos\theta$ to parametrise the detector in spherical coordinates, where $\cos\theta$ is defined with respect to the z axis ($\cos\theta = \frac{\text{Track Direction} \cdot \vec{z}}{(|\text{Track Direction}| |\vec{z}|)}$). Also, since the analysis concerns NuMI tracks another investigated variable was $\cos\theta_{\text{NuMI}}$, which has a similar definition to $\cos\theta$ where the \vec{z} direction is substituted by the NuMI direction. Plots are shown in Figures 18→24. In Figure 20 some of the most relevant mismatches can be seen. Such mismatches are probably due to some detector effects that will be included in the analysis in the future.

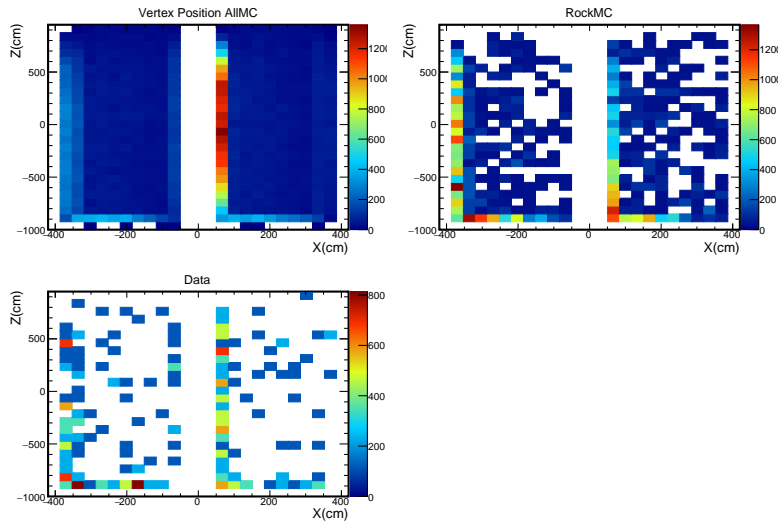


Figure 12: XZ plane projection of the vertex position of the tracks surviving the cuts.

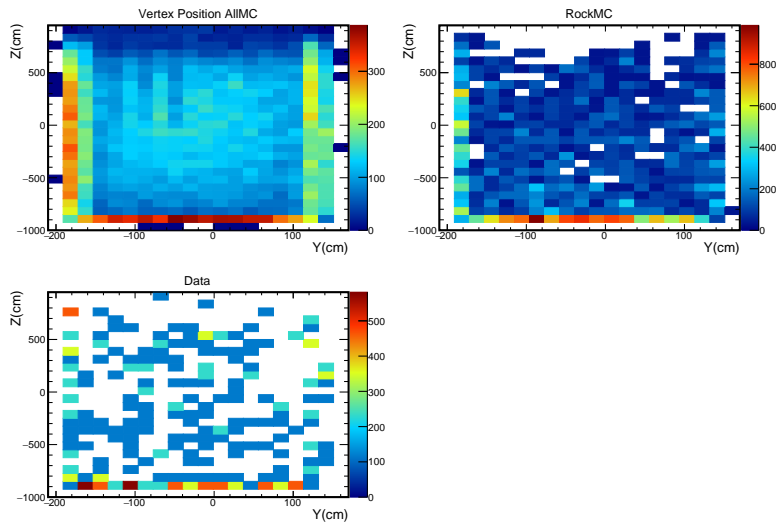


Figure 13: YZ plane projection of the vertex position of the tracks surviving the cuts.

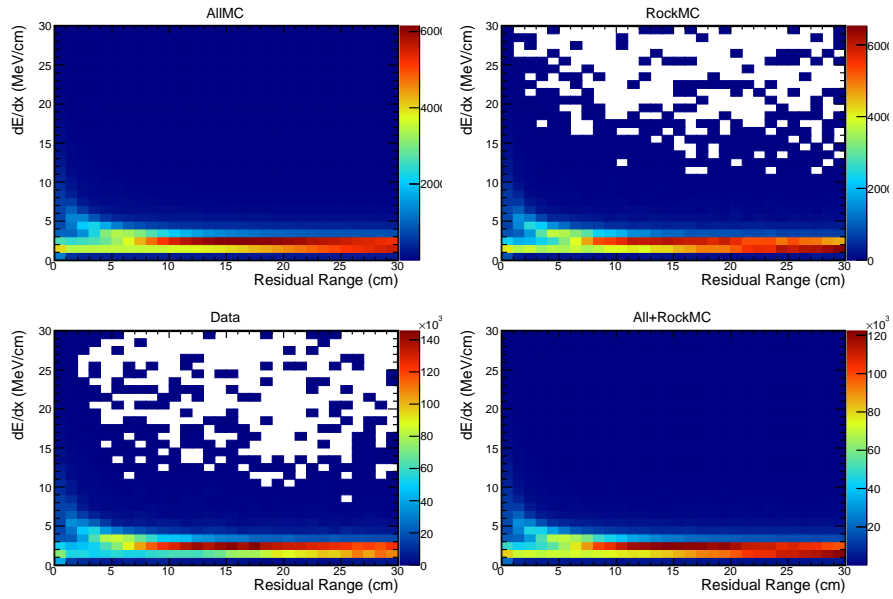


Figure 14: *Residual range of the tracks plotted against the energy loss for events surviving the selection.*

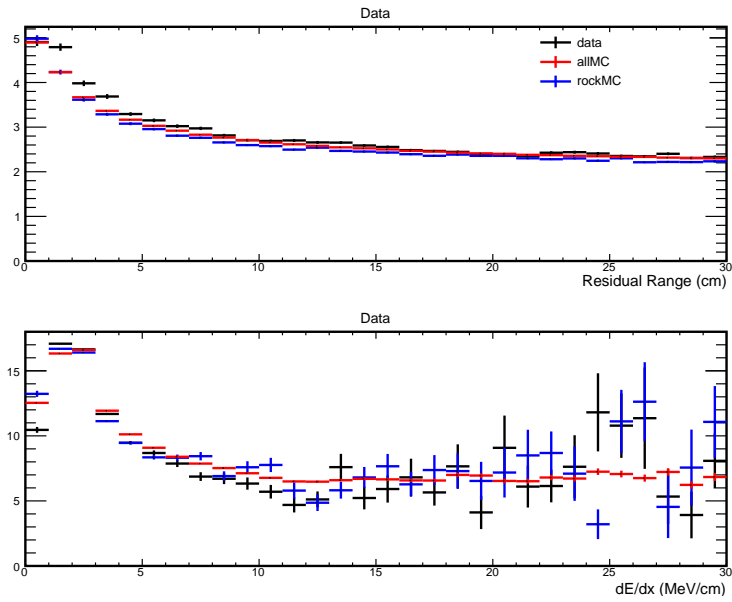


Figure 15: *Profiles for the histograms shown in Figure 14.*

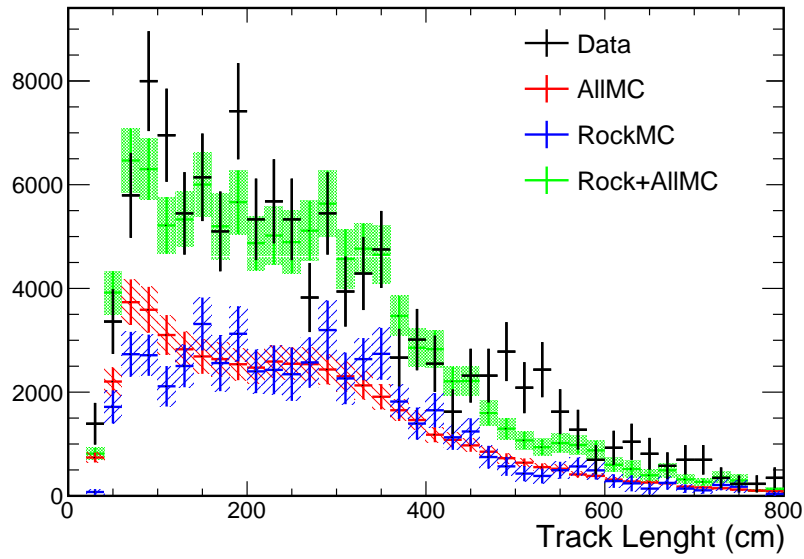


Figure 16: *Length of the tracks.*

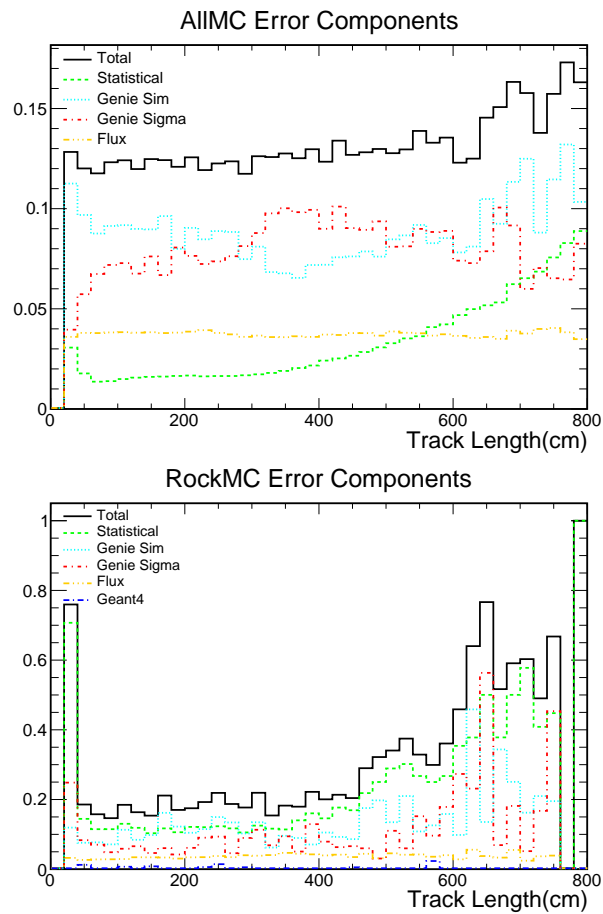


Figure 17: *Relative error components on the track length for the Monte Carlo samples.*

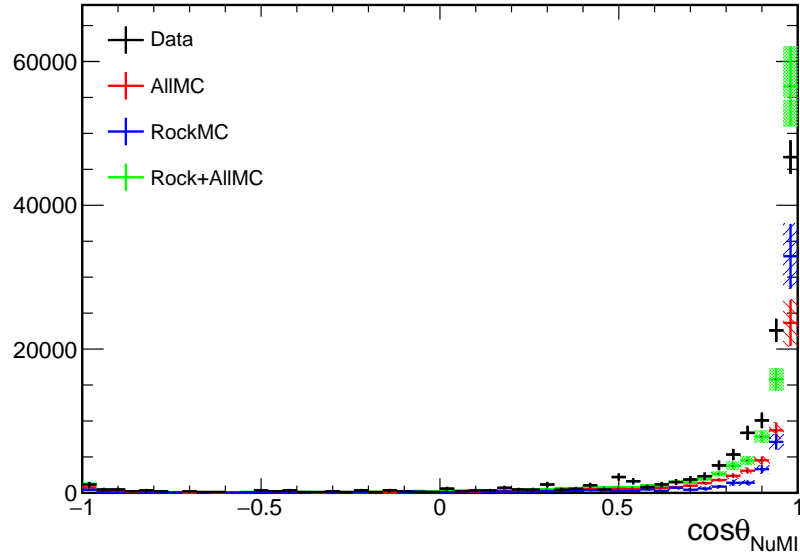


Figure 18: *Distribution of the tracks with respect to $\cos \theta_{\text{NuMI}}$.*

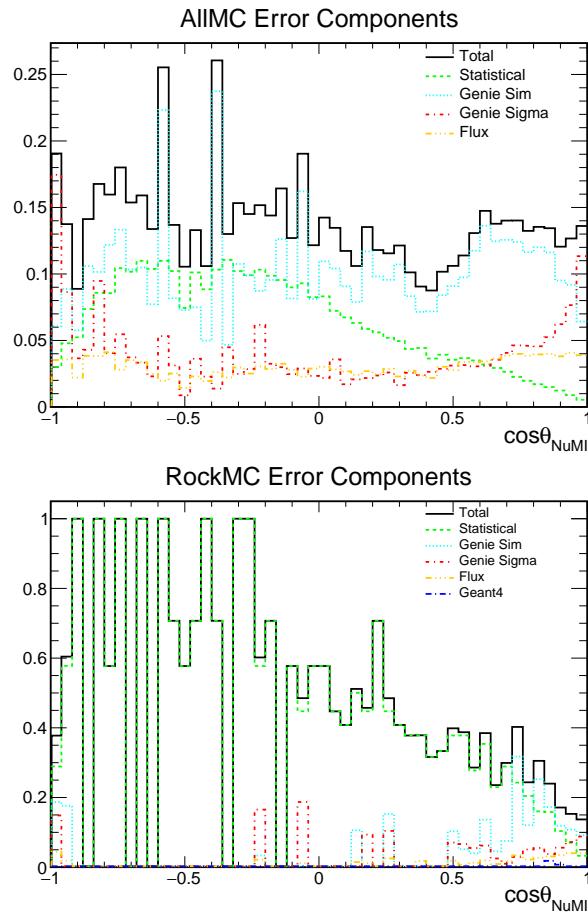


Figure 19: *Relative error components on $\cos \theta_{\text{NuMI}}$ distribution for the Monte Carlo samples.*

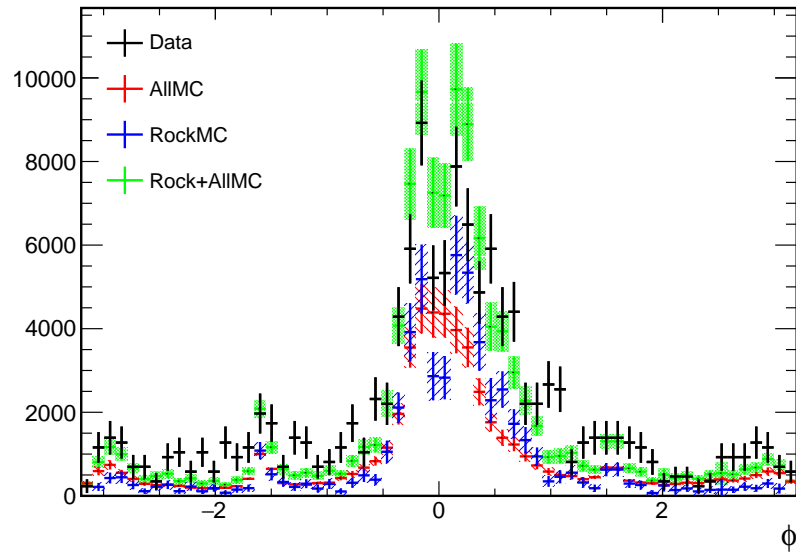


Figure 20: *Distribution of the tracks with respect to ϕ .*

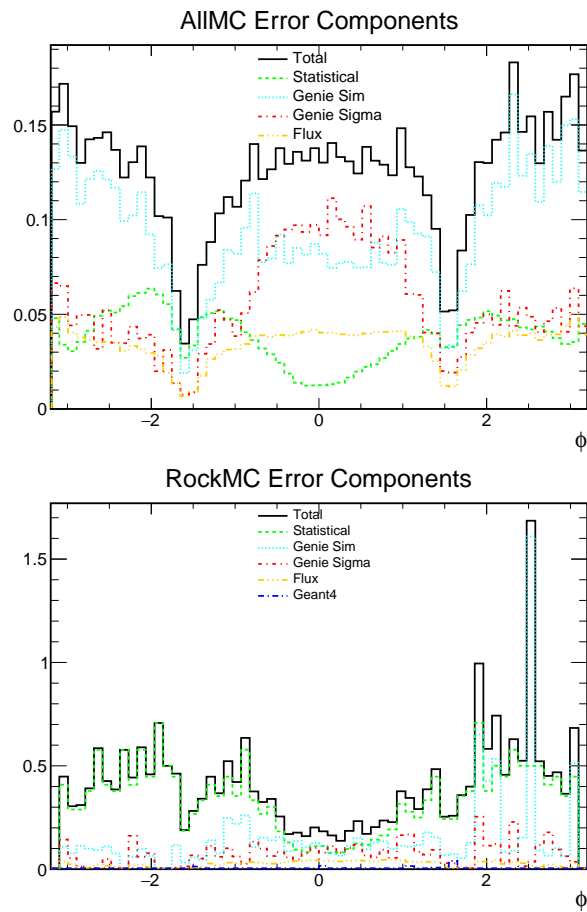


Figure 21: *Relative error components on ϕ distribution for the Monte Carlo samples.*

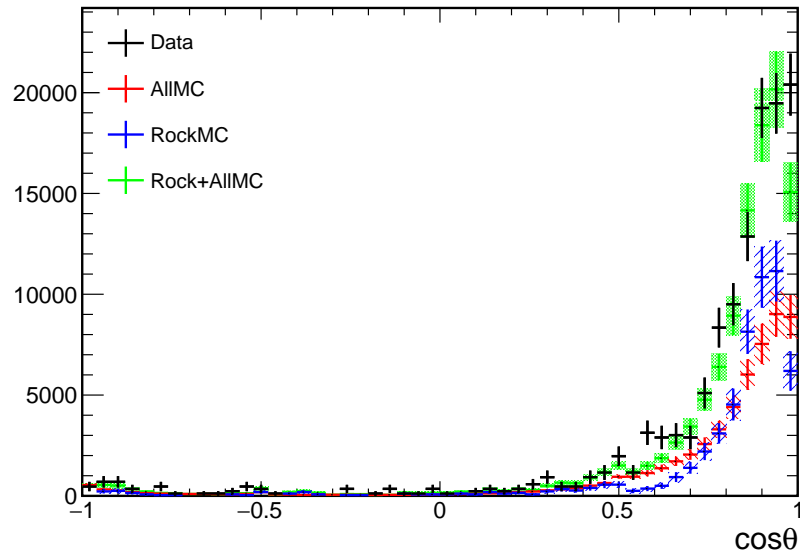


Figure 22: *Distribution of the tracks with respect to $\cos\theta$.*

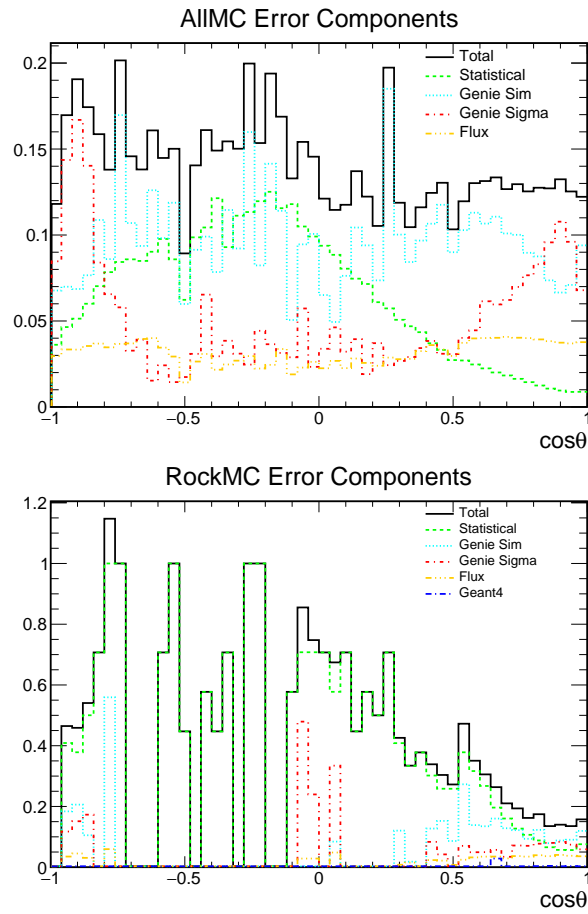


Figure 23: *Relative error components on $\cos\theta$ distribution for the Monte Carlo samples.*

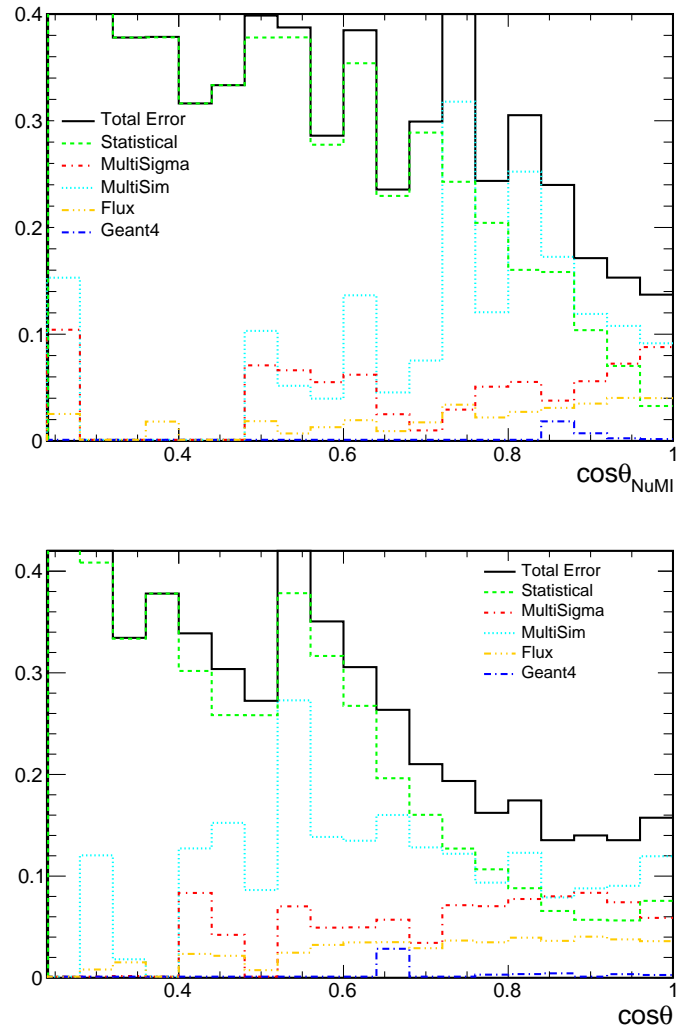


Figure 24: Zoom on the non statistic dominated regions for the errors on $\cos\theta$ and $\cos\theta_{\text{NuMI}}$ for the Rock MC.

4 Conclusions

In this work a selection for NuMI rock events for the ICARUS run 1 events has been developed. Firstly, some cuts to remove background events were established. In a second moment the most relevant systematic effects were added to the Monte Carlo samples. Finally some relevant distributions have been plotted. Some of these distributions show a somewhat good degree of compatibility, while some others show some tension between the Monte Carlo dataset and the experimental data. Such tensions will be addressed during a later stage of the analysis. In particular the current work is still missing the inclusion of in-time cosmic events and systematic detector effects.

The final selection aims at the validation of the Monte Carlo samples which will be used for the data analysis of ICARUS data. The goal of this analysis being the search of oscillation effects which may hint at the existence of a new unflavoured neutrino, namely the sterile neutrino. The observation of such effects may explain several anomalies found in the neutrino sector which, as of today, are still unsolved.

References

- [1] M. A. ACERO ET AL., *White paper on light sterile neutrino searches and related phenomenology*, 2023.
- [2] B. AHARMIM ET AL., *Independent measurement of the total active 8B solar neutrino flux using an array of 3He proportional counters at the Sudbury Neutrino Observatory*, Physical Review Letters, 101 (2008), p. 111301.
- [3] T. KAJITA ET AL., *Atmospheric neutrino results from Super-Kamiokande and Kamiokande evidence for ν_μ oscillations*, Nuclear Physics B-Proceedings Supplements, 77 (1999), pp. 123–132.
- [4] S. MELE, *The measurement of the number of light neutrino species at LEP*, in 60 Years of CERN Experiments and Discoveries, World Scientific, 2015, pp. 89–106.
- [5] F. POPPI ET AL., *The cosmic ray tagger system of the icarus detector at fermilab*, PoS ICHEP2022, 10 (2022), p. 0580.
- [6] THE GENIE COLLABORATION ET AL., *Recent highlights from GENIE v3*, The European Physical Journal Special Topics, (2021), pp. 1–19.
- [7] THE ICARUS COLLABORATION, *Design, construction and tests of the ICARUS T600 detector*, Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment, 527 (2004), pp. 329–410.
- [8] H. ZHANG, *Light sterile neutrino in the minimal extended seesaw*, Physics Letters B, 714 (2012), pp. 262–266.

A Setup and Macro Example

In order to facilitate the continuation of this work in what follows some basics to start working in the CAFana environment are included. Once the access to the Fermilab machines has been granted it is important to be able to set up the working environment.

The first thing to do is to properly setup a Kerberos account. Once this step is properly done it should be possible to create a ticket and login in to the ICARUS machines. Depending on your machine you may need proper setup for machine usage, in particular I have been working with Windows Subsystem for Linux (WSL) which require XServers like VcXsrv. Another possible solution may be working with MobaXterm.

A.1 Login

For a user `username` the Kerberos ticket creation process can look something like this

```
bash$ kinit -f username@FNAL.GOV
[Password for username@FNAL.GOV]: ****
```

The password will be provided by the service desk.

The current list of tickets can be displayed with the `klist` command.

Once a Kerberos ticket has been created it can be used for the following 24 hours, so the previous processes can be done only once each day. At this point it is possible to `ssh` login on the desired machine. For example logging in to the `icarusgpvm02` machine would become

```
bash$ ssh -kX icarusgpvm02.fnal.gov -l username
```

The `ctrl+D` keys allow to logout from the session.

Upon login users login to their home directory `/nashome/m/username`. In this directory it is convenient to store setup scripts for environment configuration. Setup operations can be very repetitive so using script can make this process faster and less tedious. In general it is not a good idea to work in this directory. A more suitable for work space can be reached with the following command.

```
bash$ cd /icarus/app/users/username/
```

Here it is possible to store macros and much more.

A.2 Environment Setup

Once you have logged inside of a machine it is possible to setup `icaruscode`. The easiest way is to use a default version. This approach is good for running small macros locally.

```
bash$ source /cvmfs/icarus.opensciencegrid.org/products/icarus/setup_icarus.sh
bash$ setup icaruscode v09_72_00 -q e20:prof
```

This setup provides with `icaruscode` version 09_72_00 with `e20:prof` qualifier.

The list of available versions can be found after `setup_icarus.sh` command with

```
bash$ ups list -aK+ icaruscode
```

where the `-aK+` flags are necessary in order to see all the instances and their relative informations.

A.3 Macro Example

The CAFana environment is good for LArTPC data analysis because of its physics oriented approach. In fact things are defined with respect to variables we want to observe and cuts we want to apply. Cuts can be defined both at spill and slice level. For code readability it may be a good idea to store all cuts and variables in an helper file (basically an header) when their number starts to become big. After proper variables and cuts have been defined they can be used to create a spectrum which will be loaded by a loader that takes as input the path to our trees (.root) containing the data we want to analyse. Once spectra have been filled they can be converted to histograms for proper display or comparison with other datasets.

The macro would look something like what follows.

To avoid issues it is recommended that

- The macro is written on a C++ file, generally .C files (NOT .c)
- The main loop has the same name as the file
(i.e. macro.C has a main function `void macro(){...}`)

```
#pragma once

#include "sbnana/CAFana/Core/Binning.h"
#include "sbnana/CAFana/Core/SpectrumLoader.h"
#include "sbnana/CAFana/Core/EnsembleRatio.h"
#include "sbnana/CAFana/Core/EnsembleSpectrum.h"
#include "sbnana/CAFana/Core/LoadFromFile.h"
#include "sbnana/CAFana/Core/Var.h"
#include "sbnana/CAFana/Cuts/TruthCuts.h"
#include "sbnana/CAFana/Systs/SBNWeightSysts.h"
#include "sbnana/CAFana/Analysis/ExpInfo.h"
#include "sbnanaobj/StandardRecord/Proxy/SRProxy.h"
#include "sbnana/SBNana/Vars/Vars.h"
#include "sbnana/SBNana/Vars/Binnings.h"
#include "sbnana/SBNana/Vars/NumuVars.h"
#include "sbnana/SBNana/Vars/NueVars.h"
#include "sbnana/SBNana/Vars/TruthVars.h"
#include "sbnana/SBNana/Cuts/Cuts.h"
#include "sbnana/SBNana/Cuts/TruthCuts.h"

#include "TFile.h"
#include "TH1.h"
//+other appropriate c++ ROOT INCLUDES

using namespace ana;

double POT = 10e8;

const Var kVertexX([](const caf::SRSliceProxy* slc){
    return slc->vertex.x;
});

const Cut kRFiducialX([](const caf::SRSliceProxy* slc) {
    return ( !isnan(slc->vertex.x) &&
        ( ( slc->vertex.x < -61.94 - 25 && slc->vertex.x > -358.49 + 25 ) ||
          ( slc->vertex.x > 61.94 + 25 && slc->vertex.x < 358.49 - 25 ) )
    );
});
```

```

const Binning kBinningX = Binning::Simple(50,-420.,420.);
const HistAxis axisXvtx ("X vtx", kBinningX, kVertexX);
void macro(){
    std::string fpath="path/to/file.root";
    SpectrumLoader loader(fpath);

    Spectrum sXPosition (loader, axisXvtx, kNoSpillCut, kRFiducialX);

    loader.Go();

    //sXPosition.OverridePOT(1000000);
    //POT override is required when doing POT normalised plots when POT
    //information is not present in the root files, for example on real data
    //MonteCarlo should in general have POT information

    TH1* hXPosition = sXPosition.ToTH1(POT);

    //+histo manipulation/plotting/saving
}

```

Obviously cuts and variables can be defined in more complicated ways. Cuts can also be combined in a very intuitive way, either defining a combined cut or in the `Spectrum` definition with `&&`. In order to negate a cut the `!` symbol can be used.

```

const Cut kCombined = kCut1 && !kCut2;
Spectrum sXPosition (loader, axisXvtx, kNoSpillCut, kRFiducialX && !kCombined);

```

Finally it is recommended to save the final histograms at the end of the code because large datasets may take from some minutes to several hours for execution.

```

TFile file ("finalHistos.root", "RECREATE");
hXPosition->Write("finalHistoName");

```

A.3.1 Macro Execution and ROOT

After proper setup configuration of the environments it is possible to run code and to use regular ROOT commands.

In order to run a CAFana macro the `cafe` command can be used.

```

bash$ cafe macro.C

```

The option `-bq` can be used to close the ROOT session at the end of the execution.

The best option to inspect `.root` files is the ROOT TBrowser which can be opened with the following commands.

```

bash$ root -l --web=off
root[0] TBrowser browserName

```

Once the ROOT session has been opened regular ROOT commands can be used without issues.

A.3.2 Systematics Effects

The inclusion of systematics effects is done at `Spectrum` level.

NuMI Flux Components

Flux components are defined as follows

```
int fluxComponents = 10; //how many flux components we are taking into
    consideration
int fluxSize = 1; //how much of a deviation we are applying
std::vector<const ISyst*> numi_flux_pcs = GetNuMIPCAFluxSysts(fluxComponents);

Spectrum* sVar = new Spectrum(loader, axis, spillCut, sliceCut, SystShifts(
    numi_flux_pcs[i],fluxSize), kGetNuMIFluxWeight);
//i the desired ith component of the flux correction
```

GENIE Components

GENIE contributions may be of two different kinds:

- Multisigma
- Multisim

For what concerns `multisigma` the procedure is the same as the one shown for the flux components:

```
int SigmaSize = 1; //how much of a deviation we are applying

const ISyst* systsSigmaKnob = new SBNWeightSyst("multisigmaComponent");

Spectrum* spectrumMultiSigma = new Spectrum(loader, axis, spillCut, sliceCut,
    SystShifts(systsSigmaKnob, SigmaSize), kGetNuMIFluxWeight);
```

On the other hand `multisim` corrections require the introduction of `EnsembleSpectrum` objects.

```
std::vector<Var> multisimWeight;
for(int i=0;i<100;i++){
    multisimWeight.push_back(GetUniverseWeight("multiSimComponent",i));
}

EnsembleSpectrum* ensSpectrum = new EnsembleSpectrum
    (loader, axis, spillCut, sliceCut, multisimWeight, kGetNuMIFluxWeight);
```

The manipulation of `EnsembleSpectrum` objects slightly differs from the `Spectrum` ones. From a `Spectrum` the `ToTH1(POT)` function is called to obtain the relative TH1 histogram. The `EnsembleSpectrum` instead gives the nominal spectrum via the `Nominal()` function (on which `ToTH1(POT)` is called). Furthermore the error bars associated are given via the `ErrorBand(POT)` function which produces a `TGraphAsymmErrors` object.

GEANT4 Components

GEANT4 components are implemented in the same way as `multisim` GENIE components.

B CAFana Cuts

B.1 Vertex Position

```
const Cut kRFiducial([](const caf::SRSliceProxy* slc) {
    return ( !isnan(slc->vertex.x) &&
        ( ( slc->vertex.x < -61.94 - 25
            && slc->vertex.x > -358.49 + 25 ) ||
```

```

        ( slc->vertex.x > 61.94 + 25
        && slc->vertex.x < 358.49 - 25 )
    )
    && !isnan(slc->vertex.y)
    && ( slc->vertex.y > -181.86 + 25
    && slc->vertex.y < 134.96 - 25 )
    && !isnan(slc->vertex.z)
    && ( slc->vertex.z > -894.951 + 30
    && slc->vertex.z < 894.951 - 50 ));
});

```

B.2 Track End Position

```

const Cut kPTrackContained([](const caf::SRSliceProxy* slc) {
    int Ind = kPTrackInd(slc);
    bool Contained(false);
    if ( Ind >= 0 ){
        auto const& trk = slc->reco.pfp.at(kPTrackInd(slc)).trk;
        Contained = ( !isnan(trk.end.x) &&
            ( ( trk.end.x < -61.94 - 10. && trk.end.x > -358.49 +
              10. ) ||
              ( trk.end.x > 61.94 + 10. && trk.end.x < 358.49 - 10.
                ) ) &&
            !isnan(trk.end.y) &&
            ( trk.end.y > -181.86 + 10. && trk.end.y < 134.96 - 10.
              ) &&
            !isnan(trk.end.z) &&
            ( trk.end.z > -894.951 + 10. && trk.end.z < 894.951 -
              10. ) );
    }
    return Contained;
});

```

B.3 Track Y Direction

```

const Cut kSlcTrkDiry([](const caf::SRSliceProxy* slc) -> double {
    if (slc->nuid.crlongtrkdiry > -0.2) {
        return true;
    } else {
        return false;
    }
});

```

B.4 Track Length

```

const Cut kSlcTk1([](const caf::SRSliceProxy* slc) -> double {
    if (kLongestTrackLength(slc) > 20) {
        return true;
    } else {
        return false;
    }
});

```

B.5 PMT Flash Times

```
const SpillCut OpFlash_MC([](const caf::SRSpillProxy* sr){
for (const auto& opflash : sr->opflashes){
    auto thistime = opflash.firsttime - sr->hdr.triggerinfo.
        trigger_within_gate;
    if (thistime > -0.1 && thistime < 9.7) {
        return true;
    }
}
return false;
});
```

C List of Systematics

This is the list of systematic correction applied to the two Monte Carlo samples.

C.1 GENIE MultiSigma

```
"GENIEReWeight_ICARUS_v2_multisigma_RPA_CCQE",
"GENIEReWeight_ICARUS_v2_multisigma_CoulombCCQE",
"GENIEReWeight_ICARUS_v2_multisigma_NormCCMEC",
"GENIEReWeight_ICARUS_v2_multisigma_NormNCMEC",
"GENIEReWeight_ICARUS_v2_multisigma_NonRESBGvpCC1pi",
"GENIEReWeight_ICARUS_v2_multisigma_NonRESBGvpCC2pi",
"GENIEReWeight_ICARUS_v2_multisigma_NonRESBGvpNC1pi",
"GENIEReWeight_ICARUS_v2_multisigma_NonRESBGvpNC2pi",
"GENIEReWeight_ICARUS_v2_multisigma_NonRESBGvnCC1pi",
"GENIEReWeight_ICARUS_v2_multisigma_NonRESBGvnCC2pi",
"GENIEReWeight_ICARUS_v2_multisigma_NonRESBGvnNC1pi",
"GENIEReWeight_ICARUS_v2_multisigma_NonRESBGvnNC2pi",
"GENIEReWeight_ICARUS_v2_multisigma_NonRESBGvbarpCC1pi",
"GENIEReWeight_ICARUS_v2_multisigma_NonRESBGvbarpCC2pi",
"GENIEReWeight_ICARUS_v2_multisigma_NonRESBGvbarpNC1pi",
"GENIEReWeight_ICARUS_v2_multisigma_NonRESBGvbarpNC2pi",
"GENIEReWeight_ICARUS_v2_multisigma_NonRESBGvbarnCC1pi",
"GENIEReWeight_ICARUS_v2_multisigma_NonRESBGvbarnCC2pi",
"GENIEReWeight_ICARUS_v2_multisigma_NonRESBGvbarnNC1pi",
"GENIEReWeight_ICARUS_v2_multisigma_NonRESBGvbarnNC2pi",
"GENIEReWeight_ICARUS_v2_multisigma_NormCCCOH",
"GENIEReWeight_ICARUS_v2_multisigma_NormNCCOH"
```

C.2 GENIE Multisim

```
"GENIEReWeight_ICARUS_v2_multisim_ZExpAVariationResponse",
"GENIEReWeight_ICARUS_v2_multisim_NCELVariationResponse",
"GENIEReWeight_ICARUS_v2_multisim_CCRESVariationResponse",
"GENIEReWeight_ICARUS_v2_multisim_NCRESVariationResponse",
"GENIEReWeight_ICARUS_v2_multisim_DISBYVariationResponse",
"GENIEReWeight_ICARUS_v2_multisim_FSI_pi_VariationResponse",
"GENIEReWeight_ICARUS_v2_multisim_FSI_N_VariationResponse"
```

C.3 GEANT4

GEANT4 corrections were only applied to the Rock Monte Carlo dataset.

```
"reinteractions_piminus_Geant4",  
"reinteractions_piplus_Geant4",  
"reinteractions_proton_Geant4"
```

D Useful Paths

In order to facilitate either the replication or the expansion of the work presented in this report in what follows are reported some useful paths on the ICARUS machines `icarusgpmv`.

D.1 Code

The `icaruscode` configuration used for running the code is stored in

► `/nashome/m/mcaporal/init.sh`

which works on a custom version of `icaruscode v09.76_00 e20`.

The modifications helped solving an issue with older `.root` trees with missing entries and allows for the inclusion of systematic effects.

The directory with the code which produces the histograms with the systematic components added is

► `/icarus/app/users/mcaporal/systematics/errorComponents.C`

For simplicity, in order to avoid clashes in the application of systematic correction, the code can run only on one dataset with systematic corrections applied. For the full comparison the code must be run separately with the `bool rock` set once on 0 (for AllMC dataset) and once on 1 (for RockMC). Since the Data dataset takes only a few tens of seconds to be processed it is not an issue to run on it twice. On the other hand, when machines are under load the processing of the AllMC dataset may require several hours.

The helper file with the definition of cuts and variables is stored in

► `/icarus/app/users/mcaporal/systematics/thisCuts.h`

After running on both Monte Carlo datasets final plots can be produced with the macro stored in

► `/icarus/app/users/mcaporal/systematics/output/plot.C`

In the `output` folder are stored all the plots and the `.root` files containing the histograms produced during the code execution.

D.2 Datasets

The Data dataset is stored at

► `/pnfs//icarus/scratch/users/jskim/data/Run1/NUMIMAJORITY/flatcaf/
DMCP2023G_Prescaled_reprod/flatcaf*.root`

The Rock Monte Carlo is instead stored at

► `/pnfs/icarus/scratch/users/gputnam/DMCP2023G/mc-reproc/D-MCNUPhase2` For a faster execution the `flatcaf*.root` files were merged in the following files

► `/pnfs/icarus/scratch/users/jskim/mc/NuMI_MC_Nu_Phase2/flatcaf/v09.72_00_03p01/
230919_Reproc/flatcaf_*.root`

The All Monte Carlo files, which were reprocessed by me using a modified version of Gray Putnam's scripts, can be found at

▶ /pnfs/icarus/scratch/users/mcaporal/DMCP2023G/mc-reproc/A-MCNuPhase2/
64671055_*/*flat.caf.root

The scripts used for reprocessing can be found at

▶ /icarus/app/users/mcaporal/DP2023G

The script for submission is

▶ /icarus/app/users/mcaporal/DP2023G/grid/mc-reproc/run_submit_recaf_MCNuPhase2.sh