

Reconstruction Efficiencies for Data & MC using the Muon Counter System

MICROBOONE-NOTE-XXXX-INT-vX

M. Bass^{*1} and S.R. Soleti^{†1}

¹*University of Oxford, Oxford, United Kingdom*

December 6, 2016

0 Purpose of Forthcoming Public Note

The message of the public note based upon this document is as follows:

- reconstructed tracks in the TPC can be matched to hits in an external cosmic ray counter (in our case the MuCS);
- the data reconstruction efficiency can be measured comparing the number of MuCS-triggered events with the number of MuCS reconstructed tracks;
- there is a systematic error in the measurement of the data reconstructing efficiency due to the presence of detector non-uniformities;
- the multiple scattering of the cosmic rays introduces an irreducible impurity in our sample of MuCS reconstructed tracks.
- taking into account these systematic effects, the data reconstruction efficiency obtained with the MuCS is in agreement with a Monte Carlo reconstruction efficiency generated uniformly across the TPC.

The figures that will be carried forward to the public note are Figg. ??, ??, ?? (pandoraCosmic only), ??-??, ??-??-??-??-??-??, ??-??-?? (pandoraCosmic only), ??, ??, ??, ??, ??, ?? and ??.

^{*}matthew.bass@physics.ox.ac.uk

[†]stefano.soleti@physics.ox.ac.uk

1 Introduction

This analysis note describes the work done to study the reconstruction efficiencies using a dataset of cosmic rays passing through the Muon Counter System (MuCS). The goal is to provide data and Monte Carlo reconstruction efficiencies that can be used to compare reconstruction performances and to show that an external cosmic-ray counter can be used to measure the data reconstruction efficiency in the LArTPC.

We measured the data reconstruction efficiency comparing the number of events triggered by the MuCS and the number of events with a MuCS-compatible reconstructed track. The Monte Carlo reconstruction efficiency, instead, was measured by comparing the number of generated cosmic rays with the number of reconstructed tracks.

The reconstruction efficiency expressed as a function of the cosmic-ray starting angle (given by the spherical angles θ and ϕ) and of the length L of its path in the TPC.

Using the `pandoraCosmic` algorithm [?] provided by the Pandora framework [?], the overall reconstruction efficiency is 96.1 ± 0.1 (stat) ± 1.1 (sys) % for data and 96.3 ± 0.1 % for Monte Carlo.

In the future, the method described in this paper will be adapted to use the data coming from the Cosmic Ray Tagger [?], which is able to tag around 80% of the cosmic rays hitting MicroBooNE LArTPC. In this way, we will be able to cover the entire (θ, ϕ, L) parameter space and measure efficiency-corrected quantities, such as the cosmic-ray flux in the LArTPC.

2 The Muon Counter System

The MuCS consists of two sets of planar modules made up of scintillator strips placed into two separate, light-tight boxes, readout via wavelength shifting fibers connected to Multi-Anode PMTs and placed on the top of the TPC. The Multi-Anode PMTs are readout by a DAQ system, separated from the DAQ system that reads out the TPC and PMT systems of the main detector.

Each planar module is made up of two sets of 24 scintillator strips, 4 cm wide, arranged into bi-layers oriented perpendicular to each other. This configuration provides two coordinates (z and x in the MicroBooNE TPC coordinate system) of the crossing points of the cosmic rays. Combining these two coordinates with the height of the modules (corresponding to the y coordinate in the MicroBooNE TPC coordinate system), it is possible to extrapolate the 3D trajectory of the cosmic ray.

This analysis has been performed on three merged dataset, acquired with different geometrical configurations. In each configuration, the two boxes have been placed at the upstream end, at the center and at the downstream end of the LArTPC, respectively.

A three-dimensional schematics of the three MuCS setups is shown in Fig. ?? and the coordinates of the two boxes for each configuration are reported in Tab. ??.

| Box | x [cm] | | y [cm] | | z [cm] | |
|-------------------|--------|-----|--------|-----|--------|------|
| | start | end | start | end | start | end |
| Upstream | | | | | | |
| Top box | -75 | -27 | 392 | 393 | 224 | 272 |
| Bottom box | -27 | 21 | 320 | 321 | 224 | 272 |
| Center | | | | | | |
| Top box | -72 | -24 | 397 | 398 | 579 | 627 |
| Bottom box | -27 | 21 | 320 | 321 | 581 | 629 |
| Downstream | | | | | | |
| Top box | -75 | -27 | 392 | 393 | 971 | 1019 |
| Bottom box | -27 | 21 | 320 | 321 | 971 | 1019 |

Table 1: Coordinates of the top and bottom box of the MuCS detector

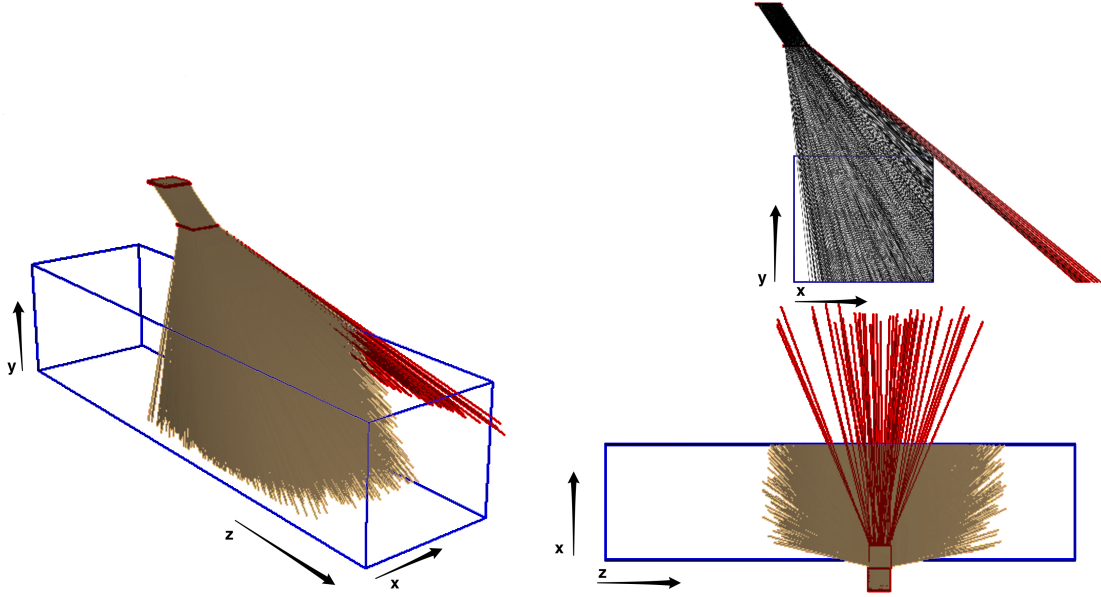


Figure 1: Monte Carlo simulation of the possible MuCS trajectories in the current configuration of the boxes. The brown tracks are those that intersect the TPC while the red tracks miss the TPC.

3 MuCS merging

The MuCS is designed to provide a trigger on through-going muons that intersect two planes of scintillator strips. The trigger is propagated to the MicroBooNE trigger system to record a full TPC and PMT readout. With the MuCS trigger in place, the t_0 for a track associated with the MuCS is known and these tracks are useful for various detector physics and reconstruction studies.

The dataset used for this study has been collected with the DAQ configured in the trigger-readout mode: the MuCS trigger is sent to the TPC readout, while the MuCS DAQ saves the hit patterns seen in the scintillator strips. The MuCS triggers at a rate of nearly 3 Hz and this rate is prescaled by 100 before sending the signal to trigger readout. In the general data stream then, a MuCS trigger is issued at a rate of 0.02 Hz (lower than 0.03 Hz because the MuCS trigger is often vetoed by beam triggers), or roughly one per minute. Given a DAQ integration window of 100 ns, the accidental coincidental rate is negligible for our study. The probability that, during the same readout window (4.8 ms), another cosmic ray hits the MuCS is, given a trigger rate of 3 Hz, 0.01% and it has also been considered negligible.

The data then follows a processing path that merges the MuCS hit patterns and reconstructed trajectory information with the TPC and PMT data stream to form a MuCS-merged dataset [?]. The TPC hits are fed to the reconstruction chain: then, the reconstructed track with the closest starting point to the extrapolated MuCS trajectory is stored in the dataset and it is tagged as a MuCS-reconstructed track (Figure). The PMT flash in time with the MuCS signal, if present, is also stored and associated to the MuCS-reconstructed track.

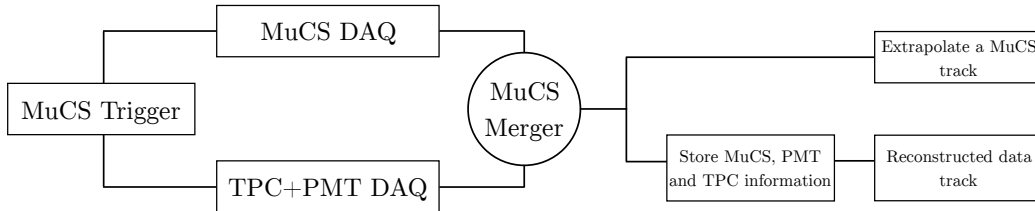


Figure 2: Flowchart showing the procedure used to generate the dataset used in this analysis. We end up with two data products: one with extrapolated coordinates obtained using MuCS-only data and one with reconstructed data information.

Our dataset will then include two different sets of information, with a one-to-one correspondence:

- MuCS-extrapolated information: using the two 2D points given by the MuCS (one for each box) we extrapolate a 3D line crossing the entire TPC. In this way we

87 obtain two extrapolated starting angles (θ_{MuCS} and ϕ_{MuCS}), an extrapolated track
 88 length $L - \text{MuCS}$ and extrapolated start/end points, using only MuCS information.

- 89 • Reconstructed TPC data information: for each event, we store the reconstructed
 90 spatial information of the MuCS-reconstructed track and of the in-time PMT flash,
 91 if present (starting angles θ_{data} and ϕ_{data} , track length L_{data} , start/end points).

92 4 MuCS-triggered cosmic rays angular distributions

93 Before comparing the MuCS-extrapolated information with the reconstructed informa-
 94 tion from Monte Carlo and data, a study on the angular distribution of the MuCS-
 95 triggered cosmic rays has been performed. Fig. ?? shows the number of MuCS-triggered
 96 events as a function of the starting angles θ_{xy} and θ_{yz} , defined as:

$$\begin{aligned}\theta_{xy} &= \arctan(\Delta y / \Delta x) \\ \theta_{yz} &= \arctan(\Delta y / \Delta z),\end{aligned}\tag{1}$$

97 where Δy (Δz) is defined as difference between the y_{top} (z_{top}) coordinate detected by top
 98 box and the y_{bottom} (z_{bottom}) coordinate detected by bottom box. Therefore, these dis-
 99 tributions use only MuCS information and are independent from the TPC reconstruction
 100 algorithms.

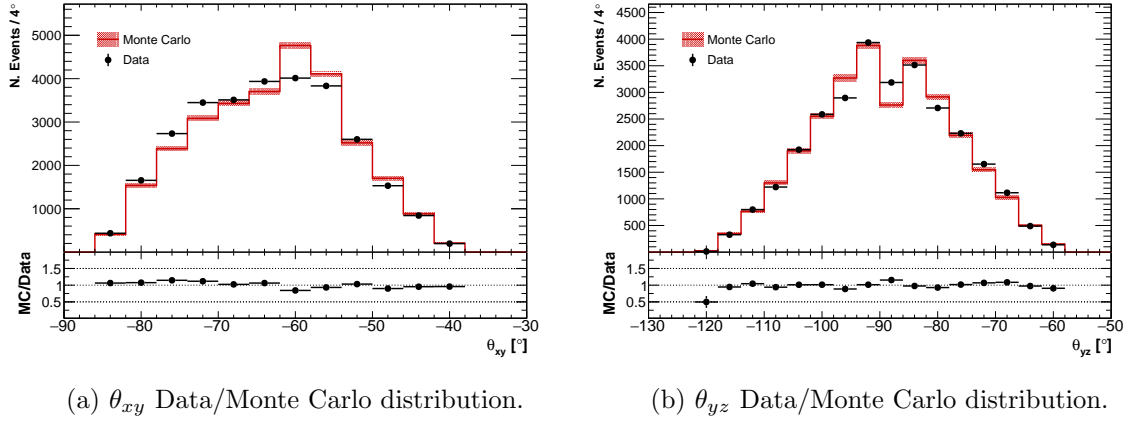


Figure 3: Data/Monte Carlo distributions of the angles of the cosmic rays hitting the MuCS, obtained extrapolating the MuCS hits. Errors are statistical only.

101 Monte Carlo distributions, in this case, have been obtained generating a CRY [?]
 102 sample of cosmic rays within the top box borders. Each event is stored only if the
 103 cosmic ray crosses also the bottom box. This Monte Carlo information must not be
 104 confused with the reconstructed Monte Carlo information described above, which is
 105 obtained reconstructing a simulated cosmic ray in the TPC.

106 The θ_{xy} distribution is peaked around -60° because of the relative shift of the two
 107 boxes in the x direction, as seen in Fig. ?. The θ_{yz} , instead, is symmetrical around

108 -90° because the two boxes are aligned on the z axis. The drop at $\theta_{yz} = -90^\circ$ is caused
 109 by of the staggered arrangement of the strips in each MuCS bi-layer: vertical cosmic
 110 rays will cause a signal in two strips relatively shifted by 2 cm, causing a bias in the
 111 extrapolation of the trajectory [?].

112 5 Minimum-ionizing particles discrimination

113 Since both MuCS boxes are positioned above the TPC, the MuCS-triggered events will
 114 correspond both to muons crossing the TPC entirely (minimum ionizing particles, MIPs)
 115 and, to a lesser extent, to muons stopping in the liquid argon. For each MuCS-triggered
 116 event, the corresponding Monte Carlo event is obtained generating a 4 GeV muon with
 117 a starting position and direction given by the hits of the data in the MuCS strips. A 4
 118 GeV muon has a negligible probability to stop in few meters of liquid argon: thus, in
 119 order to get a fair Monte Carlo/data comparison, it is necessary to remove the stopping
 120 muon events from the data.

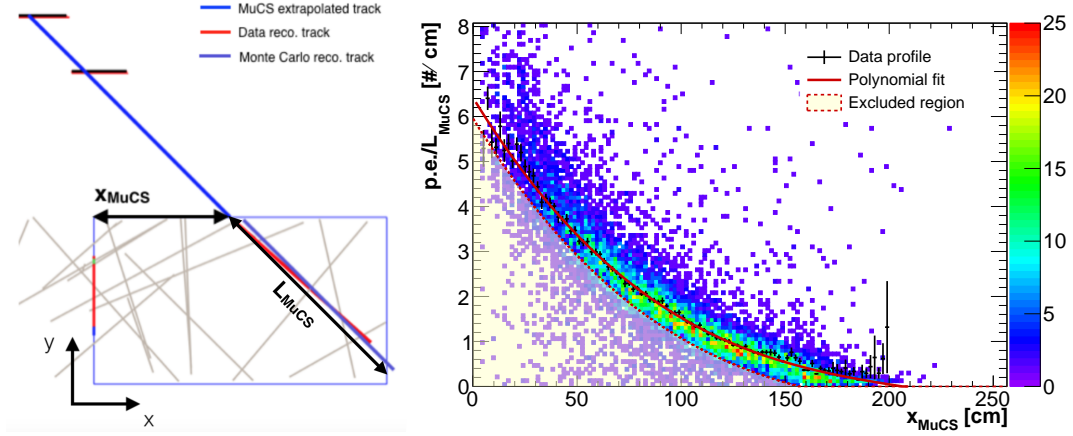


Figure 4: Left: 2D event display ($x - y$ projection) of a MIP event showing the extrapolated starting x coordinate (x^{MuCS}). Right: number of measured photoelectrons divided by the extrapolated length L^{MuCS} as a function of the extrapolated started x coordinate (x^{MuCS}). The shaded region correspond to the excluded events. L^{MuCS} and x^{MuCS} have been extrapolated from the MuCS hits coordinates.

121 It is possible to distinguish between MIPs and stopping muons using only the Micro-
 122 Boone optical system and the MuCS. For each event, the track length L^{MuCS} is calcu-
 123 lated extrapolating a straight line from the coordinates given by the MuCS. Then, the
 124 track length L^{MuCS} , which correspond to extrapolated path length of the muon through
 125 the detector (assuming it is through-going), is divided by the number of photoelectrons
 126 p.e., measured by the optical system.

127 The stopping muon events are expected to have a lower $p.e./L^{\text{MuCS}}$ ratio than the

MIP events, since their track length will be lower than the extrapolated L^{MuCS} . This ratio is plotted as a function of the extrapolated starting coordinate x^{MuCS} , to take into account the light attenuation. Since the PMTs are located at $x = 0$, cosmic rays that cross the TPC at high x coordinates will correspond to less photoelectrons in the optical system, because of smaller solid angle and light self-absorption in the liquid argon. Fig. ?? (left) shows a 2D event display with the MuCS extrapolated track and the starting coordinate x^{MuCS} .

The profile of the 2-dimensional histogram of the data points is fitted by a polynomial function. The events with a $\text{p.e.}/L^{\text{MuCS}}$ ratio below the fitting function have less light than expected for a MIP event and they are tagged as stopping muon events, corresponding to the shaded region in Fig. ?? (right).

The cut used to distinguish between MIPs and stopping muons depends on the chosen trade-off between the size and the purity of our MIP dataset. The shaded region of Fig. ?? (right) corresponds to events with at least 0.45 p.e./cm less than the expected value, given by the fitting function.

Our dataset contains 18704 MuCS-triggered events, acquired during run 3702 (November 12, 2015) and run 7106 (August 8, 2016). The MIP-selection cut tags 3158 events as stopping muons, reducing the size of the dataset by 17%. However, as the plot shows, it is not possible to discriminate between MIPs and stopping muons that hit the TPC further than 150 cm the anode, which correspond to very tilted cosmic rays ($\theta_{xy} < -60^\circ$) and very short tracks. Our dataset will then include an irreducible set of stopping muons.

6 Residual distributions

After removing the stopping muon events from the data it is possible to study the residual distributions between the data/Monte Carlo and the corresponding MuCS variables. In this study we will focus on the angular residual distributions of the starting angles θ_{xy} and θ_{yz} . The residual distributions are defined, for data and Monte Carlo, as:

$$\begin{aligned}\Delta\theta_{xy}^{\text{data/MC}} &= \theta_{xy}^{\text{MuCS}} - \theta_{xy}^{\text{data/MC}} \\ \Delta\theta_{yz}^{\text{data/MC}} &= \theta_{yz}^{\text{MuCS}} - \theta_{yz}^{\text{data/MC}},\end{aligned}$$

where the $\theta_{yz}^{\text{MuCS}}$ and $\theta_{xy}^{\text{MuCS}}$ angles are obtained extrapolating a line from the MuCS hits (??), while $\theta_{xy}^{\text{data/MC}}$ and $\theta_{yz}^{\text{data/MC}}$ are obtained reconstructing the hits in the TPC and measuring the starting angle of the track. The space-charge effect (SCE) in the data distributions has been corrected with the procedure described in [?].

For each MuCS-merged event we have simulated all the possible cosmic rays that can have the same geometrical hit pattern in the MuCS boxes. We introduced the effects of the MuCS strip widths (4 cm) into the Monte Carlo residual distributions with a Gaussian smearing. The σ parameter of the smearing Gaussian is different for each event and it corresponds to the standard deviation of the possible angles of the event hit pattern in the MuCS scintillator strips [?]. The mean σ are 0.6° and 0.7° for θ_{xy} and θ_{yz} , respectively.

The three different reconstruction algorithms used in this study are:

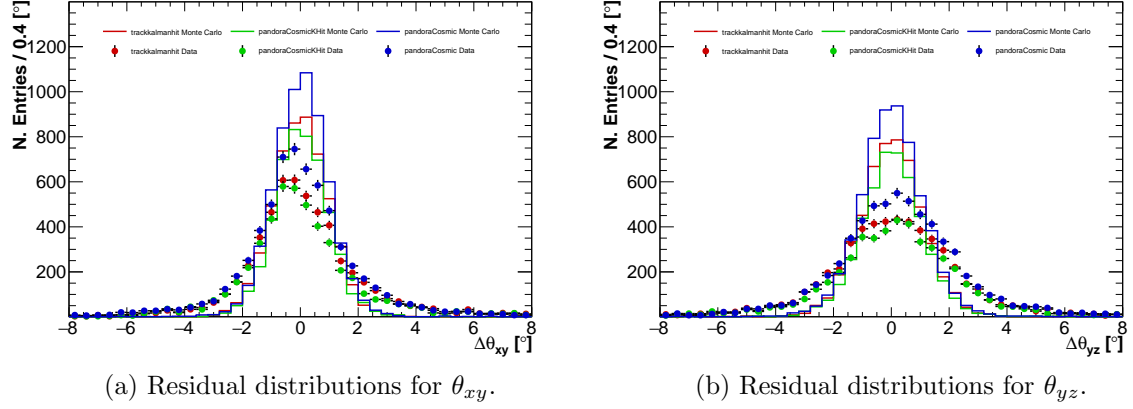


Figure 5: Monte Carlo (solid lines) and data (points) residuals distributions of θ_{xy} and θ_{yz} for **trackkalmanhit**, **pandoraCosmic** and **pandoraCosmicKHit** reconstruction algorithms.

- 162 • **trackkalmanhit**, which performs a Kalman filter [?] on the hits of 2D clusters,
 163 one for each wire plane (U, V and Y), and then merges the three 2D tracks to form
 164 a 3D track object [?];
- 165 • **pandoraCosmic**, which uses the Pandora framework [?] to reconstruct the tracks
 166 generated by cosmic rays: first, it creates three 2D clusters using a linear fit and
 167 then merges the clusters to create 3D space points. A linear fit is then performed
 168 on the space points to obtain a smooth 3D track object [?];
- 169 • **pandoraCosmicKHit**, which shares the same procedure of **pandoraCosmic**, but
 170 instead of a linear fit on the 3D space points, it performs a Kalman filter to obtain
 171 the final 3D track object [?].

172 They all give very similar residual distributions, as shown in Fig. ???. In order to
 173 normalize the data/Monte Carlo histograms, each entry correspond to an event with
 174 both a data reconstructed track and a Monte Carlo reconstructed track. The larger
 175 width of the data distributions is caused by two factors:

- 176 • multiple Coulomb scattering in the MuCS boxes, which are not included in the
 177 GEANT4 stage of the Monte Carlo simulation;
- 178 • different energy between the Monte Carlo and the data cosmic ray. Since we
 179 generate a 4 GeV Monte Carlo cosmic ray, corresponding data events with a less
 180 energetic cosmic ray will scatter more in the liquid argon, increasing the width of
 181 the residual distributions.

7 Reconstruction efficiencies

In this analysis we will compare the reconstruction efficiencies and performances of three different reconstruction algorithms mentioned above.

The following efficiencies have been measured comparing the number of MuCS-merged events with the number of Monte Carlo and data reconstructed events. Also in this case, the SCE effect on the angular coordinates has been corrected.

In this analysis, the reconstruction efficiency ϵ is defined as the ratio between the number of MuCS-tagged reconstructed events and the number of MuCS-triggered events. We will study it as a function of the starting angles, $\theta_{xy}^{\text{MuCS}}$ and $\theta_{yz}^{\text{MuCS}}$, and of the extrapolated cosmic-ray path length in the TPC, L^{MuCS} (from now on we will omit the MuCS for brevity):

$$\epsilon(\theta_{xy}, \theta_{yz}, L) = \frac{\text{N. of reco. events}}{\text{N. of MuCS-triggered events}}. \quad (2)$$

Being a function of three parameters, the reconstruction efficiency can be represented by a 3D plot, as shown in Fig. ??, where larger boxes correspond to bigger efficiencies. Each box correspond to a particular bin in the $\theta_{xy}, \theta_{yz}, L$ parameter space.

In order to better understand the reconstruction efficiency features, we are also reporting the 2D and 1D projections of the three-dimensional plot.

The three 2D projections, shown in Fig. ??, ??, ??, allow to measure the efficiency as a function of two of the three parameters ($\theta_{xy} - \theta_{yz}$, $\theta_{xy} - L$ or $\theta_{yz} - L$).

Fig. ?? shows the 1D projections of the 3D efficiency plot: in this case, it is possible to study the reconstruction efficiency as a function of the single parameter θ_{xy}, θ_{yz} or L .

The reconstruction efficiency, both for Monte Carlo and data, can be affected by several factors, including a cosmic ray path in the TPC too small to be reconstructed (Fig. ??) or a cosmic ray passing through a region of the TPC with missing wires. Fig. ?? and Fig. ?? show two examples of a MuCS cosmic ray not reconstructed because of a too short path length or because of a missing wires region, respectively.

However, while the Monte Carlo efficiency is very close to 1 for the entire parameter space, the corresponding data efficiency is significantly lower, even after the removal of the stopping muons with the procedure described in Sec. ?. So far, we identified two main factors that can contribute to this difference:

- stopping muons in the dataset that are not removed by the $p.e./L^{\text{MuCS}}$ cut;
- different energy between the Monte Carlo and the cosmic-ray data. Cosmic ray with a lower energy are scattered more in the TPC and then can be more difficult to reconstruct.

Appendix ?? shows Monte Carlo reconstruction efficiencies using 500 MeV cosmic rays. The overall efficiency in this case is $(81.4 \pm 0.4)\%$: we can safely assume that the reconstruction efficiency is proportional to the cosmic-ray energy. The Monte Carlo study described in Sec. ?? addresses this discrepancy, using a complete energy spectrum of cosmic rays instead of a fixed-energy sample.

| Reconstruction algorithm | Data [%] | | 4 GeV Monte Carlo [%] | |
|--------------------------------|----------|------|-----------------------|------|
| | avg. | err. | avg. | err. |
| <code>pandoraCosmic</code> | 84.6 | 0.4 | 96.5 | 0.2 |
| <code>pandoraCosmicKHit</code> | 64.8 | 0.5 | 93.8 | 0.2 |
| <code>trackkalmanhit</code> | 76.0 | 0.5 | 90.9 | 0.3 |

Table 2: Data and Monte Carlo overall efficiencies for three reconstruction algorithms. Errors are statistical only.

The low efficiency for $\theta_{xy} > -50^\circ$ (Fig. ??) is due both to the ineffectiveness of the MIP cut in this region and to the short path length of the very tilted cosmic rays. The short path length causes also the efficiency drops for $\theta_{yz} < -110^\circ$ and $\theta_{yz} > -70^\circ$. Also in the bi-dimensional efficiency plots (Fig. ??, ??, ??) the regions with the lowest efficiency are the ones with $\theta_{xy} > -50^\circ$.

The `trackkalmanhit` efficiency drop around $\theta_{yz} = -90^\circ$, both for data and Monte Carlo, is due to the track being aligned to the collection plane (Y) wires (Fig. ??). This geometrical configuration lowers the number of hits available by the Kalman Filter to reconstruct the cosmic ray track. The `pandoraCosmic` and `pandoraCosmicKHit` reconstruction algorithms do not exhibit this behavior.

The overall reconstruction efficiencies, obtained integrating the 3D plots, are reported in Table ?. The best reconstruction algorithm resulted to be `pandoraCosmic`, with an overall reconstruction efficiency of $(84.6 \pm 0.4)\%$ ($(96.5 \pm 0.2)\%$) for data (Monte Carlo).

- (a) 3D data efficiency with `pandoraCosmic`.

(c) 3D data efficiency with `pandoraCosmicKHit`.

(e) 3D data efficiency with `trackkalmanhit`.

(b) 3D Monte Carlo efficiency with `pandoraCosmic`.

(d) 3D Monte Carlo efficiency with `pandoraCosmicKHit`.

(f) 3D Monte Carlo efficiency with `trackkalmanhit`.

Figure 6: Data (left) and Monte Carlo (right) 3D efficiencies as a function of θ_{xy} , θ_{yz} and L for `trackkalmanhit`, `pandoraCosmic` and `pandoraCosmicKHit` reconstruction algorithms. Each box correspond to a particular combination of θ_{xy} , θ_{yz} and L . Larger boxes correspond to larger efficiencies.

- | | |
|---|--|
| (a) $\theta_{xy} - \theta_{yz}$, pandoraCosmic - data | (b) $\theta_{xy} - \theta_{yz}$, pandoraCosmic - Monte Carlo |
| (c) $\theta_{xy} - L$, pandoraCosmic - data | (d) $\theta_{xy} - L$, pandoraCosmic - Monte Carlo |
| (e) $\theta_{yz} - L$, pandoraCosmic - data | (f) $\theta_{yz} - L$, pandoraCosmic - Monte Carlo |

Figure 7: 2D scatter-plot of the reconstruction efficiencies for data (left) and Monte Carlo (right) for $\theta_{xy} - \theta_{yz}$, $\theta_{xy} - L$ and $\theta_{yz} - L$ 2D projections with the **pandoraCosmic** algorithm. Regions with lower data efficiency correspond to cosmic ray with a small path length in the TPC.

- | | |
|---|--|
| (a) $\theta_{xy} - \theta_{yz}$, pandoraCosmicKHit - data | (b) $\theta_{xy} - \theta_{yz}$, pandoraCosmicKHit - Monte Carlo |
| (c) $\theta_{xy} - L$, pandoraCosmicKHit - data | (d) $\theta_{xy} - L$, pandoraCosmicKHit - Monte Carlo |
| (e) $\theta_{yz} - L$, pandoraCosmicKHit - data | (f) $\theta_{yz} - L$, pandoraCosmicKHit - Monte Carlo |

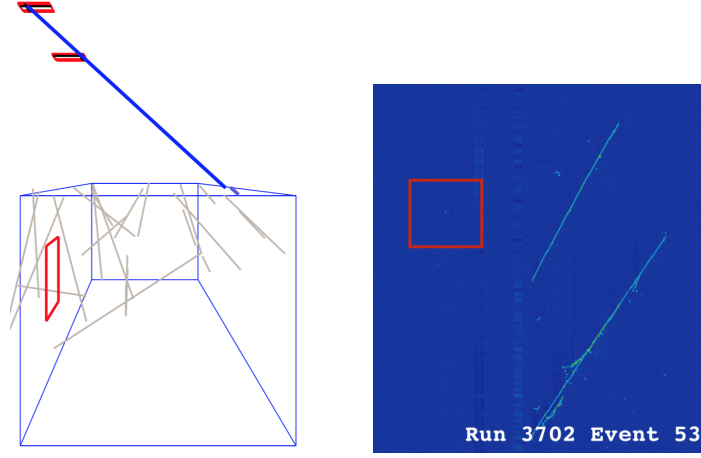
Figure 8: 2D scatter-plot of the reconstruction efficiencies for data (left) and Monte Carlo (right) for $\theta_{xy} - \theta_{yz}$, $\theta_{xy} - L$ and $\theta_{yz} - L$ 2D projections with the **pandoraCosmicKHit** algorithm. Regions with lower data efficiency correspond to cosmic ray with a small path length in the TPC.

- | | |
|--|---|
| (a) $\theta_{xy} - \theta_{yz}$, trackkalmanhit - data | (b) $\theta_{xy} - \theta_{yz}$, trackkalmanhit - Monte Carlo |
| (c) $\theta_{xy} - L$, trackkalmanhit - data | (d) $\theta_{xy} - L$, trackkalmanhit - Monte Carlo |
| (e) $\theta_{yz} - L$, trackkalmanhit - data | (f) $\theta_{yz} - L$, trackkalmanhit - Monte Carlo |

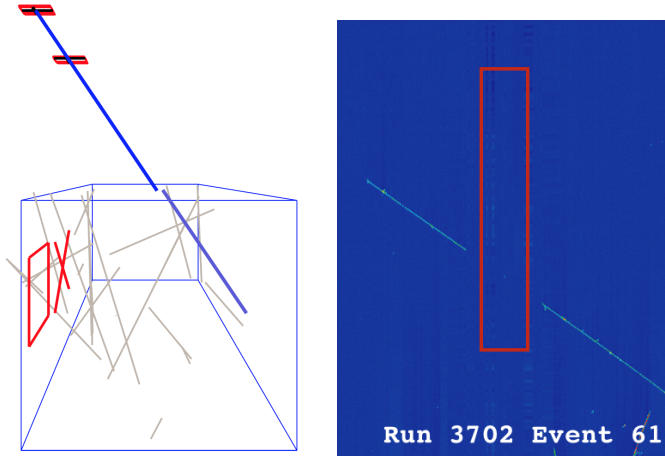
Figure 9: 2D scatter-plot of the reconstruction efficiencies for data (left) and Monte Carlo (right) for $\theta_{xy} - \theta_{yz}$, $\theta_{xy} - L$ and $\theta_{yz} - L$ 2D projections with the **trackkalmanhit** algorithm. Regions with lower data efficiency correspond to cosmic ray with a small path length in the TPC. In general, this algorithm has lower data/Monte Carlo efficiencies w.r.t. to the ones using the Pandora framework.

- | | |
|---|---|
| (a) Efficiency as a function of θ_{xy} . | (b) Efficiency as a function of θ_{yz} |
| (c) Efficiency as a function of L | |

Figure 10: Monte Carlo (boxes) and data (points) efficiencies as a function of θ_{xy} , θ_{yz} and L for **trackkalmanhit** (red), **pandoraCosmic** (blue) and **pandoraCosmicKHit** (green) reconstruction algorithms. In general, data efficiency is always lower than the Monte Carlo one. In the case of the L projection, the efficiency is proportional to the cosmic-ray path because a longer track has a larger number of hits in the TPC. The efficiency decreases when $\theta_{xy} > -40^\circ$, because this angle corresponds to cosmic ray with a short path length.



(a) Short length of the cosmic ray.



(b) Missing wires in the expected region of the cosmic ray.

Figure 11: 3D event reconstruction and corresponding event display for two MuCS-triggered events without tagged tracks. The red box (cross) in the 3D event corresponds to the data (Monte Carlo) flash. In the event on the top, the track was not reconstructed because the cosmic ray length in the TPC was too small. In the event on the bottom, the cosmic ray path was expected to be in a region with missing wires. The red boxes on the event display show the expected region for the cosmic ray tracks.

8 CRY Monte Carlo dataset

From the study of the Monte Carlo dataset generated with 4 GeV (Sec. ??) and 500 MeV (Sec. ??) fixed-energy cosmic rays it is possible to observe that the efficiency is energy-dependent. In particular, low-energy cosmic rays have a lower reconstruction efficiency, because of increased multiple Coulomb scattering.

Thus, in order to obtain a better Data/MC agreement we generated a new Monte Carlo sample using a complete cosmic-ray energy spectrum.

In this section, since our target is to obtain an efficiency-corrected cosmic-ray flux measurement, we will use the θ and ϕ angular coordinates, as defined in the MicroBooNE coordinate system [?]:

$$\begin{aligned} r &= \sqrt{x^2 + y^2 + z^2} \\ \theta &= \arccos \frac{z}{r} \\ \phi &= \arctan \frac{y}{x}. \end{aligned} \tag{3}$$

In our procedure, we generate a complete CRY simulation with 1000 events (1 cosmic ray per event) for each bin of the θ , ϕ parameter space, while the separation of the events on the L axis is given automatically by the geometry. In order to account for the angular dependence of the cosmic-ray flux, each three-dimensional bin is normalized by the number of MuCS-triggered events in the same bin, given by the data (Sec. ??). In this way, it is possible to incorporate the effects of the energy dependence in the measurement of the Monte Carlo efficiency.

For each event, we check if there is a reconstructed track that, extrapolated up to the MuCS box heights, crosses both boxes, with the same procedure used for data and described in Sec. ?. Since with this procedure we don't have a one-to-one correspondence between a MuCS-triggered event and a Monte Carlo event, the Monte Carlo efficiency is defined as:

$$\epsilon(\theta, \phi, L) = \frac{\text{N. of reco. events}}{\text{N. of generated events}}. \tag{4}$$

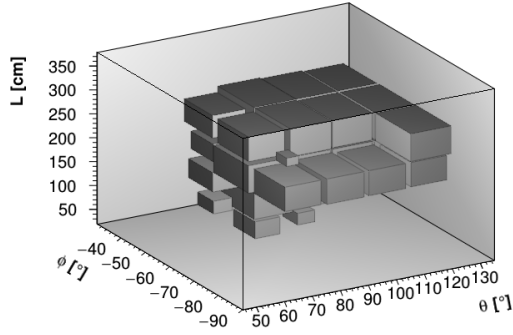
In order to remove stopping muon events, the same cut described in Sec. ?? is applied to the generated events.

The 3D θ - ϕ - L efficiency for data, Monte Carlo and data/Monte Carlo ratio, obtained with the `pandoraCosmic`, is shown in Fig. ?. Figs. ??, ?? and ?? show the θ - ϕ , θ - L and ϕ - L 2D projections, respectively, while the 1D projections are shown in Fig. ?.

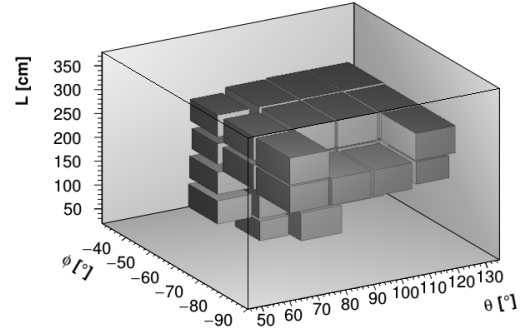
Tab. ?? reports the overall efficiencies for data, fixed-energy Monte Carlo samples (4 GeV and 500 MeV) and CRY sample. As we can see, this procedure sensibly improves the data/Monte Carlo agreement: the overall efficiency for Monte Carlo is $(85.0 \pm 0.1)\%$, compared with the $(96.5 \pm 0.2)\%$ obtained with the 4 GeV fixed-energy sample.

| | Data | | 4 GeV MC | | 500 MeV MC | | CRY MC | |
|------------|------|------|----------|------|------------|------|--------|------|
| | avg. | err. | avg. | err. | avg. | err. | avg. | err. |
| Efficiency | 84.6 | 0.4 | 96.5 | 0.2 | 81.4 | 0.4 | 85.0 | 0.1 |

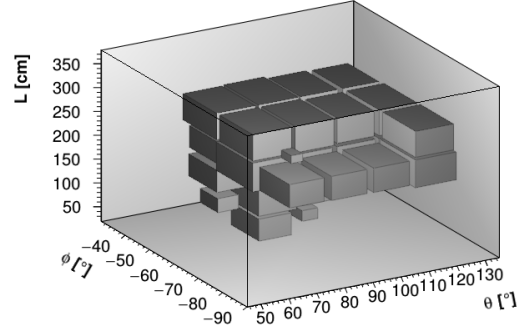
Table 3: Overall efficiencies for data, 4 GeV Monte Carlo, 500 MeV Monte Carlo and CRY Monte Carlo. Errors are statistical only.



(a) $\theta - \phi - L$, data

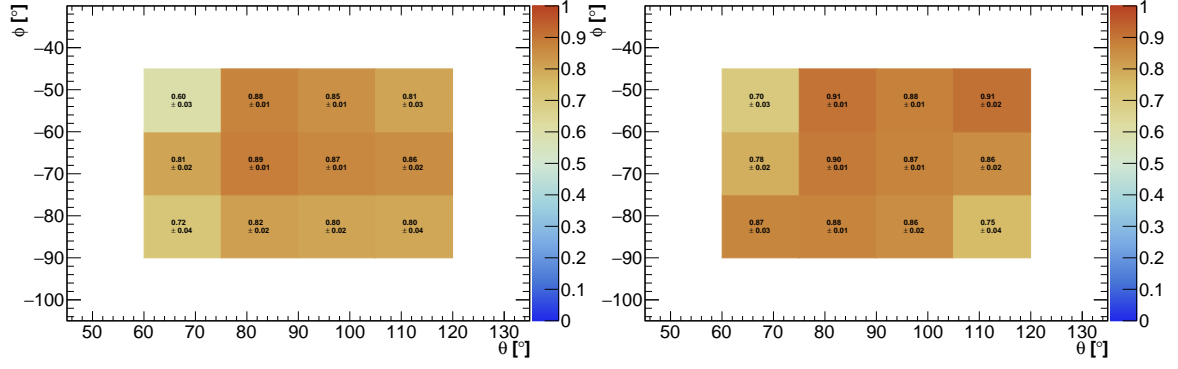


(b) $\theta - \phi - L$, Monte Carlo



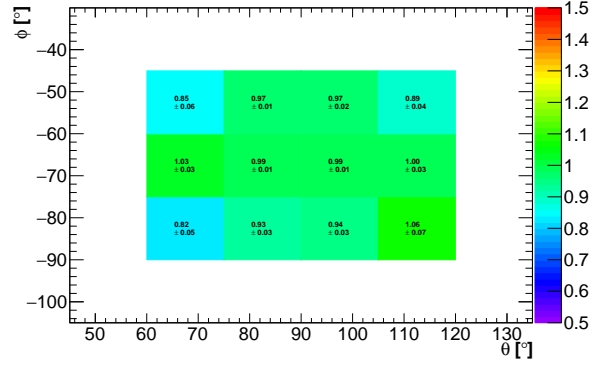
(c) $\theta - \phi - L$, data/Monte Carlo

Figure 12: 3D $\theta - \phi - L$ efficiency with the **pandoraCosmic** algorithm for data, Monte Carlo and data/Monte Carlo ratio.



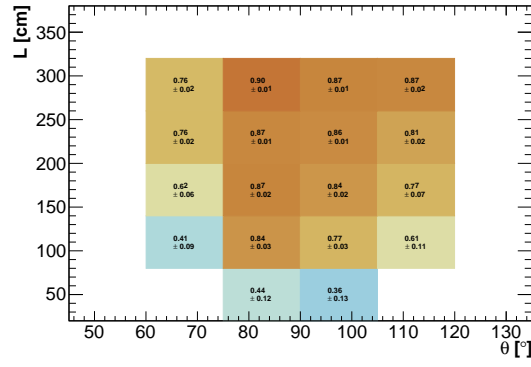
(a) $\theta - \phi$, data

(b) $\theta - \phi$, Monte Carlo

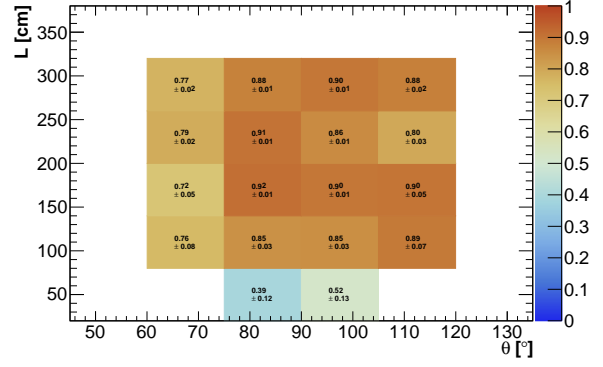


(c) $\theta - \phi$, data/Monte Carlo

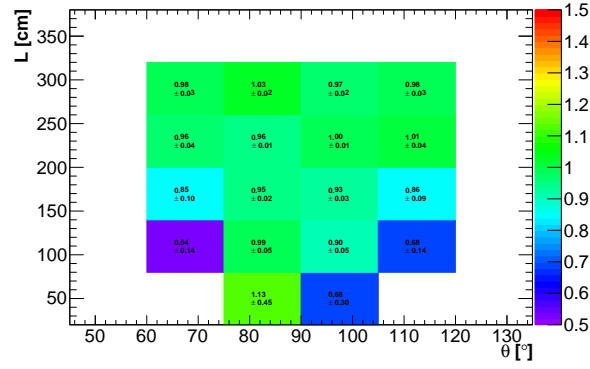
Figure 13: 2D $\theta - \phi$ efficiency with the **pandoraCosmic** algorithm for data, Monte Carlo and data/Monte Carlo ratio.



(a) $\theta - L$, data

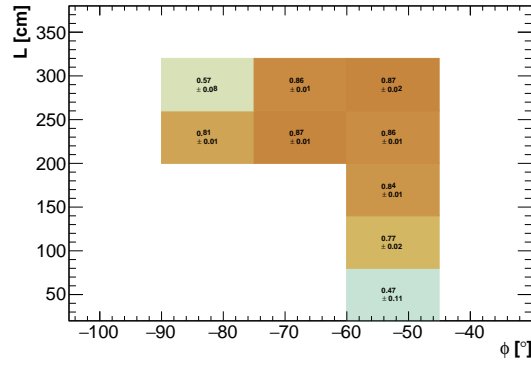


(b) $\theta - L$, Monte Carlo

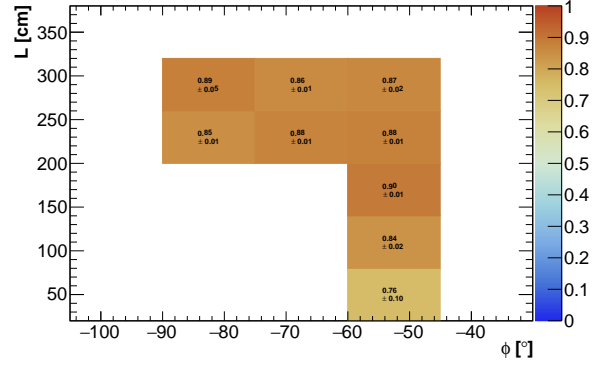


(c) $\theta - L$, data/Monte Carlo

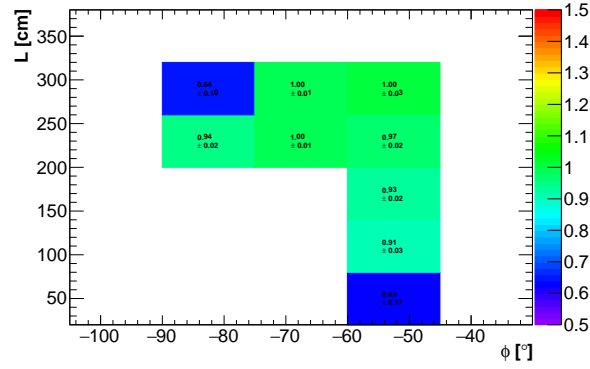
Figure 14: 2D $\theta - L$ efficiency with the **pandoraCosmic** algorithm for data, Monte Carlo and data/Monte Carlo ratio.



(a) $\phi - L$, data

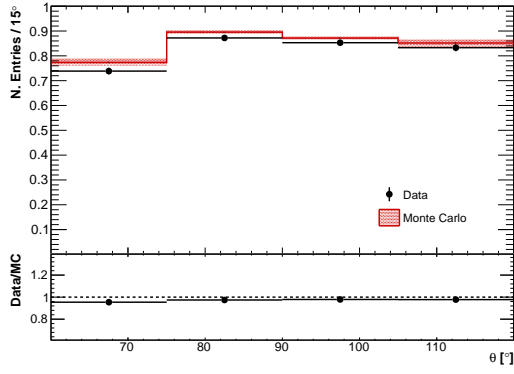


(b) $\phi - L$, Monte Carlo

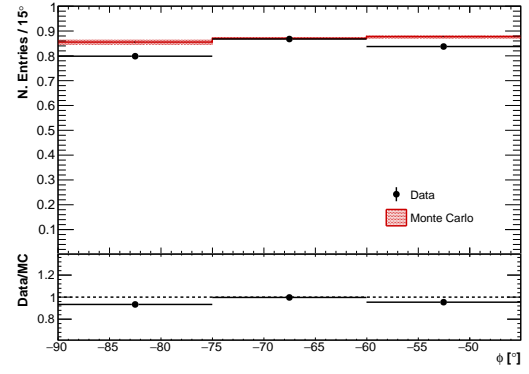


(c) $\phi - L$, data/Monte Carlo

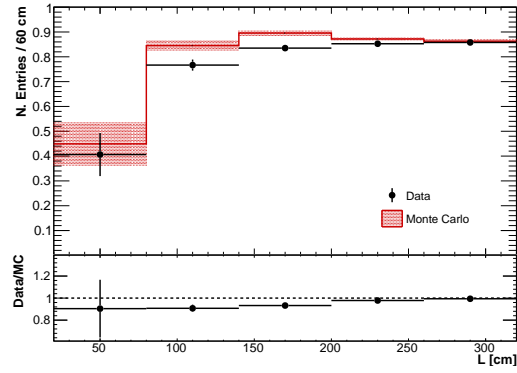
Figure 15: 2D $\phi - L$ efficiency with the `pandoraCosmic` algorithm for data, Monte Carlo and data/Monte Carlo ratio.



(a) θ , data and Monte Carlo



(b) ϕ , data and Monte Carlo



(c) L , data and Monte Carlo

Figure 16: 1D efficiencies with the `pandoraCosmic` algorithm for data and Monte Carlo.

9 Summary and plans

We compared the data/Monte Carlo reconstruction efficiencies for cosmic rays with three different algorithms, as a function of the starting angle and the path length of the cosmic rays, using a dataset of events triggered by the MuCS. We showed that the Monte Carlo reconstruction efficiency is energy-dependent and, in particular, lower energy cosmic rays correspond to a lower reconstruction efficiency. The Monte Carlo efficiency obtained with a CRY sample is in good agreement with the data. Leftover discrepancies can be caused by (1) differences between the CRY energy spectrum and data energy spectrum and (2) absence of not-MuCS cosmic rays in the Monte Carlo sample.

By moving the MuCS boxes it is possible to increase the coverage of $(\theta_{xy}, \theta_{yz}, L)$ parameter space. Our plan is to move the boxes at the center of the TPC on the z axis. Then, the boxes will be aligned in the x direction and their distance in the y will be reduced, in order to cover the region with $\theta_{xy} = 90^\circ$, $\theta_{yz} < -120^\circ$ and $\theta_{yz} > -60^\circ$, as shown in Fig. ??.

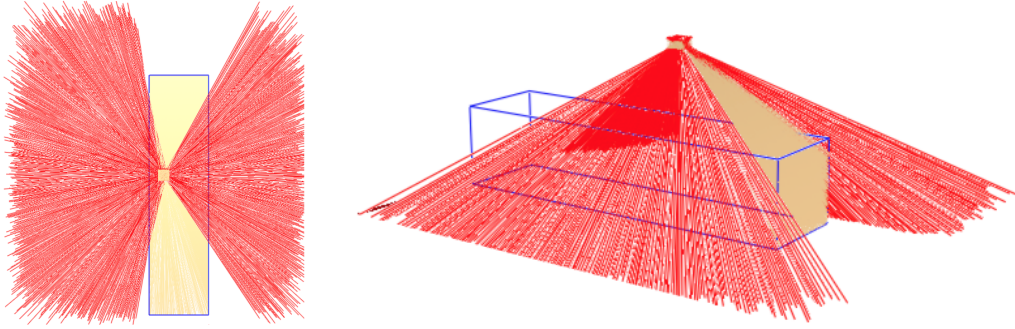


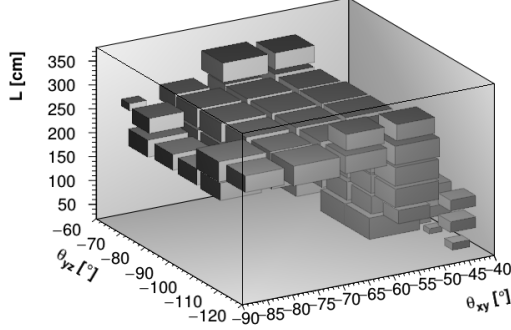
Figure 17: Monte Carlo simulation of the planned new positioning of the MuCS. Red lines show cosmic rays that trigger the MuCS but don't hit the TPC.

Once we have obtained a larger coverage of the $(\theta_{xy}, \theta_{yz}, L)$ parameter space, it will be possible to measure the efficiency-corrected cosmic-ray flux on the top of the TPC, both for data and Monte Carlo. Moreover, testing different regions of the TPC will allow us to measure eventual systematic errors caused by detector non-uniformity.

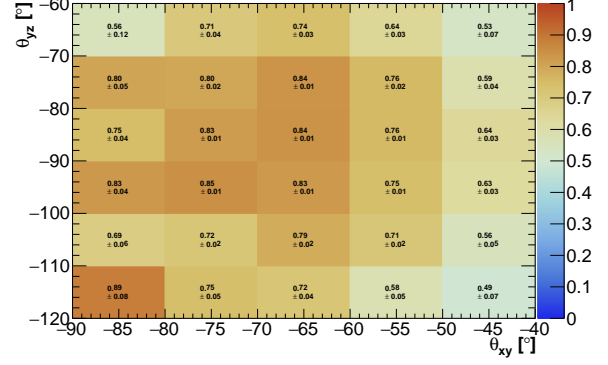
A 500 MeV Monte Carlo sample

In order to check if the differences observed between data and Monte Carlo are caused by the energy of the Monte Carlo cosmic rays, we generated a sample of 500 MeV cosmic muons and measured the efficiency in the same way described in Sec. ?. The efficiency plots, for `pandoraCosmic` only, are shown in Fig. ?? and ?. The integrated efficiency in this case is $(81.4 \pm 0.4)\%$, to be compared with $(96.5 \pm 0.2)\%$, obtained with the 4 GeV sample.

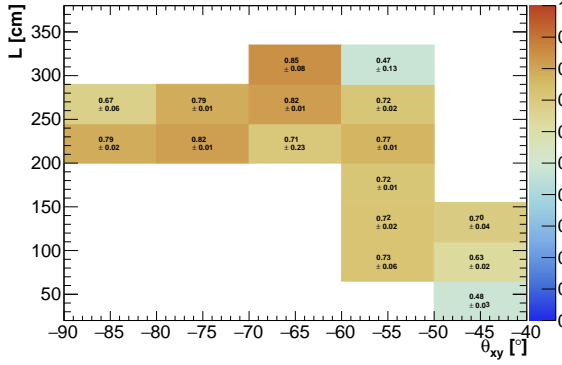
289 We can then safely assume that the reconstruction efficiency is proportional to the
 290 cosmic-ray energy, since cosmic rays with lower energy scatter more in the liquid argon
 291 and their track then more difficult to reconstruct.



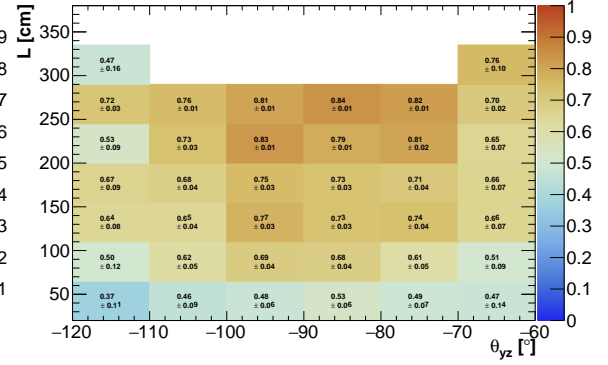
(a) 3D Monte Carlo efficiency with pandoraCosmic.



(b) $\theta_{xy} - \theta_{yz}$, pandoraCosmic - Monte Carlo

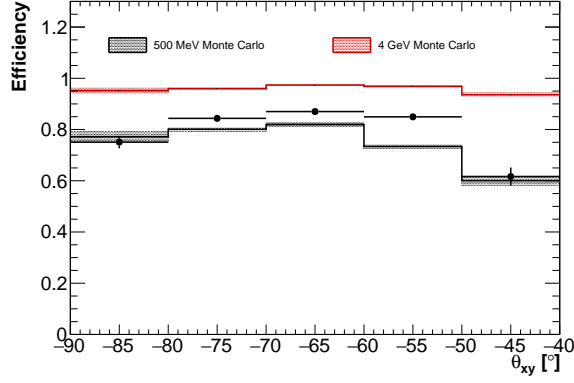


(c) $\theta_{xy} - L$, pandoraCosmic - Monte Carlo

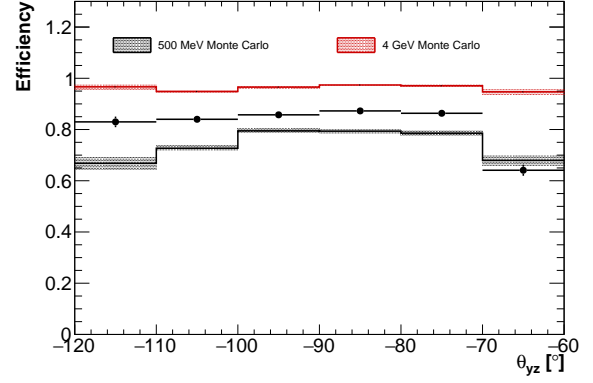


(d) $\theta_{yz} - L$, pandoraCosmic - Monte Carlo

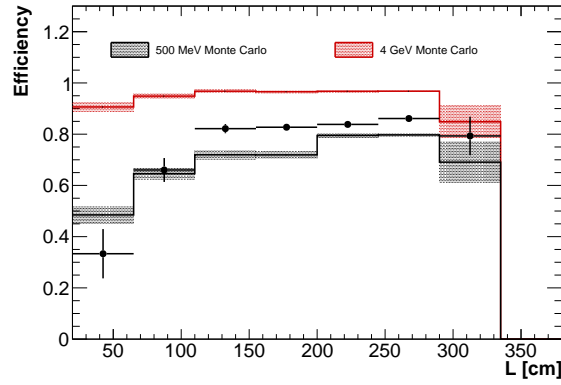
Figure 18: 3D reconstruction efficiency (top-left) and 2D scatter-plots for the 500 MeV Monte Carlo pandoraCosmic algorithm.



(a) Efficiency as a function of θ_{xy} .



(b) Efficiency as a function of θ_{yz} .



(c) Efficiency as a function of L .

Figure 19: Monte Carlo efficiencies with the 500 MeV (black) and 4 GeV (red) samples, as a function of θ_{xy} , θ_{yz} and L for the **pandoraCosmic** reconstruction algorithms.

B Technical details

The dataset used in this analysis correspond to run 3702 and run 7106. The reconstruction stages `reco1` and `reco2` use `uboonecode v05_08_00`. Their configuration is provided by the standard fcl files `reco_uboone_mcc7_driver_stage1.fcl` and `reco_uboone_mcc7_driver_stage2.fcl`. The source code of the `MuCSMerger` and `MuCSTagger` modules is available on the `uboonecode` repository.

The dataset with the merged MuCS, Monte Carlo and data information are stored in `/uboone/data/users/mibass/MuCS/MuCS/v05_08_00/MergedOuttree/`.

References

- [1] M. Bass, R. An, and L. Kalousis, *Muon Counter System Commissioning Review*, MicroBooNE docdb-4871.
- [2] M. Bass, M. Mooney, and S.R. Soleti, *Space-Charge Effect Measurements and Correction*, MicroBooNE docdb-5860.
- [3] H. Greenlee, *A Proposed C++ Toolkit for Kalman Filtering in Larsoft*, MicroBooNE docdb-3516.
- [4] R. Fruhwirth, *Application of Kalman filtering to track and vertex fitting*, Nucl.Instrum.Meth. A262 (1987) 444-450
- [5] M. Bass, and L. Kalousis, *MuCS Merger Update*, MicroBooNE docdb-5178.
- [6] M. Bass, L. Kalousis, and R. An, *MuCS Data/MC update*, MicroBooNE docdb-5025.
- [7] T. Usher, and E. Snider, *Summary of Reconstruction Updates*, MicroBooNE docdb-5723.
- [8] C. Hagmann, et al., *Cosmic-ray Shower Library (CRY)*, Lawrence Livermore National Laboratory, UCRL-TM-229453, 2012.
- [9] D. Heck, et al., *CORSIKA: A Monte Carlo code to simulate extensive air showers*, FZKA-6019, 1998.
- [10] M. Auger, et al., *A Novel Cosmic Ray Tagger System for Liquid Argon TPC Neutrino Detectors*, Instruments 2016, 1, 1.
- [11] J. S. Marshall and M. A. Thomson, *The Pandora Software Development Kit for Pattern Recognition*, Eur. Phys. J. C 75, no. 9, 439 (2015) doi:10.1140/epjc/s10052-015-3659-3 [arXiv:1506.05348 [physics.data-an]].
- [12] R. Acciarri, et al. [MicroBooNE Collaboration], *A Comparison of Monte-Carlo Simulations and Data from MicroBooNE*, MICROBOONE-NOTE-1014-PUB.