

¹ Selection of charged-current ν_μ inclusive events
² - Internal Note -
³ (MICROBOONE-NOTE-1010-INT)

⁴ Rui An⁵, Andy Blake⁷, David Caratelli³, Ben Carls⁴, Raquel
⁵ Castillo Fernández⁴, Eric Church¹⁰, Marco Del Tutto⁹, Danny
⁶ Devitt⁷, Steve Dytman¹¹, Brandon Eberly¹², Jessica Esquivel¹³,
⁷ Andrew Furmanski⁸, Sowjanya Gollapinni⁶, Nick Graf¹¹, Herb
⁸ Greenlee⁴, Pip Hamilton¹³, Chun Min (Mindy) Jen¹⁴, Libo Jiang¹¹,
⁹ David Kaleko³, Wes Ketchum⁴, Mike Kirby⁴, Adam Lister⁷, Xiao
¹⁰ Luo¹⁵, John Marshall², Varuna Meddage⁶, Tia Miceli¹⁶, Jarek
¹¹ Nowak⁷, Zarko Pavlovic⁴, Aleena Rafique⁶, Leon Rochester¹²,
¹² Brooke Russell¹⁵, Anne Schukraft⁴, Erica Snider⁴, Matt Toups⁴,
¹³ Serhan Tufanli¹⁵, Tracy Usher¹², Christoph Rudolf von Rohr¹,
¹⁴ Michele Weber¹, Jack Weston², Athula Wickremasinghe¹¹, and
¹⁵ Tingjun Yang⁴

¹⁶¹Laboratory for High Energy Physics, University of Bern,
¹⁷ Switzerland

¹⁸²University of Cambridge

¹⁹³Columbia University

²⁰⁴Fermilab

²¹⁵Illinois Institute of Technology

²²⁶Kansas State University

²³⁷Lancaster University

²⁴⁸University of Manchester

²⁵⁹University of Oxford

²⁶¹⁰Pacific Northwest National Laboratory

²⁷¹¹University of Pittsburgh

²⁸¹²SLAC

²⁹¹³Syracuse University

³⁰¹⁴Virginia Tech

³¹¹⁵Yale University

³²¹⁸New Mexico State University

³³ Editorial Board:
³⁴ Xin Qian* and Mike Shaevitz†

Abstract

This internal note is describing the chain to select a sample of ν_μ charged-current neutrino interaction from the MicroBooNE BNB data stream. The BNB data sample has been constrained to a set of good runs that fulfill data quality criteria and add to about 5e19 POT, which is MicroBooNE's unblinded data sample. The following chapters document the selection of runs, the reconstruction algorithms used, the studies on detector stability and data/MC agreement. Two independent and complementary event selections are described and (kinematic) event distributions of ν_μ CC events from both selections are presented. The work contained in this note is a snapshot towards the measurement of a CC inclusive cross section and intended to be shown at NEUTRINO2016.

Contents

1	Introduction	4
50	1.1 Context and Scope of this note	4
51	1.2 Organization of the note	4
2	Data and MC samples	5
52	2.1 Data And MC processing	5
53	2.2 Good run list	6
54	2.2.1 Detector conditions	7
55	2.2.2 Data quality	7
56	2.2.3 Beam conditions	8
57	2.2.4 Run processed	8
58	2.3 Final samples	8
59	2.4 Normalization of off-beam to on-beam data	9
60	2.5 Normalization of MC samples	9
3	Optical triggering and reconstruction	11
62	3.1 Software trigger	11
63	3.2 Flash reconstruction	12
64	3.3 Beam window	13
65	3.4 Analysis flash threshold	14
66	3.5 Impact of shortcomings in the optical simulation on this analysis	16
4	TPC reconstruction	19
68	4.1 Hit reconstruction	19
69	4.2 Clustering	20
70	4.3 Pandora	20
71	4.4 trackkalmanhit	21
72	4.5 Cosmic hit removal	21
73	4.6 Projection Matching Algorithm	21

*Brookhaven Lab

†Columbia University

75	4.7	Calorimetry	22
76	4.8	Tracking efficiency comparison	22
77	4.9	Treatment of noise and dead wires in the MC and impact on data/MC agreement	23
79	5	Data-MC comparisons	24
80	6	Event selection	26
81	6.1	Selection I	27
82	6.1.1	Event Selection Scheme (original)	27
83	6.1.2	Modified Event Selection Scheme	28
84	6.1.3	Expected Backgrounds	30
85	6.1.4	Truth Distributions of Selection I	33
86	6.2	Selection II	46
87	6.2.1	Selection IIA: track multiplicity 2 (or larger) and no con- tainment requirement	46
88	6.2.2	Selection IIB: single tracks with containment requirement	52
89	6.2.3	Selection II signal efficiency and background rejection .	53
90	6.2.4	Truth distributions for selection II	56
91	6.2.5	Passing rates for different physics processes	60
92	6.3	Overlap between selections	64
94	7	Systematic variations	66
95	7.1	BNB neutrino Beam	66
96	7.2	Model systematics	70
97	7.2.1	Model variations through reweighting	70
98	7.2.2	Model changes through re-generation	76
99	7.3	Detector systematics	80
100	7.3.1	Impact on muon tracking efficiency	81
101	7.3.2	Impact of 3 ms lifetime sample on data	83
102	8	Final event Distributions	90
103	8.1	CC inclusive event distributions for Selection I (original)	90
104	8.1.1	Thoughts on the observed excess in θ and ϕ	91
105	8.1.2	Cut on Vertical Quadrants of ϕ	93
106	8.2	CC inclusive event distributions for Selection I (modified)	96
107	8.3	BNB Systematic Uncertainties for Selection I	100
108	8.4	CC inclusive event distributions for Selection II	103
109	8.5	Event views	108
110	A	GENIE Model Uncertainties - Cross section difference plots	117
111	B	Normalized On-Beam and Off-Beam Distributions for selection	
112	I		122
113	C	Normalized On-Beam and Off-Beam Distributions for selection	
114	II		127
115	D	Selection studies with various reco algorithms	131

116 1 Introduction

117 1.1 Context and Scope of this note

118 This note contains material that presents the state of the event selection for a ν_μ
119 CC inclusive analysis by the time of the NEUTRINO2016 conference. Part of
120 these plots are intended to be shown as part of the NEUTRINO2016 conference
121 talk and posters. This note is not intended for a publication.

122 Before we plan to publish this material, we will go back and reiterate through
123 several aspects, for example the low-level agreement of data and MC in TPC
124 and PMT data, noise simulation, overlays of off-beam cosmics with BNB MC, a
125 proper flash matching algorithm, improve tracking efficiency, data-driven study
126 of tracking efficiency, tune cut variables in order to increase the selection effi-
127 ciency, and others.

128 Since many of these topics are not addressed in this note, we do not claim
129 to have understanding of systematic uncertainties. Our intend for the presen-
130 tation of this work is to apply our event selections and show on-beam minus
131 off-beam data and compare to MC predictions. These plots clearly indicate that
132 we are able to select neutrinos from the overwhelming background of cosmics
133 and that these neutrino interactions are qualitatively described by the expected
134 muon kinematics distributions. No absolute POT normalization is being done,
135 all plots are area normalized.

136

137 1.2 Organization of the note

138 Chapter 2 describes the processing chain used for data and MC. It further
139 discusses how good runs are selected, as well as beam criteria and detector
140 stability tests that were perform to further downsize the sample of used runs
141 and ensure data quality. In this analysis, the signal is obtained by subtracting
142 off-beam from on-beam selected events. This is compared to BNB + Cosmic
143 MC, where BNB simulation is produced using the baseline version of GENIE
144 2.8.6, and Cosmic particles are simulated using CORSIKA. The normalization
145 applied is also described in chapter 2.

146 Chapter 3 describes the optical triggering, reconstruction, and the analysis
147 cuts based on optical information. Only events with a flash of minimum height
148 (measured in photo-electrons (PE)) during the beam window are used. Chapter
149 3 gives details on the determination of this threshold as well as an indication of
150 the influence of shortcomings in the current optical simulation on the result.

151 Chapter 4 discusses the TPC reconstruction chain that was applied for this
152 analysis and gives references to documentation of the individual algorithms used.

153 Chapter 5 is dedicated to comparisons between off-beam data and CORSIKA
154 MC before the application of cuts. All variables that are being used downstream
155 in the analysis are being compared (see Data/MC comparison internal note,
156 DocDB 5849).

157 Chapter 6 discusses the two different event selection that were developed for
158 this analysis. Both selections have very different approaches. They use different
159 reconstruction algorithms and cuts. The chapter describes the cuts and passing
160 rates for both selections. The two selections are presented side by side, because
161 at this point it is good to have two selections to compare features and detec-

162 tor/reco effects. Different reconstruction algorithms have different strength and
163 weaknesses. As we haven't defined a standard reconstruction chain in Micro-
164 BooNE, it is useful to try different algorithms and compare the outcomes. The
165 reconstruction chains are described in chapter 4. There is no plan to merge the
166 two selections at this time.

167 Chapter 7 discusses some of the systematic parameters that have been studied
168 in this analysis. The treatment of systematic uncertainties is not complete
169 at this point. Instead, we picked some parameters which we believed had larger
170 impact, and studied their influence on our final event distributions. These pa-
171 rameters are chosen from three groups: beam uncertainties, model uncertainties,
172 and detector uncertainties.

173 Chapter 8 shows the final event distributions for both selections and com-
174 pares on-beam minus off-beam data with BNB+Cosmic MC. In addition to
175 comparing to the baseline MC, it will in a future version also compare to se-
176 lected systematic data sets. This chapter contains the main plots that we are
177 going to propose to release for the NEUTRINO2016 conference.

178 2 Data and MC samples

179 2.1 Data And MC processing

180 The data are analyzed from three streams based upon hardware and online-
181 software trigger requirements. Events came from the BNB_INCLUSIVE, EXT_BNB_INCLUSIVE,
182 and BNB_UNBIASED streams and were used for selecting signal events and for
183 determining background contributions. The BNB_INCLUSIVE stream (BNB
184 Beam On) is selected by requiring that the BNB hardware trigger bit fire and
185 that the event pass the software trigger optical requirements in a BNB-spill-sized
186 timing window (details of the software trigger can be found in Section 3. The
187 EXT_BNB_INCLUSIVE stream (BNB Beam Off) is selected by requiring that
188 the EXT hardware trigger bit fire (which is orthogonal to all beam triggers) and
189 that the event pass the software trigger optical requirement in a BNB-spill-sized
190 timing window. The BNB_UNBIASED stream is selected by applying a prescale
191 ($\approx 0.0026 \rightarrow 0.01$ Hz rate) to all events where the BNB hardware trigger bit
192 fired with no software trigger requirement.

193 The MC samples were generated with **uboonecode v04.36.00** as the base
194 release and some samples generated with patch releases based upon that. This
195 release of uboonecode is based upon the following packages:

196

- 197 • **larsoft v4.36.00**
- 198 • **GEANT v4.9.6.p04d**
- 199 • **GENIE v2.8.6d**
- 200 • **GENIE xsec v2.8.6a**
- 201 • **pandora v2.3.0a**
- 202 • **CORSIKA v7.4003**

Table 1: The simulated samples generated for Neutrino2016 analysis and used in the inclusive CC event selection.

Sample	Gen Evts	fcl file
BNB+Cosmics	189380	prodgenie_bnb_nu_cosmic_uboone.fcl
BNB Only	654400	prodgenie_bnb_nu_uboone.fcl
inTime CORSIKA	10000	prodcosmics_corsika_cmc_uboone.fcl

203 The three MC samples used were generated as GENIE BNB neutrino interactions with cosmic ray overlay (BNB+Cosmics), GENIE BNB neutrino interactions without cosmic ray overlay (BNB Only), and CORSIKA cosmic rays within the readout window (inTime CORSIKA, later referred to as *Cosmic only*). The cosmic ray overlay for the BNB+Cosmic sample are generated with CORSIKA and the fcl file. Details of the samples are shown in Table 1.

204 Both data and MC samples were processed with the same reconstruction and
 205 Analysis Tree release, **uboonecode v05.08.00**. The reconstruction of data was
 206 performed in two stages using the configuration files reco_uboone_data_Feb2016_driver_stage1.fcl
 207 and reco_uboone_data_Feb2016_driver_stage2.fcl. After reconstruction, Analysis
 208 Tree files for data were produced using the configuration file standard_ana_uboone_data.fcl.
 209 The reconstruction of MC was also performed in two stages using the configura-
 210 tion files reco_uboone_mcc7_driver_stage1.fcl and reco_uboone_mcc7_driver_stage2.fcl.
 211 The data and MC fcl files are almost identical but differ in the following three
 212 points:

- 213 • signal shaping
- 214 • Hit finder threshold
- 215 • Calorimetry constants

216 These parameters have been found to differ in data and MC for comparable
 217 results and tuning has been discussed in the reconstruction group. Parameters
 218 used can be found in the above mentioned fcl files in uboonecode tagged release
 219 v05.08.00.

220 The processing time for a data event is 224 sec/event for the reco1 stage,
 221 and 104 sec/event for the reco 2 stage.

222 Analysis Tree files for MC were produced using the configuration file stan-
 223 dard_ana_uboone.fcl.

224 2.2 Good run list

225 The good run list is assembled based on different criteria:

- 226 • **Detector conditions:** the detector must be in good operating state
- 227 • **Data quality:** basic reconstruction quantities show normal and stable
 228 behavior
- 229 • **Beam conditions:** the BNB must be on/stable, and POT per spill must
 230 be above an intensity threshold
- 231 • **Run processed:** the full run must have been processed successfully, with
 232 no missing subruns due to crashes in the data processing

238 Each of these requirements are provided on a per-run basis except the beam
239 conditions, which are applied on a per-event basis. These criteria and references
240 to additional material are given in the subsections below. A general overview
241 can be found in DocDB 5826.

242 **2.2.1 Detector conditions**

243 Detector conditions are read from the slow monitoring database and are required
244 to be within alarm thresholds. Variables used to evaluate whether or not data
245 should be used for physics analysis are the state of the DAQ, PMT HV, Drift
246 HV, wire bias, electron lifetime, and detector power. This determination is
247 made on a run-by-run basis, and if any of the criteria are not within established
248 thresholds, the run will not pass the selection. Future good future classification
249 will be done at the sub-run level. However, with the additional requirements on
250 data quality provided by looking at high-level reconstruction objects, we expect
251 the existing selection to be sufficient for this analysis.

252 Since the beginning of data taking in the fall of 2015 73% percent of all runs
253 configured to take physics quality data pass the detector conditions criteria.
254 These runs, however, comprise 86% of the total running time as many short
255 runs do not pass. These numbers rise to 88% and 98% since December, which
256 excludes the early data taking period when detector conditions were less stable.
257 A complete list of good runs can be found here: [http://www.microboone.](http://www.microboone.fnal.gov/at_work/AnalysisTools/goodruns/)
258 be found in DocDB 5740.

260 **2.2.2 Data quality**

261 High level stability of various reconstructed parameters has been monitored
262 since the beginning of detector operation (October 2015), and detailed mon-
263 itoring using the current “production” reconstruction software has been done
264 since the start of the operation of the software trigger (February 2016). All
265 data used in this analysis has been taken after the implementation of the tight
266 software trigger. The period is February 23, 2016, to May 22, 2016. A detailed
267 discussion of comparison plots for different reconstruction algorithms, data and
268 MC, BNB and external BNB triggered data covering this run range (5114-6007)
269 is presented in a separate note (uB-doc-5850). This note also describes the gen-
270 eration of a stable run list based on stability of key reconstructed variables, like
271 the average number of tracks, hits, and flashes in each event, the average length
272 of tracks, the average amplitude and area of hits, the average flash PE, and the
273 average spread of these quantities.

274 Application of these quality requirements lead to a selection of about 55%
275 of runs. We find that almost all (more than 99%) of runs that pass the data
276 quality requirements also pass the detector conditions criteria. A number of
277 runs are removed based on the demands of stable hit amplitude and integrated
278 area, which is dominated by some changes in the electron lifetime during the run
279 period. Removing conditions on these variables can increase the run selection
280 to 70% of runs. We continue to investigate the suitability of such a relaxation.

281 **2.2.3 Beam conditions**

282 Beam quality conditions are applied through a per-event art filter module that
283 checks the fraction of proton beam interacting with the target, the horn current,
284 and the intensity of protons per spill. Details can be found in DocDB 5719.
285 The filter will be run on events from the on-beam stream. A final sample
286 with $\sim 5 \times 10^{19}$ POT is anticipated, using a minimum per-spill intensity of
287 4×10^{12} POT.

288 **2.2.4 Run processed**

289 To ease the calculation of total POT, we would like to know that all events
290 (and therefore all subruns) in a run were successfully processed (that is, with-
291 out crashes in data processing). A list of successful subrun ranges has been
292 provided by the production team and can be found here: /uboone/app/user-
293 s/uboonepro/pubs-devel/dstream_prod/subruns3.txt. This is a criteria that can
294 be removed in the future if POT calculation is adjusted to being done per sub-
295 run instead of per run. However, since we have sufficient data for this study
296 using only complete runs, we are using this requirement.

297 **2.3 Final samples**

298 The good run samples using the criteria described in sections 2.2.1, 2.2.2, and
299 2.2.4 have been used to define a SAM data set. Using the BNB POT information
300 in table attached to the DocDB entry of this note (DocDB), the BNB data set
301 is including runs up to a total POT of 5×10^{19} . The SAM data set definitions
302 are

- 303 • BNB: prod_bnb_reco_neutrino2016_beamfilter_goodruns_v5
- 304 • EXT BNB: prod_extbnb_reco_neutrino2016_goodruns_v5

305 for the reco2 stage, and

- 306 • BNB: prod_bnb_anatree_neutrino2016_beamfilter_goodruns_v5
- 307 • EXT BNB: prod_extbnb_anatree_neutrino2016_goodruns_v5

308 for the Analysis Tree stage.

309 Merged Analysis Tree files can be found here:

- 310 • BNB: /pnfs/uboone/persistent/users/aschu/devel/v05_11_01/hadd/GOODBNBEM
- 311 • EXT BNB: /pnfs/uboone/persistent/users/aschu/devel/v05_11_01/hadd/GODEXTBNB

312 The number of total events in the good run sample is

- 313 • BNB: 546910
- 314 • EXT BNB: 388471

315 The beam filter described in section 2.2.3 has been finalized and is being
316 tested and placed into production. The filter will run over the existing recon-
317 structed data samples to remove events with bad beam quality or too little POT
318 intensity. After this, a slightly modified data set will be defined and we will ver-
319 ify that this data set contains an integrated POT of approximately 5×10^{19} .

320 **2.4 Normalization of off-beam to on-beam data**

321 The off-beam sample is used to measure beam unrelated backgrounds. To nor-
322 malize properly the measured backgrounds to the on-beam sample one needs
323 to know the number of total BNB spills (N_{BNB}) used for the analysis and the
324 total number of external triggers (N_{EXT}) used to obtain the off-beam sample.
325 More precisely one needs the number of triggers before the software trigger. In
326 case of the beam triggers only the ones that pass beam quality cuts count as
327 only those would be used in the analysis. The normalization factor is then given
328 with the ratio of the two N_{BNB}/N_{EXT} .

329 The final normalization constants for the sample defined in 2.3 are:

- 330 • $N_{BNB} = 11,045,346$
331 • $N_{EXT} = 8,979,753$

332 Details are reported here [1]. This means that off-beam data has to be scaled
333 with a factor $11,045,346/8,979,753 = 1.2300$.

334 Please note that in this study, no absolute normalization between data and
335 MC is attempted.

336 **2.5 Normalization of MC samples**

337 The following description is based on DocDB #5640.

338 **Normalization of BNB MC generated events to POT** The POT for
339 BNB events are stored per subrun. For a given sample of BNB events (can be
340 BNB only or BNB+Cosmic) a relation can be found between generated events
341 (no cuts applied, not even on activity inside the TPC), and the summed POT.
342 For MCC7, which has a large fraction ($\geq 50\%$) of interactions outside of the
343 TPC active volume), this relation is:

$$5 \times 10^{19} \text{ POT corresponds to } 41524.3 \text{ generated events.} \quad (1)$$

344 Note: this scaling factor is valid for MCC7 and might not hold for other pro-
345 ductions.

346 **Normalization of generated open cosmics MC to BNB MC** The open
347 cosmics sample is a sample where cosmics come at random times (like in reality)
348 and might not necessarily arrive during the beam window and therefore might
349 not trigger a readout. This sample is not used in the analysis, but since the
350 inTime sample, which is being used, is normalized relative to an open cosmic
351 sample, its normalization is described here. The POT per beam spill is a con-
352 stant number in the MCC7 BNB samples, it is $5e12$. Therefor, we can calculate
353 how many spills are required to produce a certain number of POT: It is total
354 POT times the average of $1/\text{POT}$ per spill, so in our case $5e19/5e12 = 1e7$
355 beam spills are required to produce 41524.3 events with a neutrino interaction
356 inside the cryostat volume. This is only 1 in 241 events. Therefore, each cosmic
357 only event needs to have a weight of 240.8 when comparing to BNB MC.

358 **Normalization of inTime cosmics MC to open cosmics MC** The COR-
359 SIKA samples are the most computing intense, because we need a huge amount
360 of generated events to have sufficient statistics at the analysis level. Therefore,
361 When producing inTime cosmics, two filters are applied:

- 362 • Gen stage: GenInTimeFilter
363 • Detsim stage: FilterSimPhotonTime

364 The intention of these filters is to reduce the number of events that will not
365 end up in the analysis at a very early stage in the simulation process, because
366 we want to save both simulation time and disk space. The first filter running
367 at gen stage selects events which have at least a particle with 0.1 GeV primary
368 energy crossing the TPC boundaries. This is to make sure that there is at least
369 one cosmic particle pointing towards the TPC, because the generation surface
370 for cosmics is way larger than the actual TPC and some events don't have any
371 activity of interest for us. The second filter running at detsim stage selects
372 events which do have scintillation light within the time window of 3125 to 4725
373 ns. This makes sure that the event is a candidate for passing the SW trigger.
374 Events that only contain out-of-beam-window cosmic activity are rejected by
375 this filter. The passing rate for both filters applied is 0.02125. Combining this
376 with information from open cosmics scaling, the per event weight for an inTime
377 cosmics when comparing to BNB MC in MCC7 is $240.8 \times 0.02125 = 5.12$.

378

3 Optical triggering and reconstruction

379

3.1 Software trigger

380 The vast majority of BNB spills recorded by MicroBooNE do not contain a
381 neutrino interaction. Triggering on all BNB spills received from the accelerator
382 result in very large data-rates, which would be very demanding of MicroBooNE's
383 computing resources, and produce a data-set mostly composed of events with
384 no neutrino interactions. The most effective way to quickly determine if a BNB
385 spill possibly produced a neutrino interaction in the MicroBooNE detector is to
386 require the presence of a burst of light activity on the light-collection system in
387 time-coincidence with the $1.6 \mu\text{s}$ beam-spill reaching the detector. By requiring
388 the presence of light coincident with the beam-spill, the vast majority of triggers
389 for which no neutrino interacts in the detector can be eliminated. This however
390 does not guarantee the presence of a neutrino interaction, as the same burst of
391 light could be produced by a coincident cosmic ray, or other background source.

392 In order to implement such a filter, the collaboration chose to adopt a "software-
393 trigger", which uses waveforms recorded by PMT system to decide, after data
394 for the event has been sent from the readout crates to the DAQ machines,
395 whether to keep the event or not. The software-trigger is implemented after
396 the event-builder combines data from the PMT crate and trigger fragments
397 into a single event. The algorithm adopted by the software-trigger was written
398 to model, and reproduce exactly, the results obtained by the hardware-filter
399 originally conceived to be implemented within the PMT readout via a FPGA
400 implementation. The FPGA implementation was tested and debugged on the
401 emulator and exactly reproduced and validated in the offline software.

402 The algorithm implemented within the software trigger uses the digitized
403 output of all 32 high-gain PMT channels. Only the waveform region in time
404 with the beam-spill is used to search for possible triggers. Per PMT, a waveform
405 obtained by taking the difference between the ADC value at tick t , and tick $t+s$
406 is calculated. This waveform is scanned for ADC values which are above some
407 settable threshold (defined as X_0). Once this condition is met, a discriminator
408 window (disc0) is opened for this PMT channel, and lasts for a fixed number of
409 time-ticks W_0 . Within this region of W_0 ticks, if the ADC count goes above a
410 second, larger threshold (X_3) a window of width W_3 is opened. The maximum
411 amplitude within this window is saved as the peak-amplitude, for a given PMT,
412 in the time-window defined by disc3. The amplitude of coincident disc3 windows
413 across all PMTs is summed together into a value called PHMAX. The software-
414 trigger places a final cut on this PHMAX value to determine whether to keep
415 the event or not.

416 Software trigger firing rates, as a function of the PHMAX threshold (in ADCs)
417 / 20 can be seen, for the BNB and NuMI streams respectively, in Fig. 1.

418 The chosen threshold applied on the BNB stream was set to 6.5 effective
419 PE, or a cut of 130 ADC on the PHMAX value calculated by the software-
420 trigger algorithm. A summary presenting the software-trigger configuration

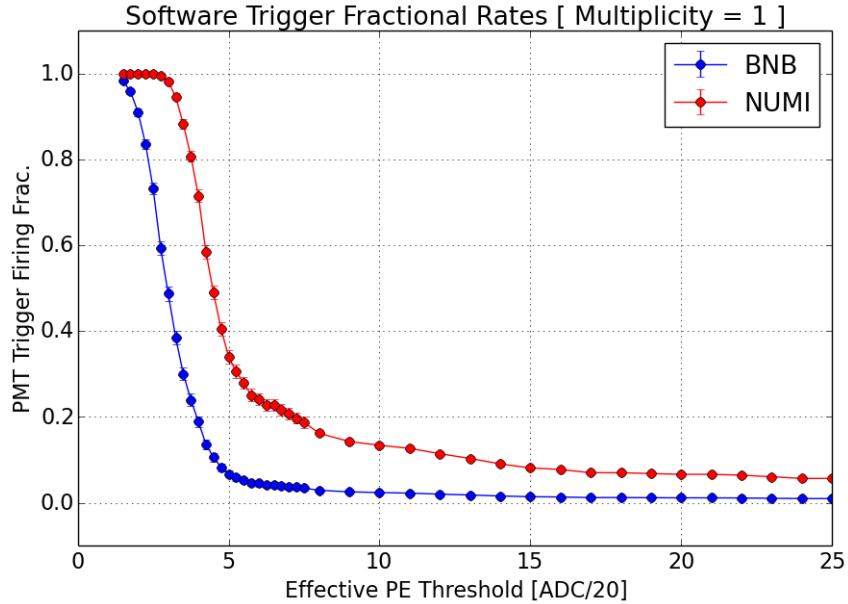


Figure 1: Fraction of events passing the software trigger as a function of the threshold applied on the PHMAX value / 20.

421 currently being applied can be found in DocDB 5403. Further documentation
 422 describing the software-trigger and its implementation, can be found in
 423 the summary prepared by the Trigger Task-Force, in DocDB 5205. This in-
 424 cludes preliminary Monte Carlo studies showing the effect of a software-trigger
 425 on the triggering efficiency for single particle, as well as neutrino samples.
 426 The configuration settings used for the BNB data-stream are documented in
 427 the appendix. The config-ID used by the software trigger is Config ID 443,
 428 PHYSICS_BNB_NUMI_MUCS_EXT3Hz_SWTrigger_c0. The specific values set for (X_0 ,
 429 X_3 W_0 , W_3) are, listed in the same order (5 ADC, 10 ADC, 6 Ticks, 6 Ticks).

430 3.2 Flash reconstruction

431 Light is collected on MicroBooNE’s 32 PMTs either in a continuous, un-
 432 biased readout window of $23.4 \mu\text{s}$ which is opened by the beam-gate signal
 433 received on the trigger-board, or in discriminated pulses (referred to as cosmic-
 434 discriminator) each $\sim 1 \mu\text{s}$ in duration activated if, for any given tube, the ADC
 435 count goes above an 80 ADC (roughly 4 PE) count. These two different formats
 436 are saved in an event as output waveforms. Each PMT provides two output
 437 streams, one for the high-gain channel ($\sim 20 \text{ ADC/PE}$) and one for the low-gain
 438 ($\sim 2 \text{ ADC/PE}$) channel. The first step performed in the optical reconstruction is
 439 the merging of these two streams into a “saturation-corrected” waveform which
 440 tries to correct saturating high-gain pulses by using information from the low-
 441 gain channel. Documentation for how high and low-gain information is merged
 442 can be found in DocDB 4979.

443 The saturation-corrected beam-gate unbiased readout window is used to re-
444 construct optical hits. Each PMT's waveform is scanned for hits which are lo-
445 cated by first applying a rolling-baseline estimation, in order to be less sensitive
446 to possible fluctuations in the signal baseline. At this stage a threshold-based
447 hit-reconstruction algorithm is applied, requiring pulses of a minimum area in
448 order for a hit to be reconstructed. Each reconstructed hit is associated with a
449 PMT number, a time (with respect to the trigger time, in μs), and a PE count.
450 Details on the optical reconstruction tools in use can be found in DocDB 5218

451 Once, for a given event, optical hits are reconstructed on all 32 PMTs, in-
452 formation from all PMTs is combined into optical flashes, which are meant to
453 represent interactions in the detector, as seen by the PMT system. Each re-
454 constructed flash contains information about the total amount of light seen in
455 the interaction (in PEs) as well as the distribution of light across all 32 PMTs.
456 The flash-time, with respect to the trigger time (in μs), as well as spatial in-
457 formation in the Y-Z plane of the detector pertaining to the flash. Flashes
458 are reconstructed by requiring a time-coincidence of $\sim 1\mu\text{s}$ between hits on all
459 PMTs. The total PE measured across all time-coincident hits is summed to-
460 gether and flashes are reconstructed if the amount of light is larger than 2 PE.
461 A safeguard is placed to avoid reconstructing pulses from late scintillation light
462 as independent flashes. Given a time-coincident group of optical hits, subse-
463 quent potential flashes are reconstructed as such only if their total PE value is
464 more than 3σ above the expected amount of light produced by late-scintillation
465 photons given the total light in the early pulse, a $1.6\mu\text{s}$ late scintillation-light
466 decay-time, and a 4:1 ratio in late-to-prompt light. If this condition is not met,
467 the subsequent photons are recognized as late-light from the preceding flash,
468 and not reconstructed as an independent flash. A thorough description of the
469 flash-finding algorithm is provided in DocDB 3735.

470 We conclude this section on flash-reconstruction by showing the time-distribution
471 of optical flashes reconstructed using BNB unbiased-triggered events in which
472 a clear excess in coincidence with the expected arrival time of neutrinos is ex-
473 pected (Fig. 2). The information shown in this plot was produced with the
474 same flash-reconstruction tools used in this note, and serves as validation for
475 this reconstruction tool. The document describing how this plot was produced
476 can be found in DocDB 5390

477 3.3 Beam window

478 The distribution of flashes for on-beam and off-beam data events and MC sam-
479 ples is shown in Fig. 3. Note that the software trigger has been applied to both
480 data streams. The application of the software trigger causes the characterisitic
481 shape of the time distribution of flashes: The pile-up of flashes just after $0\mu\text{s}$
482 is a feature caused by a flash finding at its veto effect with different settings
483 for the window from 0 to $23.4\mu\text{s}$. The pile-up consists of low-PE flashes and
484 is removed when placing a low threshold (see the plot on the right which has a
485 20 PE cut applied). It is not understood at this point in time why this feature
486 is missing in the CORSIKA inTime sample. Since we further only care about
487 flashes inside the beam window and since the CORSIKA inTime sample is only

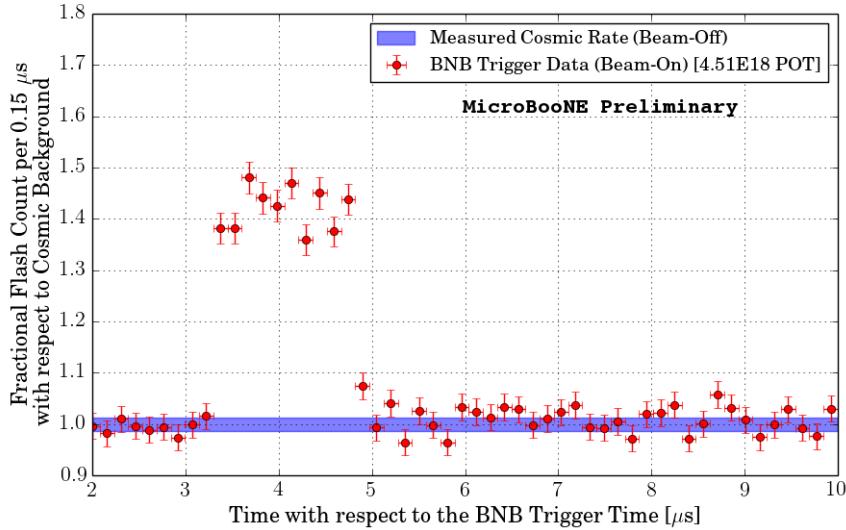


Figure 2: Time-distribution of reconstructed optical flashes with a PE value of 50 or more, for a sample of unbiased BNB triggered events. A clear excess in coincidence with the BNB spill is observable.

488 used for purity estimation but not for any final event distributions of we are not
 489 following up on this any further here.

490 The plot shows that the time windows are somewhat shifted between the
 491 samples. This is caused by different hardware configurations and it is not a
 492 problem, i.e. doesn't cause us to lose any neutrinos. However, the shift needs
 493 to be taken into account when selecting neutrino candidate flashes for the event
 494 selection. The CORSIKA inTime and the other MC samples have slightly dif-
 495 ferent beam windows. The reason for this difference is not understood, but we
 496 are deciding here to define the beam window size empirically. The windows that
 497 were chosen are

- 498 • On-beam: 3.3 to 4.9 μ s
- 499 • Off-beam: 3.65 to 5.25 μ s
- 500 • CORSIKA inTime: 3.2 to 4.8 μ s
- 501 • BNB only: 3.55 to 5.15 μ s
- 502 • BNB+Cosmic: 3.55 to 5.15 μ s

503 It is important that all windows have the same duration of 1.6 μ s.

504 3.4 Analysis flash threshold

505 The software trigger has been introduced as a trigger decision based on the light
 506 information. It is running on the data stream, but it is not calculated in MC.
 507 It is critical to map out the correlation of the software trigger threshold value

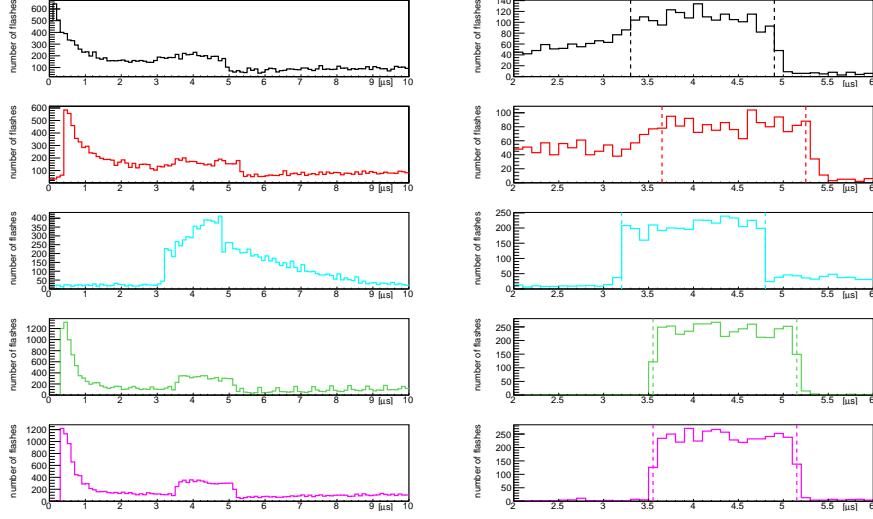


Figure 3: Flash time distribution for all flashes (left plot) and flashes $> 20\text{PE}$ (right plot). The different curves are (from top to bottom): on-beam data (black), off-beam data (red), CORSIKA inTime MC (light blue), BNB only MC (green), and BNB+Cosmic MC (purple). The dashed vertical lines mark the time window that is used for each sample

508 and the analysis cut on light information which is used in ν_μ CC study. The
509 two questions that need to be addressed are

- 510 1. Whether the analysis cut is higher or lower than the software trigger
511 threshold
- 512 2. Whether there are losses of ν_μ CC events due to application of the software
513 trigger in data.

514 Unbiased BNB data from Feb 21st to April 13rd are collected. There are
515 36059 such events which corresponds to the expectation based on the scaled
516 software trigger event rate ($\approx 50\text{day} \times 24\text{hours} \times 3600\text{second} \times 0.01\text{Hz} = 43200$). In
517 this study, the flash information is inherited from Analysis Tree files with flash
518 reconstruction module opflashSat and the software trigger data is calculated
519 from swizzled data using an extraction module written by T. Wongjirad.

520 The correlation between flash and software trigger is shown in Fig4. Flash.PE
521 on X-axis is the max flash PE found in the software trigger window (190-310
522 PMT ticks w.r.t T_0). On Y-axis is the PHMAX value formed in that event and
523 when it is above 130 ADC (Tight), a software trigger is fired. Since we used the
524 unbiased BNB data for analysis, most of events have small Flash.PE values and
525 corresponding small PHMAX values. However, the pattern is clear that approx-
526 imately 1 FLASH.PE returns a PHMAX value of 20. In this case, the analysis
527 cut of 50 PE (1000 ADC) is much higher than software trigger threshold value
528 (130 ADC for tight trigger).

529 On the other hand, we do have events with a max flash PE above 50 PE in the
530 trigger window but a PHMAX value lower than 130 ADC in 5. These events will
531 be lost for the analysis due to the application of the software trigger. However,

the effect is rather small. The lost event count due to the software trigger is $5/458 < 1.1\%$. These are events happening late in the software trigger window (an example PMT waveform, Figure 6). In this case the PMT waveform has not risen high enough and therefore only a fraction of discriminators are included in PHMAX calculation that then turns out too small to pass the threshold.

The code/module of software trigger special application for offline data is developed by Taritree and lives in github. Check the repository for more details: github.com/twongjirad/fememulator

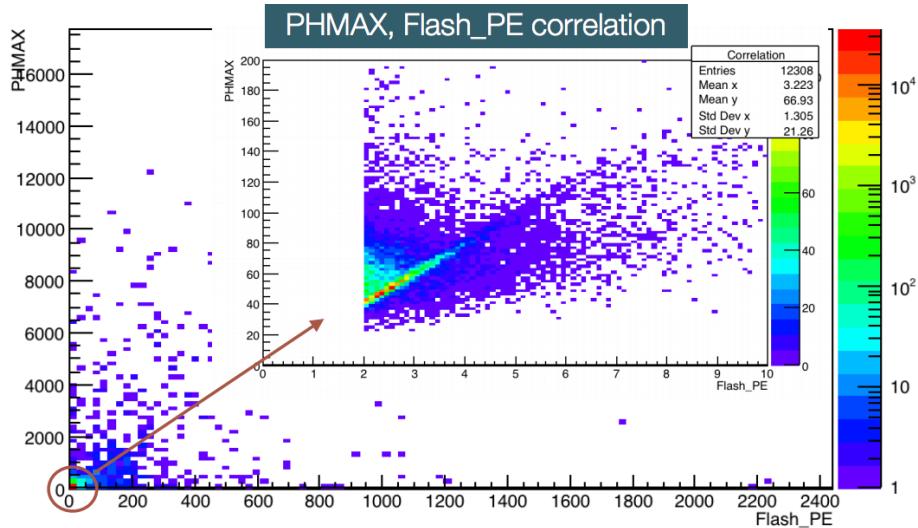


Figure 4: Correlation of software trigger and maximum flash found in trigger window. Hard cut on flash >0 is involved.

3.5 Impact of shortcomings in the optical simulation on this analysis

The previous section has shown that a threshold of 50 PE for a flash in the beam window is a sophisticated choice for the analysis. In particular it was shown, that the bias through the application of the Software Trigger is negligible. In MC, no software trigger is applied, but the same analysis threshold cut of 50 PE is chosen. However, previous studies have shown that a 50 PE flash in data does not correspond to the same value in MC (see DocDB 5722). We currently think that a flash of 50 PE in data corresponds to a flash of a factor 1.5 to 2 higher in simulation. This is not just a normalization offset, but is expected to have mostly unknown shape dependence. The subject needs further investigation and won't be fixed before the NEUTRINO 2016 conference.

In order to estimate the impact on the final event distributions of ν_μ CC inclusive events, we plot final event distributions as a function of muon track length, muon angle θ and muon angle ϕ with varied cut values for the PE cut. This study has only been done for Selection I (Original). The details of this event selection are described later in this note. Figures 7, 8 and 9 show the distributions. These show, that a normalization and shape dependent bias is to

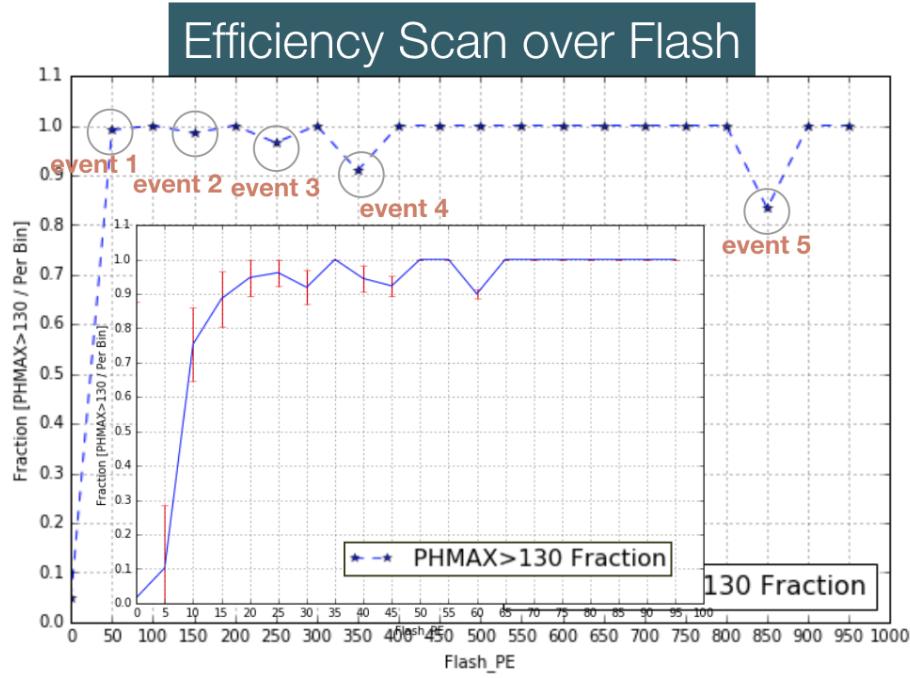


Figure 5: Fraction(events with $\text{PHMAX} > 130$ ADC / events per bin) scanning over Flash_PE. Plot is also zoomed into Flash_PE between [0,100]PE. The events that didn't pass the software trigger have been labeled and studied individually. However, this is not discussed here.

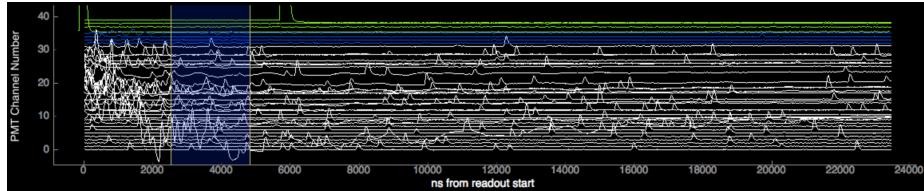


Figure 6: All 32 PMT waveforms for an event that has a flash with $\text{PE} > 50$ but $\text{PHMAX} < 130$.

be expected with the current optical simulation. This bias is at the few percent level, and it is not further taken into account in the distributions part of this note. The plan is to present the MC distributions as they are and labeled as "uBooNE baseline MC", with a statement that systematics (besides the ones explicitly discussed) are not addressed.

Note that although this study has only been done for one of the selections presented later, we expect the impact to be comparable for all selections. We are not making a quantitative statement on this anywhere later.

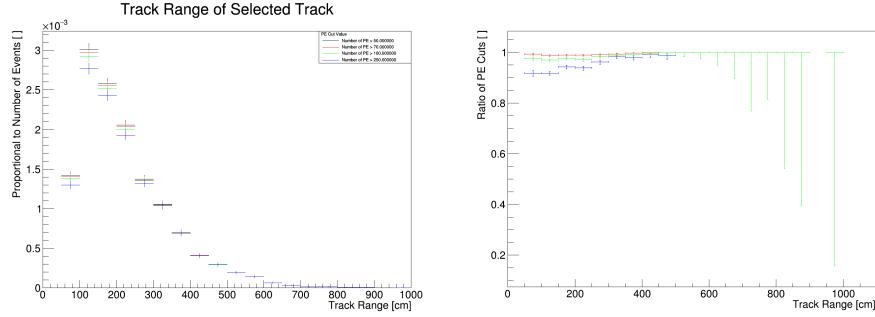


Figure 7: Distribution of track length for all simulated BNB+Cosmic events passing **Selection I (Original)**. The black curve shows the standard 50 PE plot, the red curve a 70 PE cut, the green curve a 100 PE cut, and the blue curve a 250 PE cut. Ratios are taken wrt the black curve and the other curves keep the color coding.

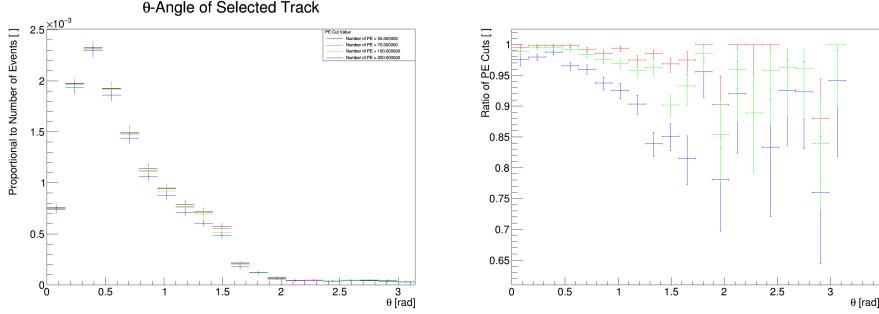


Figure 8: Distribution of muon angle θ for all simulated BNB+Cosmic events passing **Selection I (Original)**. The black curve shows the standard 50 PE plot, the red curve a 70 PE cut, the green curve a 100 PE cut, and the blue curve a 250 PE cut. Ratios are taken wrt the black curve and the other curves keep the color coding.

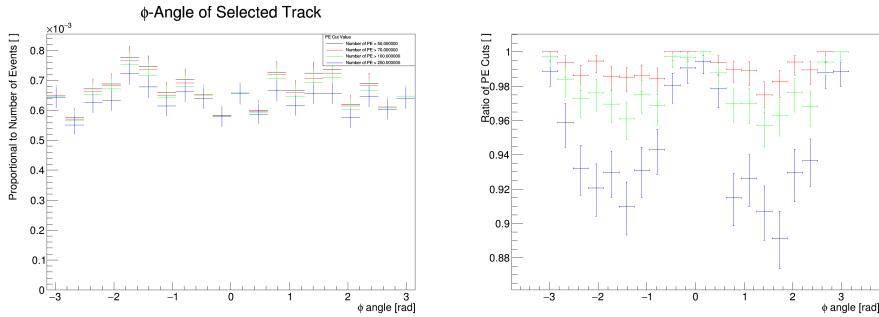


Figure 9: Distribution of muon angle ϕ for all simulated BNB+Cosmic events passing **Selection I (Original)**. The black curve shows the standard 50 PE plot, the red curve a 70 PE cut, the green curve a 100 PE cut, and the blue curve a 250 PE cut. Ratios are taken wrt the black curve and the other curves keep the color coding.

566 4 TPC reconstruction

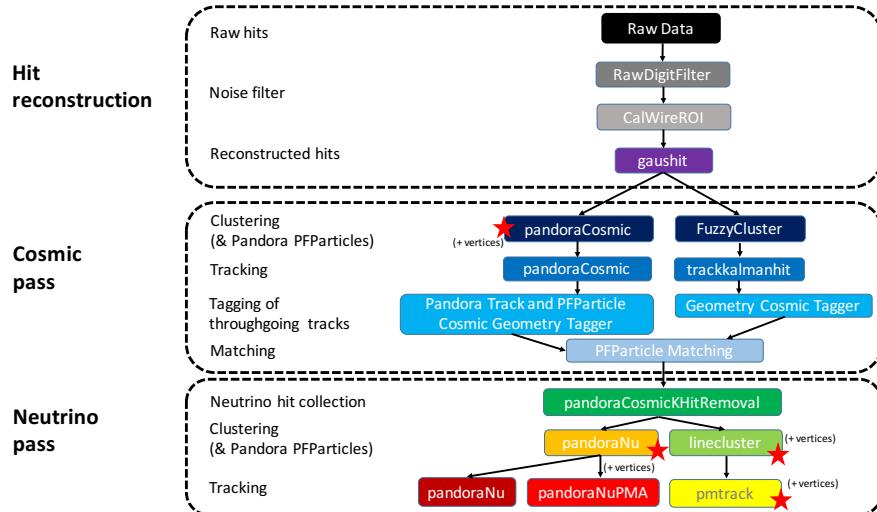


Figure 10: Reconstruction chain run for data and MC processing. The red stars on some of the boxes mean that the algorithms returns reconstructed 3D vertices.

567 Figure 10 summarizes the reconstruction chain applied to data and MC in
 568 this analysis. The hit reconstruction level is followed by a cosmic pass. Cosmic
 569 pass means that all objects are attempted to be reconstructed. All hits asso-
 570 ciated to through-going tracks in the cosmic pass, are then removed, and the
 571 neutrino reconstruction only runs on the remaining hits.
 572

573 4.1 Hit reconstruction

574 The input data to the reconstruction consists of waveforms in drift time of
 575 charged induced or deposited on the sense wires. The first step in the hit
 576 reconstruction is to pass these waveforms through a filtering algorithm which
 577 is designed to remove or reduce to the extent possible the noise introduced on
 578 to these waveforms by the electronics and through pick up on the sense wires
 579 themselves. As the input waveforms represent the largest single volume of data
 580 in the event, this filter is also used to reduce the length of the waveforms from
 581 their input 9600 ticks to 6400 ticks.

582 After noise filtering the waveforms are passed to the deconvolution algorithm
 583 which aims to “remove” the effects of the drift field and electronics response in
 584 order to get a measurement of the number of ionized electrons created by the
 585 incident track. At the same time “Regions of Interest” (ROI) are identified and
 586 “snipped” out of the waveforms in order to further reduce the overall output
 587 data volume.

588 The ROI’s are then passed to the hit finding algorithm which has as its goal
 589 to identify candidate peaks in the waveforms and fit them with a Gaussian shape
 590 in order to obtain a “hit” representing the charged deposited on that wire by

591 the incident track. Hits are represented as objects with a peak time and width
592 and serve as the basic input to the downstream reconstruction algorithms.

593 4.2 Clustering

594 The aim of the fuzzy cluster algorithm is to associate hits, in 2D, which cor-
595 respond to physics objects - tracks, showers, etc. To achieve this, a three step
596 process is used. The first step associates hits into common groups using a fuzzy
597 clustering algorithm - where hits are given a “degree of belonging” to the clus-
598 ters they are associated to. The second step then employs a Hough transform
599 to find hits associated to candidate tracks and/or showers within each of the
600 clusters returned in the first step. Finally, the last step is to merge common
601 candidate tracks and showers into larger clusters. This step also associates
602 unclustered hits that are nearby (important for shower reconstruction) to the
603 larger clusters. The resulting output is a set of clusters of hits representing
604 either candidate tracks or showers separately for each plane.

605 At the current time the trackkalmanhit algorithm does not make full use of
606 the output of the fuzzy cluster algorithm, it currently uses it as a “hit filter”
607 where only hits associated to clusters are used by the Kalman filter tracking
608 algorithm for finding and fitting tracks.

609 The pandora pass has its own clustering algorithm, which is described in
610 the pandora tech note (see section 4.3). The other clustering algorithm used is
611 linecluster. The linecluster algorithm reconstructs 2D line-like clusters in each
612 wire plane. Hits on nearby wires are fitted to a line, which is then extrapolated
613 to neighboring wires. Hits on these wires are attached to the leading edge of the
614 cluster if the hit position and charge are consistent with the previously added
615 hits. 2D vertices are formed in each plane by the intersection points near the
616 ends of clusters. The 2D vertices are matched between all three planes to create
617 3D vertices in space. The position of partially matched 2D vertices is used to
618 refine the reconstruction of clusters in the third plane.

619 4.3 Pandora

620 The pandora reconstruction algorithms are described in a separate note. The
621 DocDB reference for this note is DocDB 5828.

622 The pandoraCosmic and pandoraNu reconstruction algorithms are differ-
623 ent. pandoraCosmic is aimed at reconstructing long downward-going muon-like
624 tracks. pandoraNu is aimed at reconstructing neutrino interactions.

625 Both algorithms start at the clustering level and return a so-called PFPar-
626 ticle hierarchy. In a PFParticle hierarchy, reconstructed objects (called PFPar-
627 ticle) are linked to their daughters and parents. PFParticles can be track or
628 shower-like, and do have a vertex. For tracks, the pandora-internal track fit-
629 ting is available to be used as track reconstruction output and in downstream
630 analysis. Vertices are associated with each PFParticle. Pandora also assigns
631 the primary neutrino vertex to the primary of the PFParticle Hierarchy. For
632 downstream analyses, this primary vertex can be used. Alternatively, we can
633 use all vertex candidates from pandora and chose the primary vertex using our
634 own selection.

635 **4.4 trackkalmanhit**

636 The trackkalmanhit algorithm takes 2D hits as input and returns track objects as
637 output. In our case, it is using all hits from all 2D clusters returned by Fuzzy-
638 Cluster as an inputs. The track output is not sorted in a hierarchy structure as
639 it is in pandora. All tracks are independent from each other. There is no vertex
640 reconstruction associated with it. A detailed documentation on trackkalmanhit
641 can be found in DocDB 3516.

642 **4.5 Cosmic hit removal**

643 As discussed in paragraph 4.3, there are two pandora “passes” employed in
644 the event reconstruction, the first, aimed at reconstructing primarily downward
645 going muon-like tracks, runs on all hits reconstructed as described in paragraph
646 4.1 and the second, optimized to reconstruct neutrino interactions, is run on a
647 subset of hits which are not clearly associated to candidate Cosmic Ray tracks.

648 At the end of the pandora first pass the output PFParticle hierarchy is passed
649 to a Cosmic Ray tagging algorithm that looks at all tracks and hits associated
650 to the hierarchy to determine its start and end points. If these start and end
651 points are consistent with a trajectory that both enters and exits the TPC then
652 all hits are marked as candidates for removal from the second pass. For the
653 purposes of this algorithm, a trajectory is considered to enter or exit the TPC
654 if it contains hits which have drift times outside the neutrino drift window or
655 track end points which are outside of a fiducial volume contained within the
656 active volume of the TPC (20 cm from the top or bottom of the TPC, 10 cm
657 from the TPC ends). The fiducial volume was determined based on a study
658 presented in DocDB 5423. However, in the Cosmic Hit Removal the fiducial
659 volume cut in drift direction was released and the border is 0 cm. There are
660 additional containment cuts later in the event selection applied to tracks and/or
661 vertices with cut values described in the chapter in event selection.

662 An additional algorithm also takes the output of the trackkalmanhit stage
663 described in paragraph 4.4 and matches hits on these tracks to hits in the PF-
664 Particle hierarchies, thereby matching trackkalmanhit tracks to one or more
665 PFParticles. An algorithm similar to the Cosmic Ray tagging algorithm de-
666 scribed in the previous paragraph then makes the same basic comparisons and
667 if a PFParticle hierarchy is considered to be both entering and exiting then all
668 hits associated to the hierarchy are also marked as candidates for removal.

669 All hits which are candidates for removal are checked to see if they could
670 be shared with PFParticle hierarchies which have been deemed to be contained
671 and if so removed from the removal candidate list. At the completion of this
672 check all remaining candidate hits are removed from the input hit collection
673 and the remaining hits passed to the neutrino optimized pass of the Pandora
674 reconstruction.

675 **4.6 Projection Matching Algorithm**

676 The projection matching algorithm (PMA) has been inherited from ICARUS,
677 and is implemented in LArSoft. While most other 3D reconstruction algorithms
678 are attempting to match the 2D objects in all three planes by the drift time,
679 this algorithm uses a different approach. During the fit, the track hypothesis is

680 projected on each plane and distance between the hits and the projections in
681 all three planes are being minimized simultaneously. Details on the algorithm
682 are published in [M. Antonello et al., Adv. High Energy Phys. (2013) 260820;
683 <https://inspirehep.net/search?p=find+eprint+1210.5089>].

684 4.7 Calorimetry

685 On all track reconstruction output objects, calorimetry is being run (not in the
686 flow chart). A general introduction to the MicroBooNE calorimetry module
687 can be found in DocDB 2994. Data driven tuning of calorimetry calibration
688 constants has been run for version v05_08_00 on data and MC separately, using
689 crossing muons (see DocDB 5716).

690 The following calorimetry output is available for downstream analyses:

- 691 • The total kinetic energy of the track in each plane
- 692 • The total track range of the track in each plane
- 693 • The charge loss dq/dx associated with a hit on a track in a plane
- 694 • The energy loss dE/dx associated with a hit on a track in a plane
- 695 • The residual range from the end of track associated with a hit on a track
696 in a plane

697 In the downstream analysis, we choose the plane with most hits as the best plane
698 and use information from this plane. Typically, this is the collection plane.

699 4.8 Tracking efficiency comparison

700 Figure 11 shows the tracking efficiency obtained from matching reconstructed
701 tracks to MC truth tracks using the TrackingEfficiency module provided by
702 DUNE (see DocDB 5533). This module defines the origin of a track by using
703 backtracker to find the MC truth particle that caused the hits that are part of
704 the reconstructed track. The MC truth particle that contributed the majority
705 of the hits in a reconstructed track is defined as the origin. The efficiency is
706 then calculated as

$$\epsilon = \frac{\text{Events w/neutrino induced reco muon track}}{\text{Generated events w/ true neutrino induced muon track}}. \quad (2)$$

707 Only events with a true neutrino interaction vertex inside the fiducial volume
708 are considered in the calculation.

709 Above a momentum threshold of ~ 300 MeV, the tracking efficiency is high
710 ($\geq 80\%$) for all algorithms. Trackkalmanhit and pandoraCosmic show a very
711 good tracking performance for high momentum muon tracks. The tracking
712 efficiency is lower for the algorithms going through the neutrino pass, because
713 tracks tagged as going through crossing the detector fiducial volume (with the
714 fiducial volume being defined by a 10 cm border z, and a 20 cm border in y)
715 are being removed from that pass, which causes (unfortunately) an inefficiency
716 in detecting muons from neutrino interactions. Therefore, trackkalmanhit and
717 pandoraCosmic are both used as input for the cosmic removal.

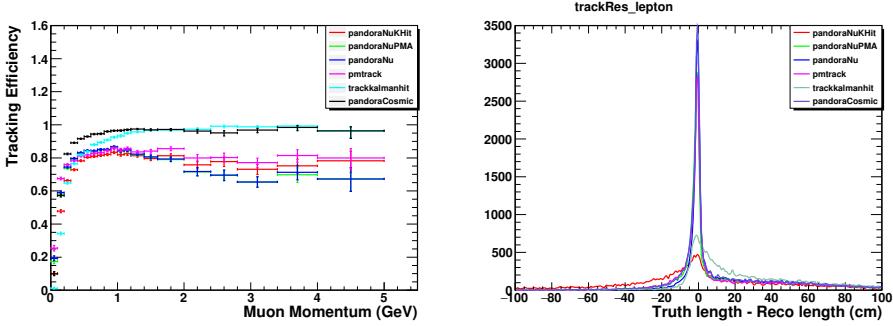


Figure 11: Tracking efficiencies and track length resolution for muons from neutrino interactions in a BNB+Cosmic simulation. No cuts applied.

718 4.9 Treatment of noise and dead wires in the MC and 719 impact on data/MC agreement

720 A variety of different sources and forms of noise on the TPC wires has been
721 observed in the MicroBooNE data. A detailed study characterizing the noise
722 can be found in [DocDB 5854].

723 The noise model used in MC simulation with version v04_36_00 (version used
724 for the detector simulation stage of all MC in this note) however does not reflect
725 all different sources observed in data. It is a simple random noise model, that
726 in particular doesn't account for correlated noise. This is a clear deficiency of
727 current MC simulation and studies are on the way to improve noise modeling
728 in the future.

729 Software noise filtering is applied to both, data and MC, in the same way.
730 The noise filtering, done using the RawDigitFilter module, has selection cuts on
731 particular forms of noise, which – if these are not present in MC because they
732 were not simulated – are not always effective for MC. Some noise sources, e.g.
733 the so called chirping noise is not present nor removed in the MC, other noise
734 sources, e.g. the 37 kHz line removal in the U plane, apply also to MC and this
735 can be seen in an FFT of the waveforms.

736 Since there was no sophisticated noise modeling by the time of the production
737 of these MC samples, there is no quantitative study on how noise affects the
738 reconstruction of high level variables from tracking and vertexing, which are
739 used in the numu CC event selection. These effects will be included in the
740 future and we will carefully study their impact as these features get added to
741 the MC.

742 The data and MC comparisons shown in the internal note [DocDB 5849]
743 show that there is reasonable agreement in particular for tracking variables.
744 Largest disagreement is observed in the calorimetric variables. The largest con-
745 tributors to the calorimetry calibration factors are expected to be the field
746 response function used in deconvolution, the electron lifetime calibration, and
747 the performance of the hit finder. The role of the noise on this should be lim-
748 ited (other than the touching wires, which again modify the local E-field, thus
749 modifying the field response functions). We developed a new procedure of sig-
750 nal processing, and we are in the process of improving the simulation on this
751 [DocDB 5808].

752 The calibration variables are used in selection II for cosmic removal, but they
753 are not used anywhere for calorimetric measurements. Selection II has been
754 tested for robustness towards varying overall scaling of calorimetry calibration
755 constants and has shown that the impact on the passing rate is 1.2%.

756 The detector also has a set of wires which are considered dead, because the
757 wires are either very noisy, the channels on the ASICs are dead or the wires
758 don't seem to be connected to the ASICs. The total number of wires considered
759 dead is 862 [DocDB 5854], which corresponds to about 10% of the total number
760 of wires in the detector. The majority of these are permanently dead. About
761 10% show intermittent behavior, which means varying amount of chirpingness
762 observed over time.

763 The MC simulation uses a static dead channel list, and channels on this list
764 are treated as non-existing during the reconstruction. This is done in both, data
765 as well as MC reconstruction. One problem that arises from dead wires in the Y
766 plane is that this forces the V plane wires in front of them to act as a collection
767 plane, while the hit reconstruction still treats these as induction plane wires.
768 This is treated incorrectly in the MC.

769 It has been observed that dead wires in the detector lead to so called bro-
770 ken tracks, which arises from clusters not being merged across dead wire gaps.
771 Therefore, tracks are often reconstructed as too short, since part of them get
772 missing in a dead wire region.

773 Both reconstruction algorithms used here, pmtrack and pandora, have been
774 taking into account the problem of dead wires planes, and therefore only require
775 two functioning planes instead of three in order to reconstruct 3D objects such
776 as tracks and vertices. The pandora algorithms even have been modified to take
777 into account the fact if the third plane has a dead region in the questionable
778 range when assembling tracks from 2D cluster objects of two planes. This has
779 increased the efficiency for long tracks [DocDB 5828].

780 The data/MC comparisons [DocDB 5849] suggest that we achieve a similar
781 treatment of dead wires in data and simulation. However, at this point in time
782 there is no reconstruction performance study that isolates the effect of dead
783 wires on e.g. tracking efficiency or track length resolution, but we are currently
784 working on producing a MC sample with dead wires turned back alive in order to
785 provide some material. There is a study on the impact for shower reconstruction
786 [DocDB 5858].

787 Figure 12 shows the impact of having $\sim 10\%$ dead wires on the reconstruction
788 efficiency of muon tracks from neutrino interactions. You can see that for both
789 algorithms shown (trackkalmanhit and pandoraCosmic) there is a small decrease
790 in tracking efficiency caused by broken wires. This decrease gets larger for
791 lower muon momentum. Note, that this definition of efficiency does not take
792 into account if the track length is reconstructed correctly. It only checks for
793 a reconstructed track matched to the true muon track. This is an effect that
794 needs to be studied separately.

795 5 Data-MC comparisons

796 Data and MC comparisons have been factorized into a separate note that is cur-
797 rently being reviewed in the reconstruction group. The note contains off-beam
798 data and inTime CORSIKA comparisons for track reconstructions, vertex re-

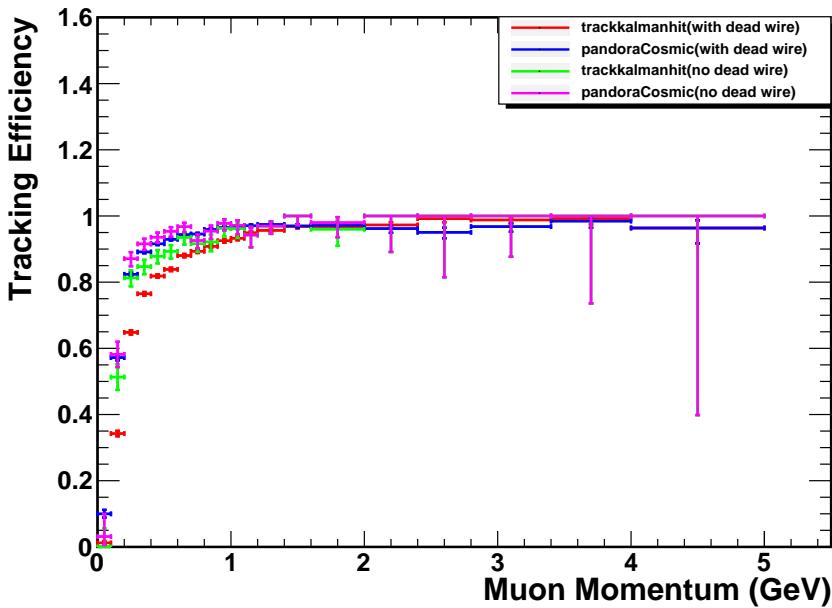


Figure 12: Tracking for muons from neutrino interactions in a BNB+Cosmic simulation for two different algorithms (trackkalmanhit and pandoraCosmic). The comparison shows reconstruction with dead wires implemented in the simulation as known from the data, and simulation with all wires working perfectly. All other parameters are kept the same. The reconstruction version used in v05_08_00. No cuts applied.

⁷⁹⁹ constructions, flash reconstruction, and calorimetry. The internal note is located
⁸⁰⁰ at DocDB 5849.

Selection I	Selection II	
orig and mod	Selection IIA	Selection IIB
pandoraNu vertices	pmtrack vertices	
pandoraNu tracks	pandoraNuPMA tracks	
containment	no containment	containment
minimum track length	no minimum track length	minimum track length
no calorimetry used	calorimetry used	
any multiplicity	multiplicity ≥ 2	multiplicity == 1

Table 2: Comparing some features of the two event selections developed to select CC inclusive events.

801 6 Event selection

802 For all selections described below, the data samples used are identical. The
803 samples are

- 804 • Data BNB on-beam: 567157 events
- 805 • Data BNB off-beam: 400675 events
- 806 • BNB+Cosmic MC: 191362 events
- 807 • inTime CORSIKA only: 4804 events (only for studies and purity estimate)

808 Both on-beam and off-beam data have been software triggered. Note that this
809 chapter is only describing the selection and passing rates. Final event distribu-
810 tions are shown in chapter 8.

811 Two different selections have been developed to filter CC inclusive events.
812 Table 2 lists some of the differences. At this point it is very useful to have two
813 different approaches and compare the resulting distributions. We are presenting
814 them side by side in the following sections. The two selections are not being
815 merged.

816 The track reconstruction algorithms used in these selections are pandoraNu
817 for selection I, and pandoraNuPMA for selection II. Both of them originate from
818 the same stream that went through the cosmic removal pass before the track
819 reconstructions were applied. This is illustrated in Figure 10 and explained in
820 section 4.5. In short: The entire event is reconstructed with pandoraCosmic and
821 trackkalmanhit. In the next step, all tracks that are completely throughgoing
822 (where throughgoing means crossing the border region between active volume
823 and fiducial volume twice) in either pandoraCosmic or trackkalmanhit (or both)
824 are picked and their associated hits are removed from the hit collection. Then,
825 the pandoraNu reconstruction is run on the remaining neutrino hit collection,
826 which returns 2D clusters in the wire planes as well as 3D vertices and tracks,
827 and assembles an event hierarchy (for pandora see section 4.3). Alternatively,
828 the PMA (Projection Matching Algorithm) track reco (see section 4.6) is run on
829 the pandoraNu clusters, and again produces 3D vertices and tracks. The third
830 option is to use the neutrino hit collection and run the linecluster algorithm to
831 obtain 2D cluster objects in each plane, followed by the PMA algorithm, which
832 returns tracks and vertices in 3D. This output is commonly called pmtrack. The
833 vertices from pmtrack are used in selection II. The track reconstruction from
834 pmtrack is not used in the following.

835 **6.1 Selection I**

836 This selection is largely based on the MC performance study that was reviewed
837 in November 2015 and is documented in DocDB 4925. This DocDB also contains
838 more details on the cut motivation and distributions based on MC.

839 The algorithms used in the selection described in this paragraph are

- 840 • Flash reconstruction: opflashSat
841 • Track reconstruction: pandoraNu
842 • Vertex reconstruction: pandoraNu

843 The reconstruction chain and these algorithms are described in section 4. The
844 choice of algorithms was made after an initial selection was run with all com-
845 binations of available vertex and track reconstruction algorithms. Most give
846 similar results for efficiency and purity (see appendix D).

847 **6.1.1 Event Selection Scheme (original)**

848 ν_μ CC events are selected in seven steps of which most are related to the TPC
849 topology of the event.

850 The next requirement is that the event has at least one scintillation light flash
851 (opflashSat) in the beam trigger window. At least one of these flashes are then
852 required to produce more than 50 photo electrons (PE) on all photo cathodes
853 of the MicroBooNE light detection system combined. Of these flashes the most
854 intense one is considered to originate from a neutrino interaction candidate,
855 meaning only this flash will be used downstream.

856 After these first stages, the TPC reconstruction is used to perform further
857 topology cuts. The next requirement is placed on the vertex positions. We
858 require at least one of the reconstructed vertices to be within the fiducial volume
859 (FV) confines at t_0 of the MicroBooNE detector. The fiducial volume boundaries
860 chosen here are 10 cm in x (drift direction) and z (beam direction), and 20 cm in
861 y (vertical direction) from the edges of the active volume of the TPC. Selected
862 vertices are further required to possess at least one adjacent track start or end
863 point within a 5 cm radius. This cut passes most events but there are cases
864 where a shower instead of a track is associated with the vertex (due to nue
865 contamination or just mis-reconstructed) and these are not further considered
866 in this selection. Of all the track-vertex pairs found in this step, only the pair
867 with the longest track is selected and used further downstream.

868 The next cut is some basic flash matching. This is a placeholder for a
869 more sophisticated flash matching algorithm that is currently being developed
870 but might not be available by the time of NEUTRINO2016. For the flash
871 matching, only the position of the flash in z-direction, and the track z-projection
872 are compared. The spacing of PMTs in y is very sparse and has shown to not add
873 additional information. The flash reconstruction is described in section 3. The
874 z-position of the flash is compared to the z-positions of track starts and ends.
875 It has to be within 80 cm of the track end points. If the flash is between track
876 start and end point, its distance to the track is defined to be 0 cm. This flash
877 matching cut was optimized to separate cosmic only background from neutrino
878 events (see DocDB 5342).

879 After, we require the track to be fully contained in the FV and at last
880 the track has to be a greater range than 75 cm. The range is defined as the 3D
881 distance between the track's start and end point. This length cut has been taken
882 from the MC performance study, where 75 cm was found to be the optimum
883 cut to remove both, additional cosmic and NC backgrounds.

884 A track getting through all these cuts is considered to be the muon of a
885 ν_μ CC event. For better readability a short version of the selection scheme
886 described above is provided in this list:

- 887 1. At least one flash of > 50 PE within the beam gate.
- 888 2. At least one reconstructed vertex in the FV.
- 889 3. At least one track originating within 5 cm around any vertex.
- 890 4. For further cuts only the longest track is considered as a muon candidate
- 891 5. The longest track is within 80 cm (z-axis only) of the flash.
- 892 6. The longest track is fully contained.
- 893 7. The track range is larger than 75 cm

894 More plots for cut distributions can be found in DocDB 5561.

895 Table 3 lists the passing rates for MC events for the selection scheme de-
896 scribed above.

897 The same table for on-beam and off-beam data is shown in table 4. The
898 passing rate for the first cut that requires at least one flash ≥ 50 PE during
899 the beam window is very different for data and Cosmic simulation. This is not
900 perfectly understood, but it is known that the optical modeling of the data is
901 not accurate. Additionally, the software trigger is not being run on simulation.
902 This will be followed up on in a future study.

903 6.1.2 Modified Event Selection Scheme

904 The before described scheme has one weak point: The longest track is selected
905 very early in the process (before the flash matching) and therefore can often
906 be of cosmic origin. The kind of problems arising with this, can be illustrated
907 by the following example. Let's assume we have a neutrino event with a true
908 vertex outside of the FV, which does create a flash. Now the second cut will
909 remove this reco vertex from the candidate list, but other reco vertices (starts of
910 tracks, decay knots, etc) still remain and can be (wrongly) selected as neutrino
911 candidate vertex. If now a cosmic event associated to such a vertex passes all
912 cuts by coincidence it will be picked as a neutrino candidate. In order to reduce
913 this mis-identified cosmic background, a new selection scheme was worked out.

914 The cut is the 50 PE flash within beam gate cuts. This is the same as above
915 and described in 6.1.1.

916 In the first selection modification, all vertices with a track originating within
917 5 cm are listed as possible neutrino vertex candidates. From there on all tracks
918 associated to a vertex are used to calculate a track length weighted average of
919 the θ -angle. Now, only the vertex with the most forward-going θ -angle average
920 of all associated tracks is considered as the neutrino vertex candidate. This is

Table 3: **Selection I: Original** The table shows passing rates for the above described event selection. Numbers are absolute event counts and Cosmic background is not scaled appropriately. The BNB+Cosmic sample contains all events, not just ν_μ CC inclusive. The selected events are further broken up in the following subsection. The numbers in brackets give the passing rate wrt the step before (first percentage) and wrt the generated events (second percentage). In the BNB+Cosmic MC Truth column shows how many true ν_μ CC inclusive events (in FV) are left in the sample. This number includes possible mis-identifications where a cosmic track is picked by the selection instead of the neutrino interaction in the same event. The cosmic only sample has low statistics, but please note that it is not used in any plots, it is just for illustrating the cut efficiency. The last column Signal:Cosmic only gives an estimate of the ν_μ CC events wrt the cosmic only background at each step. For this number, the cosmic background has been scaled as described in chapter 2. Note that this numbers is not a purity, since other backgrounds can't be determined at this step.

	BNB + Cosmic Selection		MC-Truth		Cosmic only	Signal: Cosmic only
Generated events	191362		45273		4804	1:22
≥ 1 flash with ≥ 50 PE	136219	(71%/71%)	44002	(97%/97%)	2979 (62%/62%)	1:14
≥ 1 vertex in FV	131170	(96%/69%)	43794	(99%/97%)	2805 (94%/58%)	1:13
≥ 1 track within 5 cm of vertex	129784	(99%/68%)	43689	(99%/97%)	2756 (98%/58%)	1:13
flash matching of longest track	44775	(34%/23%)	23647	(54%/52%)	647 (23%/13%)	1:5.7
track containment	10114	(23%/5.3%)	6882	(29%/15%)	61 (9.4%/1.3%)	1:1.9
track ≥ 75 cm	7358	(73%/3.8%)	5801	(84%/13%)	31 (51%/0.6%)	1:1.1

Table 4: **Selection I: Original** The table shows passing rates for the above described event selection applied to on-beam and off-beam data. The numbers in brackets give the passing rate wrt the step before (first percentage) and wrt the generated events (second percentage). Off-beam data has been scaled with a factor 1.23 to normalize to the on-beam data stream.

	on-beam	off-beam
Triggered	546910	477819
≥ 1 flash with ≥ 50 PE	135923 (25%/25%)	96748 (20%/20%)
≥ 1 vertex in FV	126817 (93%/23%)	89342 (92%/19%)
≥ 1 track within 5 cm of vertex	123779 (98%/23%)	86975 (97%/18%)
flash matching of longest track	32492 (26%/5.9%)	20940 (24%/4.4%)
track containment	7091 (22%/1.3%)	3643 (17%/0.8%)
track ≥ 75 cm	4731 (67%/0.9%)	2337 (64%/0.5%)

achieved by picking the largest track range weighted average of $|\cos(\theta)|$ (since $\cos(\theta) = 1$ is beam direction). In the next step, the FV cut is applied to the vertex candidate, i.e. the vertex is required to be within the confines of the FV. For all further cuts only the longest of all tracks associated with the vertex candidate is considered as the muon candidate of the neutrino event.

From here on the cuts are similar to the last three stages of the original event selection scheme and are described in detail in 6.1.1. Again the basic flash matching, the track containment cut, and the track range cut are applied with the same values as above. A track getting through this whole selection is considered to be the muon of a ν_μ CC event. For better readability a short version of the selection scheme described above is provided in this list:

1. At least one flash of > 50 PE within the beam gate.
2. At least one track originating within 5 cm around a vertex.
3. For further cuts only the vertex with the flattest tracks is used as the vertex candidate.
4. Vertex candidate in the FV.
5. For further cuts only the longest track associated with the vertex candidate is used.
6. The longest track is within 80 cm (z-axis only) of the flash.
7. The longest track is fully contained.
8. The track range is larger than 75 cm

The selection scheme for the modified selection is also illustrated in Figure 25. Table 5 lists the passing rates for MC events for the modified selection scheme described above.

The same table for on-beam and off-beam data is shown in table 6. Note again that normalization factors to be applied between on-beam and off-beam data are currently still under development and will be described in section 2.

6.1.3 Expected Backgrounds

The majority of the selected background events are of cosmic origin. We distinguish between two different cosmic-ray backgrounds. One is triggered by a cosmic-ray event occurring in the beam gate time window (so-called cosmic only events), the other is triggered by a beam induced interaction in the cryostat followed by a misidentification of a cosmic event as a neutrino event (so-called cosmics in BNB events). The first, cosmic-ray triggered, background can be subtracted from the selected events using the off-beam BNBEXT sample with a suited normalization, since the off-beam sample is purely a cosmic-ray triggered sample. The second background, can currently only be modeled by a MC simulation, where the BNB + Cosmic MC sample is used. Note, that at this point BNB + off-beam overlay samples are not yet available. In the future, this will be the sample desired for this type of background estimation.

Other backgrounds are originating in the neutrino beam contaminants. Here a major contribution is given by neutral current (NC) neutrino events, e.g. a

Table 5: **Selection I: Modified** The table shows passing rates for the above described event selection. Numbers are absolute event counts and Cosmic background is not scaled appropriately. The BNB+Cosmic sample contains all events, not just ν_μ CC inclusive. The selected events are further broken up in the following subsection. The numbers in brackets give the passing rate wrt the step before (first percentage) and wrt the generated events (second percentage). In the BNB+Cosmic Mc Truth column shows how many true ν_μ CC inclusive events (in FV) are left in the sample. This number includes possible mis-identifications where a cosmic track is picked by the selection instead of the neutrino interaction in the same event. The cosmic only sample has low statistics, but please note that it is not used in any plots, it is just for illustrating the cut efficiency. The last column Signal:Cosmic only gives an estimate of the ν_μ CC events wrt the cosmic only background at each step. For this number, the cosmic background has been scaled as described in chapter 2. Note that this numbers is not a purity, since other backgrounds can't be determined at this step.

	BNB + Cosmic Selection		MC-Truth		Cosmic only	Signal: Cosmic only
Generated events	191362		45273		4804	1:22
≥ 1 flash with ≥ 50 PE	136219	(71%/71%)	44002	(97%/97%)	2979	(62%/62%)
≥ 1 track within 5 cm of vertex	135830	(99%/71%)	43974	(99%/97%)	2975	(99%/62%)
vertex candidate in FV	79112	(58%/41%)	34891	(79%/77%)	1482	(50%/31%)
flash matching of longest track	40267	(51%/21%)	25891	(74%/57%)	340	(23%/7.1%)
track containment	19391	(48%/10%)	11693	(45%/26%)	129	(38%/2.7%)
track ≥ 75 cm	6920	(36%/3.6%)	5780	(49%/13%)	17	(13%/0.4%)

Table 6: **Selection I: Modified** The table shows passing rates for the above described modified event selection applied to on-beam and off-beam data. The numbers in brackets give the passing rate wrt the step before (first percentage) and wrt the generated events (second percentage). Off-beam data has been scaled with a factor 1.23 to normalize to the on-beam data stream.

	on-beam	off-beam
Triggered	546910	477819
≥ 1 flash with ≥ 50 PE	135923	(25%/25%)
≥ 1 track within 5 cm of vertex	134744	(99%/25%)
vertex candidate in FV	74827	(55%/14%)
flash matching of longest track	22059	(29%/4.0%)
track containment	10722	(49%/1.9%)
track ≥ 75 cm	3213	(30%/0.6%)

Table 7: **Selection I: Original** Signal and background event numbers at final selection level estimated from a BNB+Cosmic sample and Cosmic only sample normalized to 5×10^{19} PoT. The last column gives the fraction of this signal or background type to the total 2604 selected events.

		#Events	Fraction
Signal	ν_μ CC events with true vertex in FV	1135	39.7%
Backgrounds	Cosmics only events	1322	46.2%
	Cosmics in BNB events	300	10.5%
	NC events	40	1.4%
	ν_e and $\bar{\nu}_e$ events	1	0.1%
	$\bar{\nu}_\mu$ events	12	0.4%
	ν_μ CC events with true vertex outside FV	50	1.7%

Table 8: **Selection I: Modified** Signal and background event numbers at modified selection level estimated from a BNB+Cosmic sample and Cosmic only sample normalized to 5×10^{19} PoT. The last column gives the fraction of this signal or background type to the total 2189 selected events.

		#Events	
Signal	ν_μ CC events with true vertex in FV	1168	53.8%
Backgrounds	Cosmics only events	725	33.4%
	Cosmics in BNB events	144	6.6%
	NC events	75	3.5%
	ν_e and $\bar{\nu}_e$ events	4	0.2%
	$\bar{\nu}_\mu$ events	15	0.7%
	ν_μ CC events with true vertex outside FV	40	1.8%

963 charged pion track misidentified as a muon. Also ν_e -like and anti-muon-neutrino
964 ($\bar{\nu}_\mu$) events contribute. Anyhow, compared to the cosmic misidentification back-
965 ground, these beam related backgrounds are smaller by an order of magnitude.
966 They also can not be subtracted and are estimated through the MC truth of
967 our BNB + Cosmic MC model, as already described above.

968

969 Based on these numbers, the following performance values of the original
970 and the modified selection can be calculated:

- 971 • Efficiency: Number of selected true ν_μ CC events divided by the number
972 of expected true ν_μ CC events with interaction in the FV.
 - 973 – Original selection: (12.0 ± 0.2) %
 - 974 – Modified selection: (12.3 ± 0.2) %
- 975 • Purity: Number of selected true ν_μ CC events divided by sum of itself and
976 the number of all backgrounds.
 - 977 – Original selection: (39.7 ± 3.4) %
 - 978 – Modified selection: (53.8 ± 4.4) %
- 979 • Correctness: Fraction of selected true signal events where the selected
980 track originates from the true CC ν_μ muon.

- 981 – Original selection: (67.0 ± 1.3) %
982 – Modified selection: (67.4 ± 1.3) %

983 The 1135 (original) selected ν_μ CC signal events compose of 54 % quasi-
984 elastic, 35 % resonant, 10 % deep inelastic, and 1 % coherent. For the 1168
985 (modified) events we have 51 % quasi-elastic, 37 % resonant, 11 % deep inelastic,
986 and 1 % coherent.

987 **6.1.4 Truth Distributions of Selection I**

988 This section contains distributions of MC truth variables before and after the
989 selection, and the selection efficiencies calculated from these. The overall effi-
990 ciency, containing selection efficiency and acceptance is calculated for all ν_μ CC
991 signal events with a true interaction vertex within the FV. Since this selection
992 requires containment of the muon candidate track, we also calculate the effi-
993 ciency w.r.t. events with a fully contained muon. Therefore, all ν_μ CC signal
994 events with a true interaction vertex within the FV AND a fully contained muon
995 track were used. The before and after selection graphs were then used to calcu-
996 late the efficiency of the selection, see figures 13 through 18. The total efficiency
997 was also broken up into CCQE, CCRES, and CCDIS physics processes. These
998 plots are shown in Figures 19 through 24).

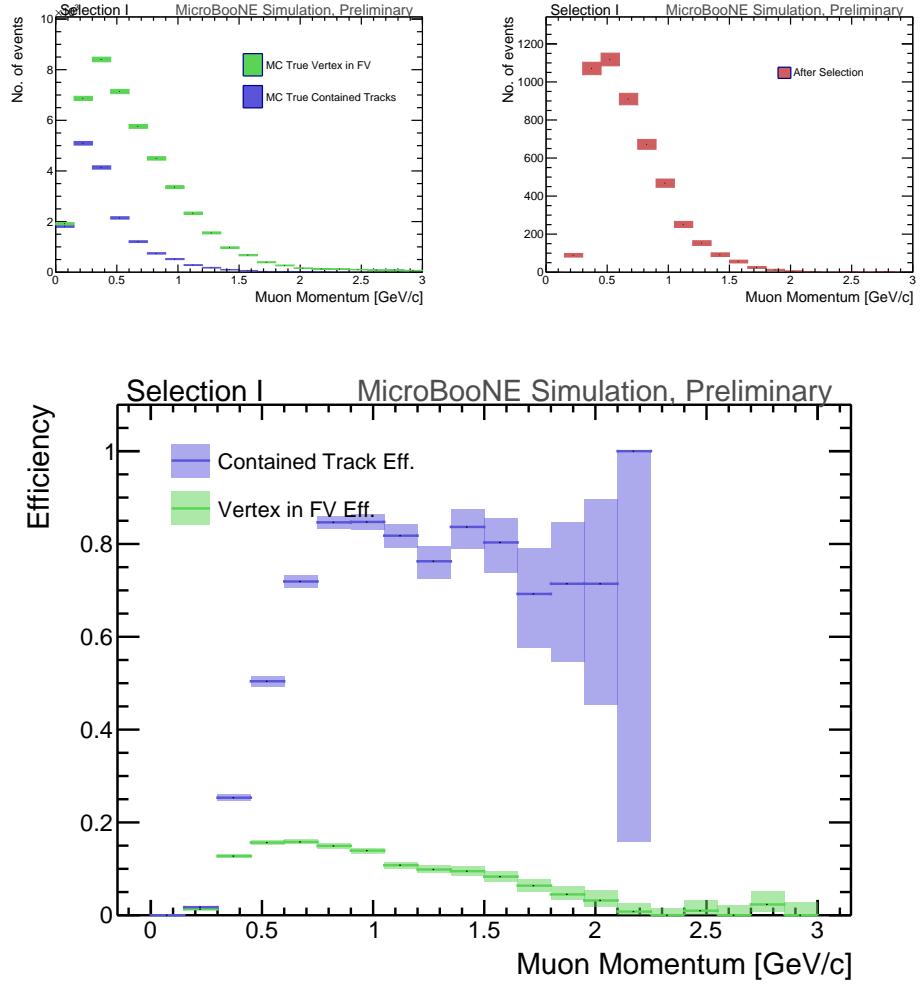


Figure 13: **Selection I:** Original MC momentum distributions of the muon originating from a ν_μ CC interaction. On the upper left we see the momentum distributions of events with the vertex in the FV (green) and the events with a fully contained track (blue) before selection. Illustrated on the upper right side is the momentum distribution of selected tracks (red). Due to reconstruction inaccuracies this curve contains less than 2% of uncontained events. The lower plot shows the selection efficiencies for both truth cases.

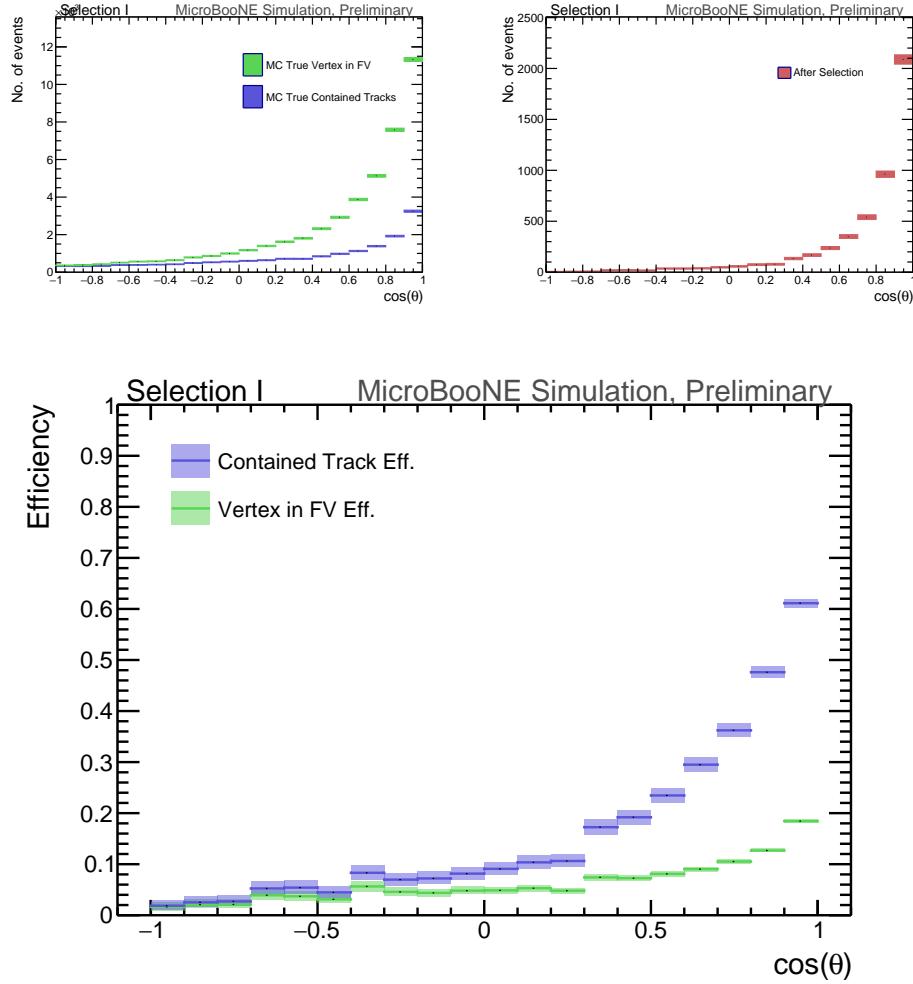


Figure 14: **Selection I:** Original MC $\cos \theta$ distributions of the muon originating from a ν_μ CC interaction. On the upper left we see the $\cos \theta$ distributions of events with the vertex in the FV (green) and the events with a fully contained track (blue) before selection. Illustrated on the upper right side is the $\cos \theta$ distribution of selected tracks (red). Due to reconstruction inaccuracies this curve contains less than 2% of uncontained events. The lower plot shows the selection efficiencies for both truth cases.

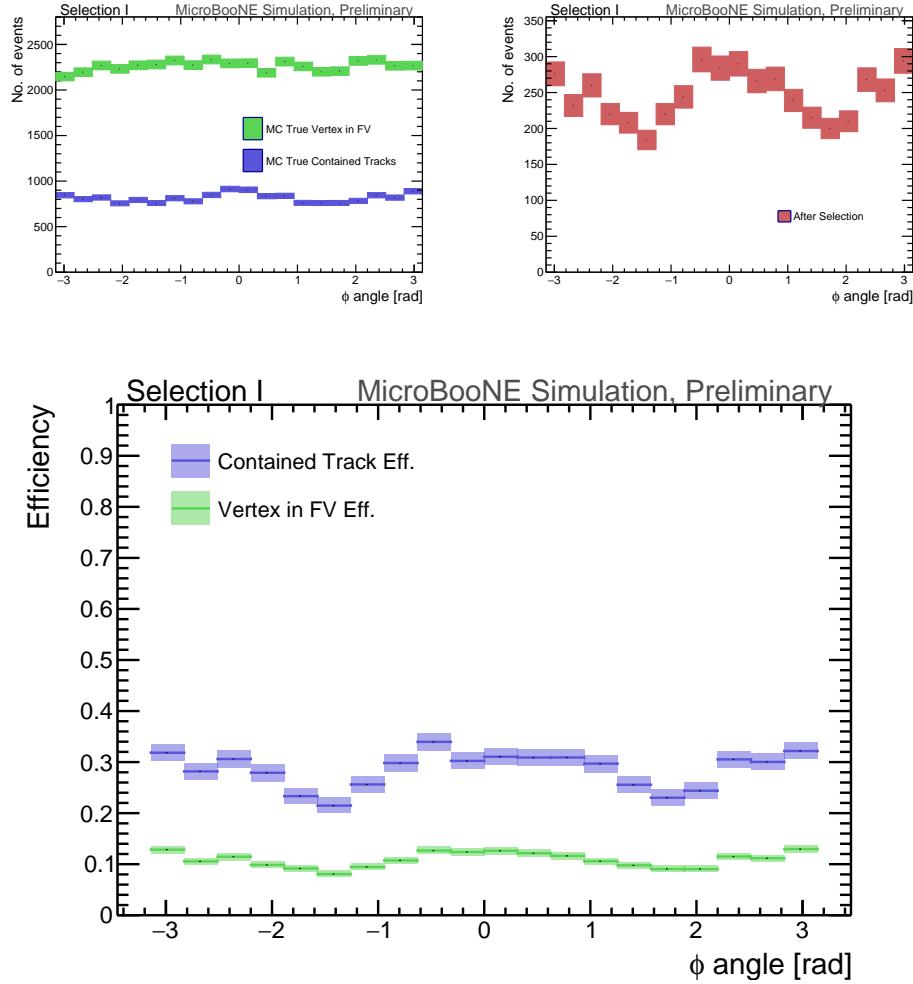


Figure 15: **Selection I:** Original MC ϕ distributions of the muon originating from a ν_μ CC interaction. On the upper left we see the ϕ distributions of events with the vertex in the FV (green) and the events with a fully contained track (blue) before selection. Illustrated on the upper right side is the ϕ distribution of selected tracks (red). Due to reconstruction inaccuracies this curve contains less than 2% of uncontained events. The lower plot shows the selection efficiencies for both truth cases.

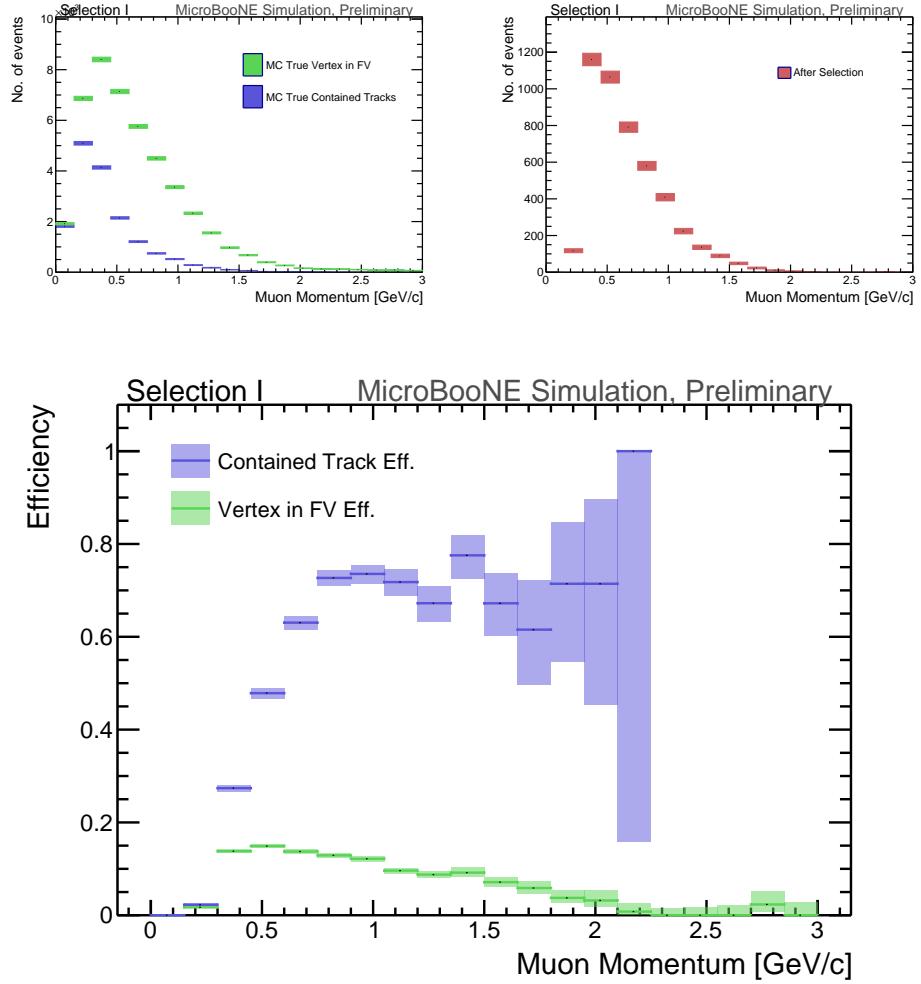


Figure 16: **Selection I: Modified** MC momentum distributions of the muon originating from a ν_μ CC interaction. On the upper left we see the momentum distributions of events with the vertex in the FV (green) and the events with a fully contained track (blue) before selection. Illustrated on the upper right side is the momentum distribution of selected tracks (red). Due to reconstruction inaccuracies this curve contains less than 2% of uncontained events. The lower plot shows the selection efficiencies for both truth cases.

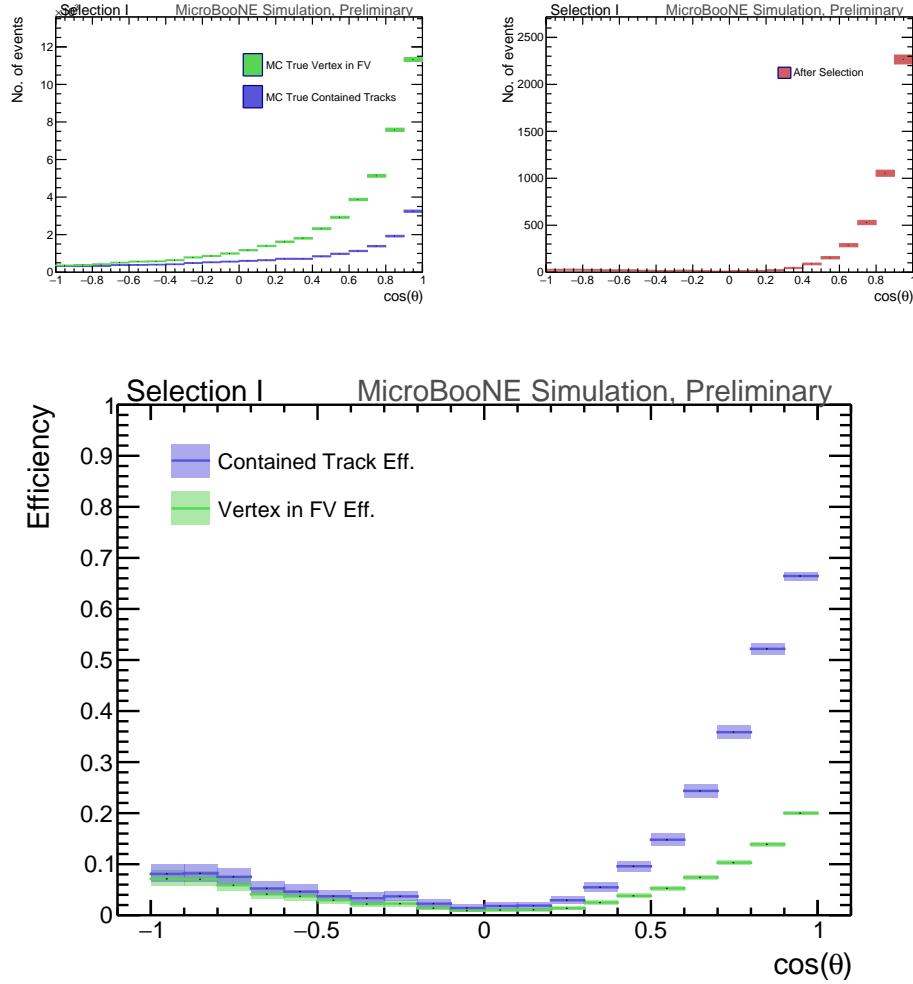


Figure 17: **Selection I: Modified MC** $\cos \theta$ distributions of the muon originating from a ν_μ CC interaction. On the upper left we see the $\cos \theta$ distributions of events with the vertex in the FV (green) and the events with a fully contained track (blue) before selection. Illustrated on the upper right side is the $\cos \theta$ distribution of selected tracks (red). Due to reconstruction inaccuracies this curve contains less than 2% of uncontained events. The lower plot shows the selection efficiencies for both truth cases.

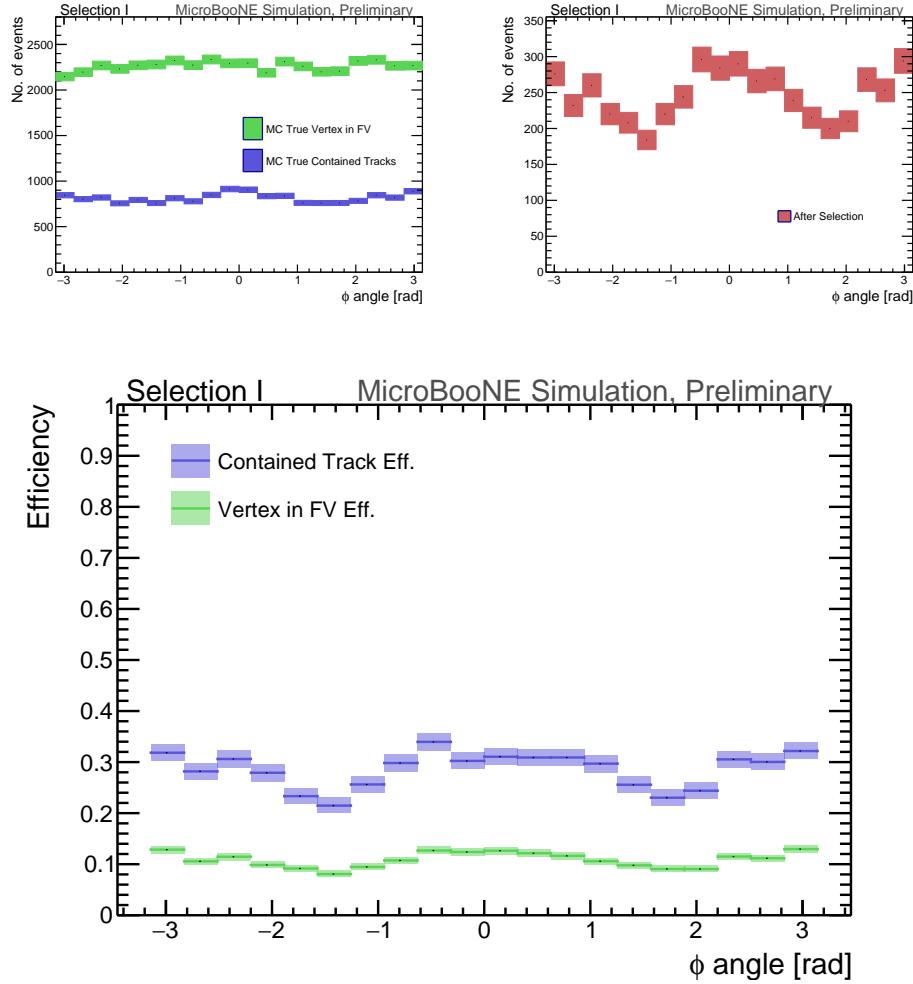


Figure 18: **Selection I: Modified** MC ϕ distributions of the muon originating from a ν_μ CC interaction. On the upper left we see the ϕ distributions of events with the vertex in the FV (green) and the events with a fully contained track (blue) before selection. Illustrated on the upper right side is the ϕ distribution of selected tracks (red). Due to reconstruction inaccuracies this curve contains less than 2% of uncontained events. The lower plot shows the selection efficiencies for both truth cases.

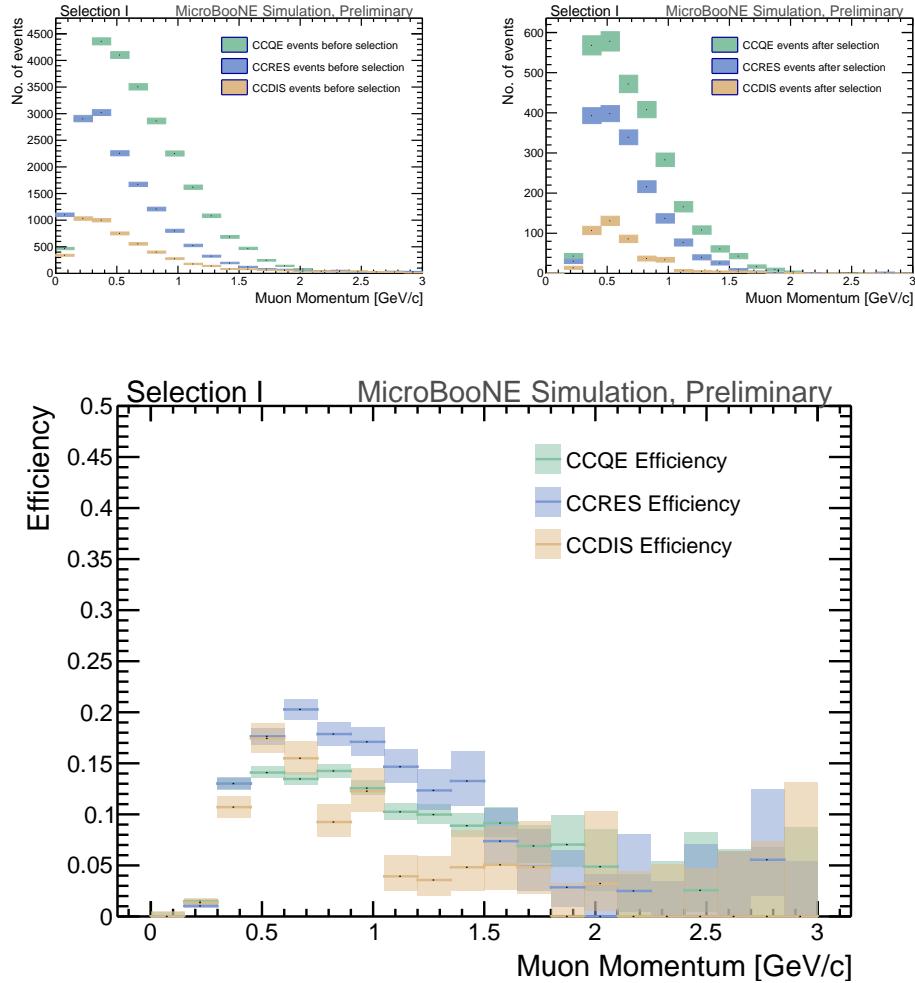


Figure 19: **Selection I: Original** MC momentum distributions of the muon originating from a ν_μ CC interaction. On the upper left we see the momentum distribution of events with the vertex in the FV split up into CCQE (red), CCRES (yellow), and CCDIS (green) before selection. Illustrated on the upper right side are momentum distributions of selected tracks (same color code). The lower plot shows the selection efficiencies for all three interaction types. Coherent interactions are not shown due to small statistics. Please note that the definition of QE, RES and DIS is according to GENIE and not based on e.g. a cut on the invariant mass W . This varies from generator to generator and explains why MiniBooNE (using NUANCE) reported a smaller DIS fraction ($\sim 1\%$).

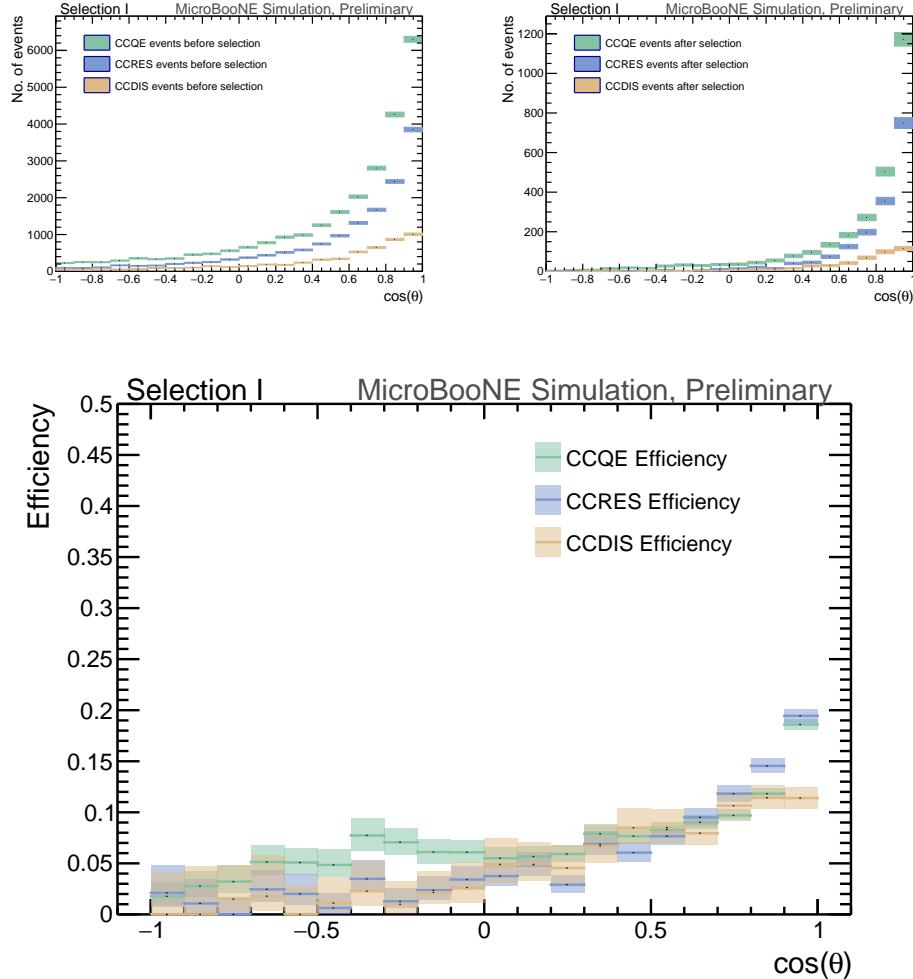


Figure 20: **Selection I: Original MC** $\cos \theta$ distributions of the muon originating from a ν_μ CC interaction. On the upper left we see the $\cos \theta$ distribution of events with the vertex in the FV split up into CCQE (red), CCRES (yellow), and CCDIS (green) before selection. Illustrated on the upper right side are $\cos \theta$ distributions of selected tracks (same color code). The lower plot shows the selection efficiencies for all three interaction types. Coherent interactions are not shown due to small statistics. Please note that the definition of QE, RES and DIS is according to GENIE and not based on e.g. a cut on the invariant mass W . This varies from generator to generator and explains why MiniBooNE (using NUANCE) reported a smaller DIS fraction ($\sim 1\%$).

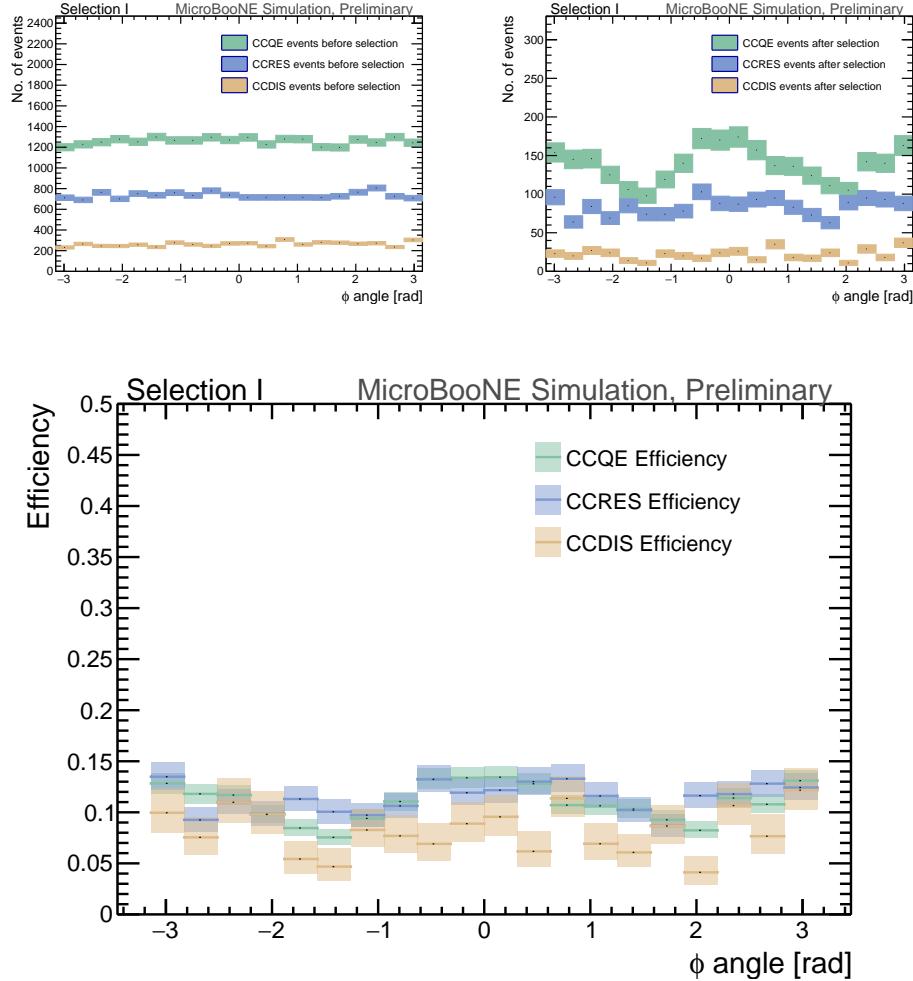


Figure 21: **Selection I: Original** MC ϕ distributions of the muon originating from a ν_μ CC interaction. On the upper left we see the ϕ distribution of events with the vertex in the FV split up into CCQE (red), CCRES (yellow), and CCDIS (green) before selection. Illustrated on the upper right side are ϕ distributions of selected tracks (same color code). The lower plot shows the selection efficiencies for all three interaction types. Coherent interactions are not shown due to small statistics. Please note that the definition of QE, RES and DIS is according to GENIE and not based on e.g. a cut on the invariant mass W . This varies from generator to generator and explains why MiniBooNE (using NUANCE) reported a smaller DIS fraction ($\sim 1\%$).

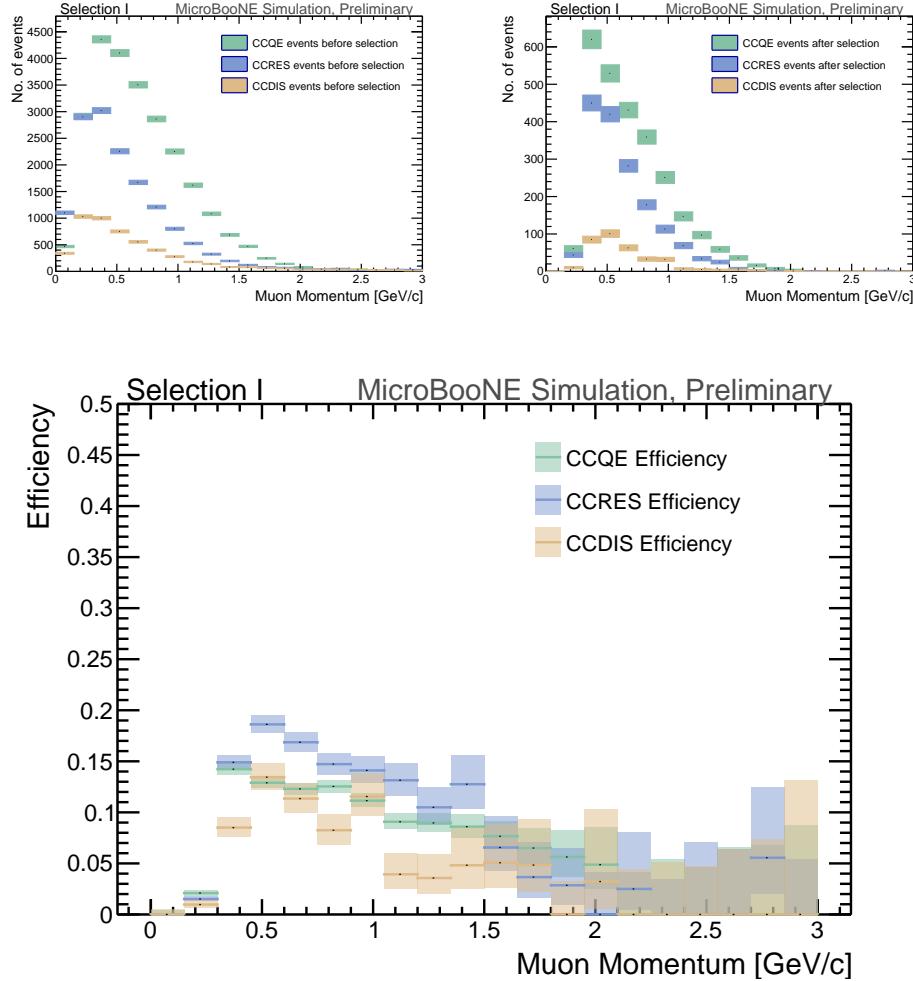


Figure 22: **Selection I: Modified** MC momentum distributions of the muon originating from a ν_μ CC interaction. On the upper left we see the momentum distribution of events with the vertex in the FV split up into CCQE (red), CCRES (yellow), and CCDIS (green) before selection. Illustrated on the upper right side are momentum distributions of selected tracks (same color code). The lower plot shows the selection efficiencies for all three interaction types. Coherent interactions are not shown due to small statistics. Please note that the definition of QE, RES and DIS is according to GENIE and not based on e.g. a cut on the invariant mass W . This varies from generator to generator and explains why MiniBooNE (using NUANCE) reported a smaller DIS fraction ($\sim 1\%$).

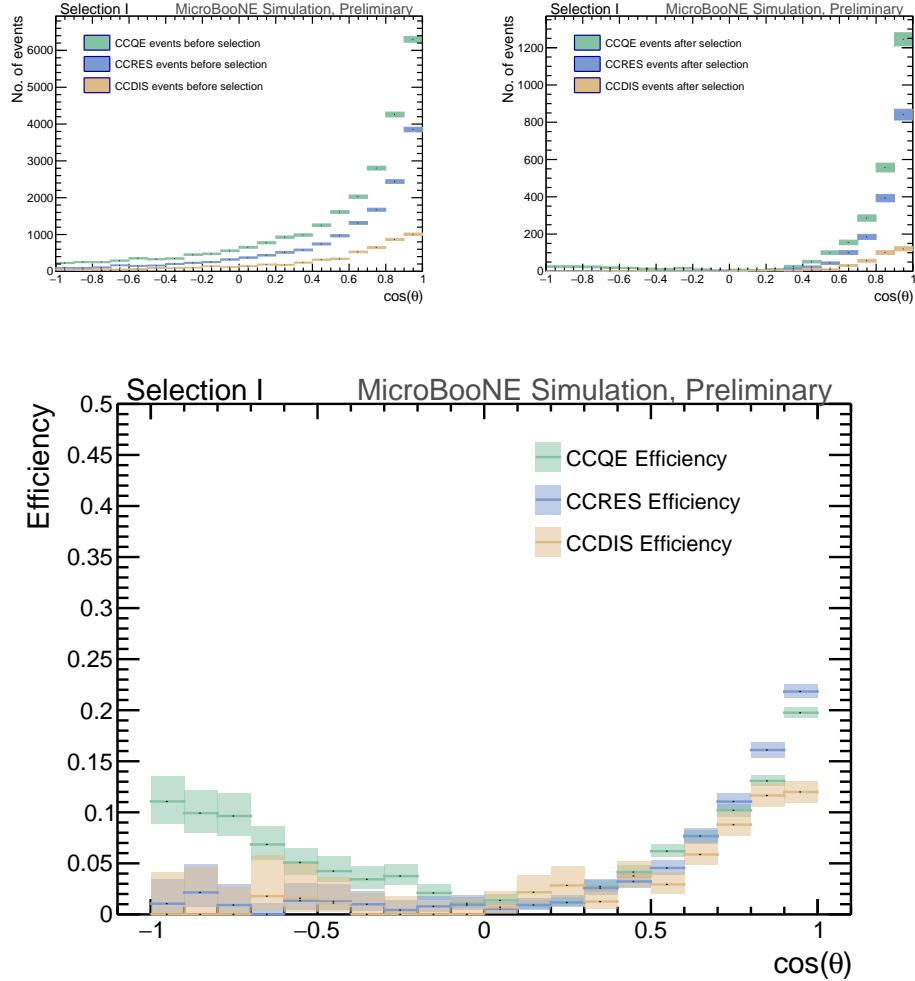


Figure 23: **Selection I:** Modified MC $\cos \theta$ distributions of the muon originating from a ν_μ CC interaction. On the upper left we see the $\cos \theta$ distribution of events with the vertex in the FV split up into CCQE (red), CCRES (yellow), and CCDIS (green) before selection. Illustrated on the upper right side are $\cos \theta$ distributions of selected tracks (same color code). The lower plot shows the selection efficiencies for all three interaction types. Coherent interactions are not shown due to small statistics. Please note that the definition of QE, RES and DIS is according to GENIE and not based on e.g. a cut on the invariant mass W . This varies from generator to generator and explains why MiniBooNE (using NUANCE) reported a smaller DIS fraction ($\sim 1\%$).

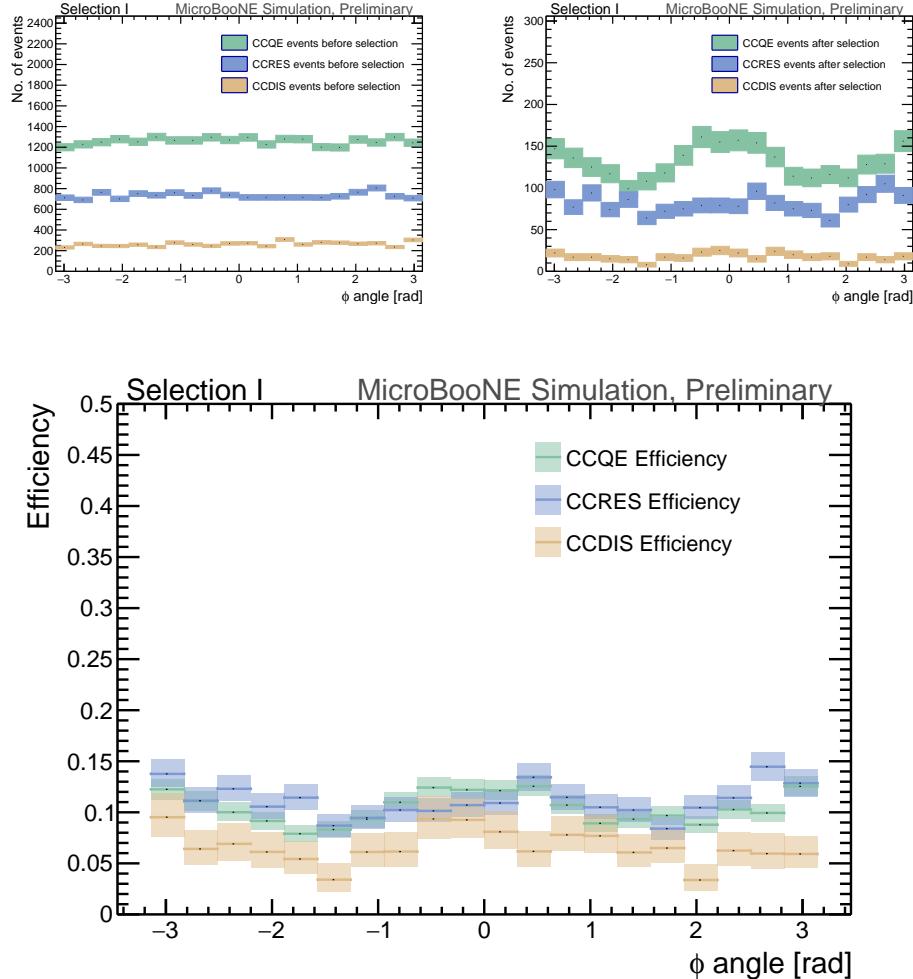


Figure 24: **Selection I: Modified MC ϕ distributions of the muon originating from a ν_μ CC interaction.** On the upper left we see the ϕ distribution of events with the vertex in the FV split up into CCQE (red), CCRES (yellow), and CCDIS (green) before selection. Illustrated on the upper right side are ϕ distributions of selected tracks (same color code). The lower plot shows the selection efficiencies for all three interaction types. Coherent interactions are not shown due to small statistics. Please note that the definition of QE, RES and DIS is according to GENIE and not based on e.g. a cut on the invariant mass W . This varies from generator to generator and explains why MiniBooNE (using NUANCE) reported a smaller DIS fraction ($\sim 1\%$).

999 **6.2 Selection II**

1000 **6.2.1 Selection IIA: track multiplicity 2 (or larger) and no contain-
1001 ment requirement**

1002 This selection scheme, unlike the previous one, doesn't require track contain-
1003 ment. In order to enhance the CC inclusive signals and suppress the large cosmic
1004 ray background, the main idea is to require a minimum track multiplicity (the
1005 number of the tracks associated with vertex).

1006 The selection algorithm takes place post analysis tree, which consist of dif-
1007 ferent reconstructed physics objects such as flashes, tracks and vertices. The
1008 reconstruction algorithm which are used for this selection are:

- 1009 • Flash: opflashSat
1010 • Track: pandoraNuPMA
1011 • Vertex: pmtrack

1012 The selection cuts are optimized using the BNB + Cosmic and Cosmic in
1013 time MC samples. For the BNB + cosmic sample, each event has one neu-
1014 trino interaction overlaid with cosmic ray background. The relevant MC truth
1015 variables stored in the analysis trees are **ccnc_truth** (0 is charge current, 1 is
1016 neutral current); **nuPDG.truth** (14 and -14 are ν_μ and $\bar{\nu}_\mu$, 12 and -12 are ν_e
1017 and $\bar{\nu}_e$) and **mode_truth** (0 is CCQE, 1 is resonant, 2 is DIS, 3 is coherent).
1018 Additionally, by referencing the **trkorigin** (0 as tracks from the neutrino inter-
1019 action, and 1 as tracks from cosmic), we can choose the cut values to optimize
1020 the selected neutrino cosmic ratio.

1021 The order of the MC selection cuts are the following:

- 1022 1. The event must have no less than one flash, which is higher than 50 PE,
1023 inside of the beam gate window (same beam gate window as defined in
1024 section 3.3 and that is also used in selection I). The highest flash met this
1025 condition is chosen as the flash for later flash track matching reference.
- 1026 2. Reconstructed track vertex association: for all the reconstructed vertices
1027 inside the fiducial volume (defined as inner box of TPC 20 cm from X
1028 and Y boundaries, and 10 cm from Z boundaries of the active volume) in
1029 this event, scan through all the tracks and associate the tracks with one
1030 end within 3cm distance from the current vertex. The number of tracks
1031 associated with this vertex is track multiplicity. This is not a event cut.
- 1032 3. For all the tracks in this event, if there is at least one track with a closest
1033 approach distance of 70 cm from the flash (chosen from the previous step)
1034 center in the Z direction, this event will survive.
- 1035 (a) For track multiplicity = 1, the contained cut will be applied. Details
1036 see next section.
- 1037 (b) For track multiplicity = 2, only keep the event if it is not tagged
1038 as muon decay to Michel. After the Michel tagging, an additional
1039 minimum angle between the two tracks is applied which is useful in
1040 particular to remove broken cosmic tracks. The details for the Michel
1041 tagging and the angular cut are described below.

- 1042 (c) For track multiplicity > 2 , all the events will pass
- 1043 4. For multiplicity ≥ 2 , the longest reco track associated to the selected
1044 vertex in the surviving event is the muon candidate.
- 1045 The selection scheme is also illustrated in Figure 25.

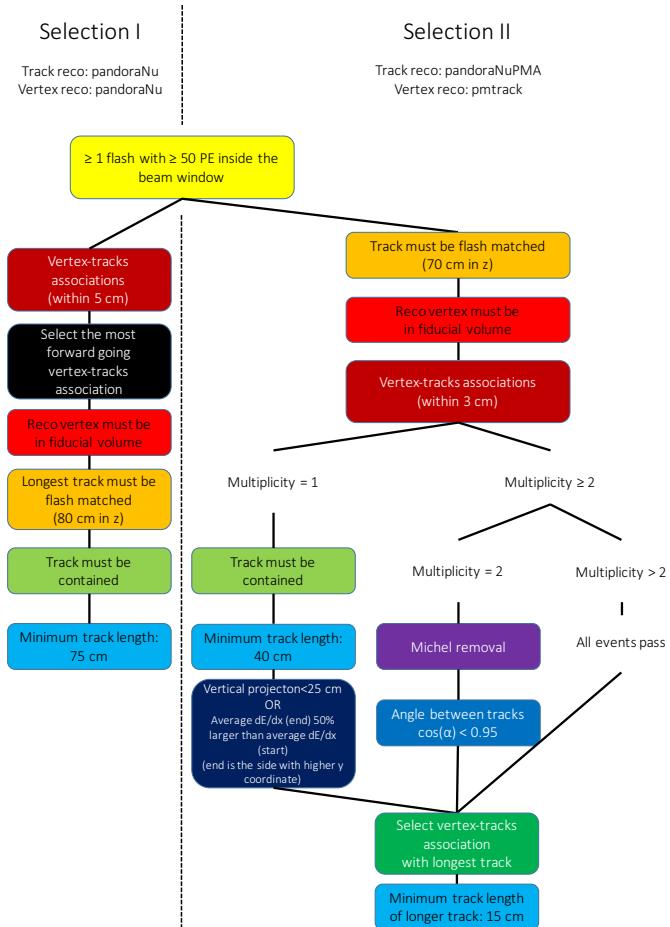


Figure 25: Event selection diagram for selections I and II, illustrating the different paths. Both selections start with the exact same cut requiring a flash in the beam window using the same flash reconstruction. Boxes in the same color symbolize similar cuts (not necessarily the same cut values) for different paths.

1046 The cuts in step 1, 2, 3 are straightforward. The next event cut is the simple
1047 track flash matching cut. The definition of the track flash distance is already
1048 described in the previous section. Figure 26 shows the distribution of the
1049 distance between the flash and reconstructed tracks. The tracks from neutrino
1050 interaction peak at distance = 0 and drop fast as distance increases. Cosmic
1051 tracks, however have a more uniform distance distribution, which is expected.
1052 Therefore, to amplify the signal to background ratio, the distance ≤ 70 cm is

1053 chosen as the basic flash track matching cut. Only the events with at least one
 1054 track satisfying the basic flash track matching are kept.

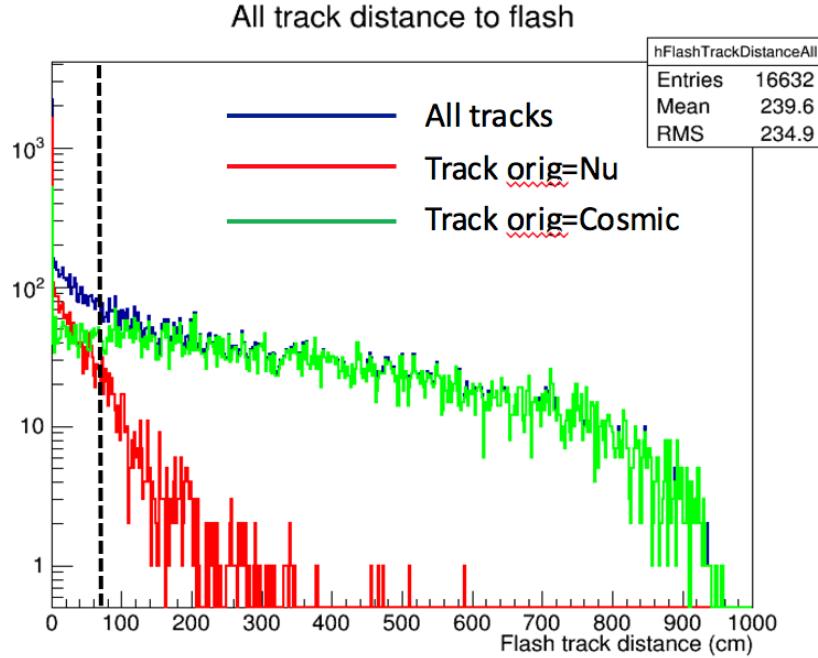


Figure 26: Track Flash distance distribution. The blue histogram includes all tracks from the BNB + Cosmic sample, the red histogram shows the tracks from neutrino interactions, and the green histogram shows tracks from cosmics. A ≤ 70 cm (shown as the black dash line) is applied as the track flash matching cut.

1055 Before applying the track multiplicity cut, track vertex association needs
 1056 to occur. This is done only for reconstructed vertices inside the FV. All re-
 1057 constructed tracks (contained and uncontained) are taken into account. The
 1058 association is demonstrated at the top graph of Figure 27. The 3 cm cut is cho-
 1059 sen to avoid the contamination from the broken tracks. The 3 cm cut was not
 1060 specifically optimized. The track multiplicity after the association is shown at
 1061 the bottom of Figure 27. Obviously, cosmic backgrounds (green) are suppressed
 1062 for multiplicities ≥ 2 . By scanning some event displays it was discovered that
 1063 the remaining cosmic background is mostly from cosmic muon decays producing
 1064 Michel electrons (illustrated by the red circle in Figure 27). Therefore, Michel
 1065 tagging and removal are necessary to further suppress the background.

1066 The following list describes the Michel tagging criteria. These cuts make
 1067 use of reconstructed calorimetry information per track. In order to correct a
 1068 mis-calibration between data and MC in the reconstructed energy loss per unit
 1069 path dE/dx , the values for data are being corrected before the cuts are applied.
 1070 The correction factors are global scaling factors per plane. They are derived
 1071 based on studies performed in the context of the data/MC comparison note
 1072 [DocDB 5849]. Gaussian distributions were fitted to both data and MC and
 1073 a scaling factor was derived to correct for shift in mean (see Figure 28. The
 1074 scaling factors are:

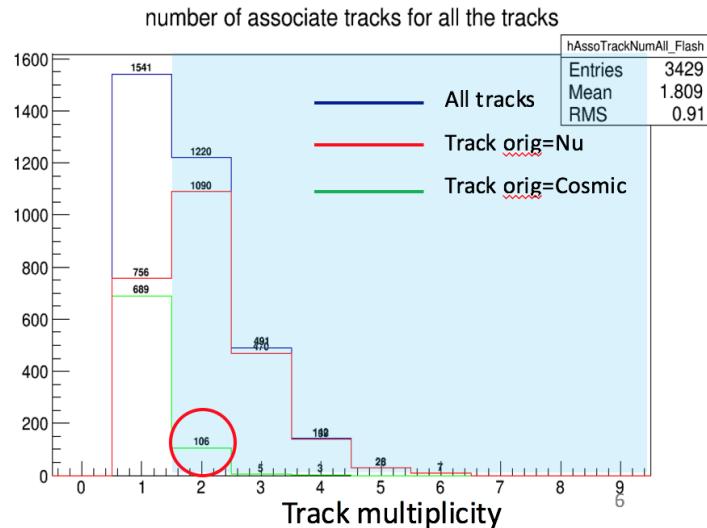
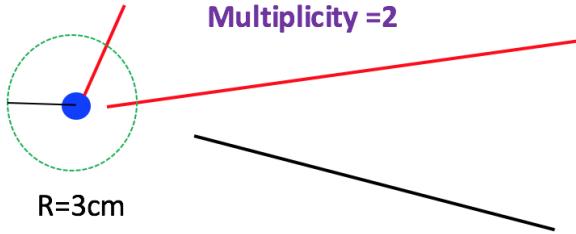


Figure 27: Track vertex association (top), and track multiplicity distribution for BNB+Cosmic MC. This signal (red) are tracks originating from a neutrino (any flavor, CC interaction). The cosmic background (green) are cosmics overlaying neutrino interactions. The blue is the sum of neutrino tracks (red), the cosmic tracks (green) plus tracks with undefined origin (e.g. noise). Pure cosmic background is not shown. This explains why the green integral is smaller than the red. Each track has one entry in this histogram. E.g. if there are three tracks associated with one common vertex, then there will be three entries at multiplicity equal to 3.

- 1075 • U plane: 1.118
- 1076 • V plane: 1.308
- 1077 • Y plane: 1.194

1078 Please note that the effect on the passing event number for selection II was
 1079 compared for the cases with and without these scaling factors. It is only around
 1080 1%.

1081 Events with a tagged Michel will be removed. Michel tagging is only applied
 1082 for multiplicity=2. The following three conditions are used to identify a vertex
 1083 as a “Michel event”:

- 1084 1. Calorimetric information of the plane with the most reconstructed hits
 1085 (usually the collection plane): the energy loss (dE/dx) at the track start

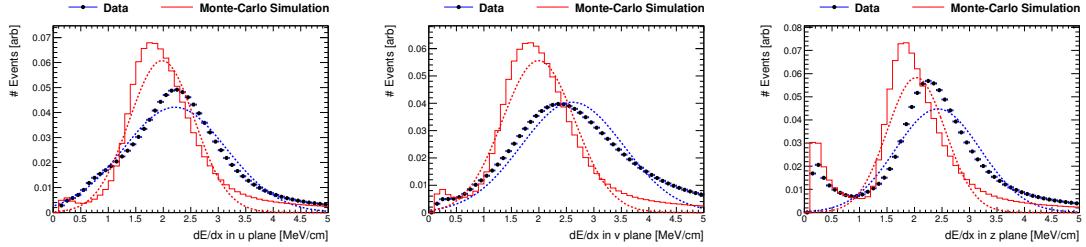


Figure 28: The distributions show the reconstructed dE/dx values along all tracks in an event for data and MC. For data, events from the off-beam stream are used. For MC, the inTime cosmic only sample is used. Both histograms are area normalized. Gaussian distributions have been fitted to the data and MC in order to determine a first order correction factor for the offset between data and MC energy calibration.

1086 is larger than at the track end. Here, the start is defined as the track
 1087 end closer to the associated vertex. The dE/dx is the average of the 10
 1088 hits of each end if the total number of the hits in the track is no less
 1089 than 20, otherwise dE/dx is the average of the hits from half track in each
 1090 side (see Figure 29). Additionally, it is checked if the dE/dx at the track
 1091 start is larger than 2.5 MeV/cm and the dE/dx at the track end is less
 1092 than 4 MeV/cm. These values are chosen based on dE/dx start versus
 1093 end comparison between tracks from neutrino and cosmic rays, shown in
 1094 Figure 29. It is worth noting that there are still quite a lot of neutrino
 1095 tracks inside the Michel tagging box. These events are dominated by the
 1096 mis-assigned dE/dx when the two tracks have a small angle. This is a
 1097 known reconstruction issue.

- 1098 2. The stopping point's Y position (further end from the vertex) of the longer
 1099 of the two tracks is above the fiducial boundary (20 cm from the active
 1100 volume). This criteria corresponds to partial containment in Y during the
 1101 Michel tagging.
- 1102 3. The track length of the shorter track is less than 30cm. This is motivated
 1103 by the Michel electron being a MIP particle with dE/dx approximately
 1104 2 MeV/cm, and the Michel energy spectrum peaking around 53 MeV.
 1105 Therefore, the range of the Michel should be no more than 30 cm.

1106 If the vertex meets the above conditions (1 OR 2) AND 3, a Michel tagging
 1107 flag is set as true, and this event will be removed from the selection.

1108 After Michel removal, an additional angle cut is applied for events with mul-
 1109 tiplicity==2. Figure 30 shows the dot product of the two track start unit
 1110 vectors from the events after Michel removal. 0 means the two tracks are per-
 1111 pendicular to each other, and 1 means that the angle between the tracks is 0.
 1112 There is an excess of events in the last bin from the cosmic distribution, and
 1113 no obvious excess observed from CC interaction. Therefore, only events with
 1114 $\cos(\alpha) < 0.95$ pass the selection, where α is the angle between the two tracks.

1115 The last check on events with multiplicity ≥ 2 after passing all the cuts
 1116 described above is the following. The longest track originating from the selected
 1117 vertex is considered for the muon kinematic distributions and in the event tables
 1118 below.

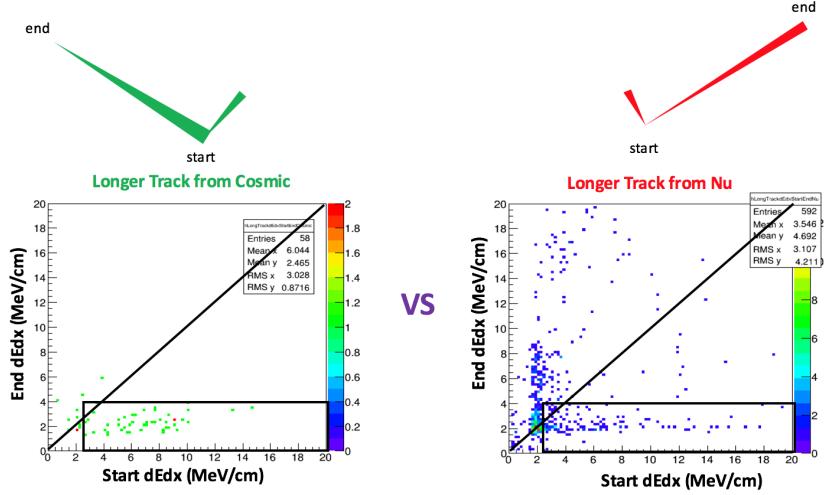


Figure 29: The top graphs are the cartoons qualitatively demonstrating the energy deposition at the start and end of the tracks for cosmic muon decays to Michel (left) and CC neutrino interaction (right). The longer tracks, which suppose to be muon tracks in both cases, have opposite dE/dx trend from start to end for cosmic ray induced muons and neutrino induced muons. The 2D histograms underneath are generated from BNB+Cosmic MC and show the energy loss dE/dx at the start vs the end for the longer track. The events inside of the black cut box without the small left upper corner are satisfying the calorimetry criteria, which is part of the Michel tagging algorithm.

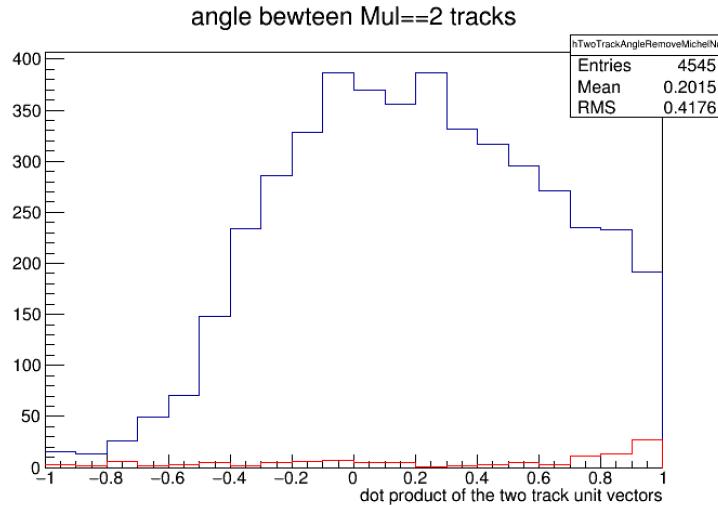


Figure 30: Dot product of the two track direction vectors for events with multiplicity == 2. Blue curve shows two-track events from CC interaction, the red curve shows two-track events that are of cosmic origin. This background is what is referred to as Cosmic in BNB events background. The cosmic only background is not shown here but is expected to have a similar distribution. This background is even larger at this stage.

1119 **6.2.2 Selection IIB: single tracks with containment requirement**

1120 In this subsection, we describe the selection of single-track events, which is
 1121 referred to as track multiplicity = 1 in the previous subsection (selection 5(a)).
 1122 There are many cases where the neutrino interaction is reconstructed as a single
 1123 track:

- 1124 • All the charged hadrons are absorbed in the final state interaction.
- 1125 • The low energy charged hadrons are not reconstructed as tracks or some
 1126 of them are merged into the muon track.

1127 The main background in the single track sample is the cosmic ray muons, in
 1128 particular, the cosmic ray muons that stop in the TPC and appear to originate
 1129 from the inside of the TPC and exit the TPC. There could also be short single
 1130 tracks generated by the neutral particles (neutrons, photons etc.). To suppress
 1131 the background, we select only single tracks that fulfill the following:

- 1132 1. The track should be fully contained in the TPC fiducial volume. The FV
 1133 is 20 cm boundary in X and Y, and 10 cm boundary in Z from the active
 1134 volume.
- 1135 2. The track length is required to be at least 40 cm.
- 1136 3. The projected track length in the vertical direction is less than 25 cm or
 1137 the average dE/dx near the track end is at least 50% higher than the
 1138 average dE/dx near the track start. The track end is defined as the end
 1139 with higher y (vertical) coordinate. Average dE/dx is calculated using 10
 1140 hits if there are at least 20 hits on the track and half of the hits if there
 1141 are less than 20 hits on the track.

1142 The last cut intends to remove cosmics, because cosmic ray muons go down-
 1143 ward and leave a long projection in the vertical direction. Also, a stopping
 1144 cosmic ray muons deposits more energy before towards the end of the track,
 1145 which is in the inside of the TPC. This cut is illustrated in Fig. 31.

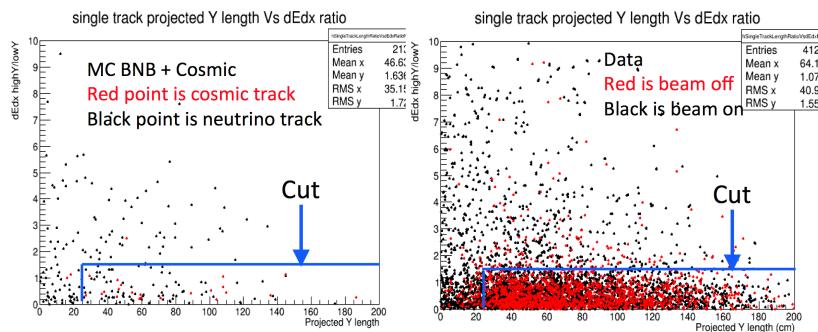


Figure 31: dE/dx ratio vs projected track length in the vertical direction. Left plot is made using the neutrino overlaid with CORSIKA cosmic background MC, where black points are for neutrino tracks while red points are for cosmic tracks. Right plot is made using data, where black points are for on-beam data while red points are for off-beam data.

1146 **6.2.3 Selection II signal efficiency and background rejection**

1147 After the selection described above are applied, several neutrino candidate ob-
 1148 jects are allowed per events. In the final step, a decision is made to select
 1149 one candidate object per event. This decision is based on choosing the object
 1150 with the longest track. It can be from the multiplicity ≥ 2 or multiplicity == 1
 1151 sub-branch.

1152 Update: To further reduce the “noise”, we decide to make the final track
 1153 length cut on the selected muon candidate from above cuts. This length cut
 1154 value is chosen based on Figure 32. A 15 cm minimum track length requirement
 1155 can suppress “noise” by more than a factor of 2, as well as reducing NC and
 1156 cosmic background but still maintain the efficiency of the signal selection.

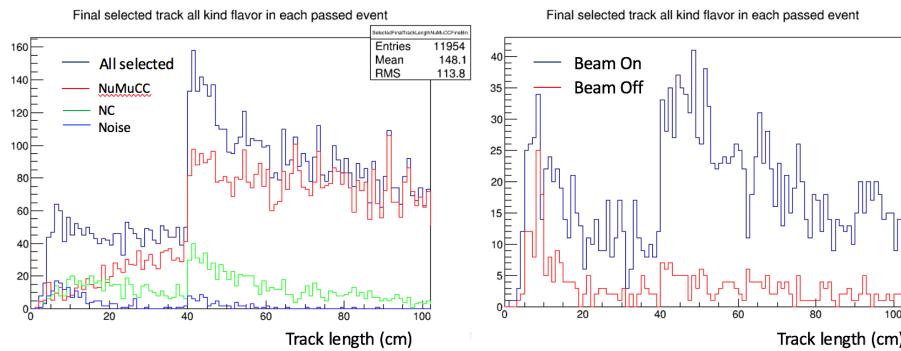


Figure 32: Zoomed in of the selected track length distribution with 1 cm binning. Left is different channels from MC BNB+Cosmic sample. Right is from beam on and beam off data. To reduce noise and other backgrounds, a 15cm minimum track length cut is made on the final selected track.

1157 Table 9 shows the signal and background event breakdown after applying
 1158 this final decision. While the cosmic background is smaller than in selection I,
 1159 the NC background is larger. The reason for this is that the NC background is
 1160 mostly reduced by the track length cut. The minimum track length requirement
 1161 is 75 cm for selection I, but only 15/40 cm for selection II.

1162 The efficiency of the selection is calculated as $11827/41261 = 28.7\% \pm 0.3\%$.

1163 Table 10 shows the event breakdown for the beam on and beam off data
 1164 events (updated from the final good run list). After all the cuts, we select 3228
 1165 events in the beam on data sample and 528 events in the beam off data sample
 1166 (scaled). The ratio from the beam on and beam off data is 5.2:1. This ratio
 1167 reflects selection efficiency on the signals and effective cosmic rejection.

1168 Table 11 shows the backgrounds expected at final selection level with selec-
 1169 tion II. Based on these numbers, the purity is calculated to be 65%.

1170 The selected ν_μ CC signal events compose of 43% quasi-elastic, 42% reso-
 1171 nant, 14% deep inelastic, and 1% coherent. More studies on the passing rate
 1172 for the different processes are discussed in subsection 6.2.5.

Table 9: **Selection II** The table shows passing rates for the above described event selection. Numbers are absolute event counts and Cosmic background is not scaled appropriately. The BNB+Cosmic sample contains all events, not just ν_μ CC inclusive. The events are further broken up in table 11. The numbers in brackets give the passing rate wrt the step before (first percentage) and wrt the generated events (second percentage). In the BNB+Cosmic Mc Truth column shows how many true ν_μ CC inclusive events (in FV) are left in the sample. Note that the FV in X differs for selection I and II. Other than selection I, this number does NOT includes possible mis-identifications where a cosmic track is picked by the selection instead of the neutrino interaction in the same event. The cosmic only sample has low statistics, but please note that it is not used in any plots, it is just for illustrating the cut efficiency. The last column Signal:Cosmic only gives an estimate of the ν_μ CC events wrt the cosmic only background at each step. For this number, the cosmic background has been scaled as described in chapter 2. Note that this numbers is not a purity, since other backgrounds can't be determined at this step.

	BNB + Cosmic Selection		MC-Truth		Cosmic only	Signal: Cosmic only
Generated events	191362		41261		4804	1:24
≥ 1 flash with ≥ 50 PE	136213	(71%/71%)	40263	(98%/98%)	2978	(62%/62%)
Flash track matching	108862	(80%/57%)	35209	(87%/85%)	2214	(74%/46%)
Track multiplicity $>=2$ (after Michel removal)	10753	(9.9%/5.6%)	9337	(27%/23%)	12	(0.5%/0.2%)
Track multiplicity $=1$	3512	(3.2%/1.8%)	2490	(7.1%/6.0%)	8	(0.36%/0.17%)
Final (combined track mult.)	14265	(13.6%/7.8%)	11827	(34.0%/29.0%)	20	(0.9%/0.4%)
						1:0.34

Table 10: **Selection II:** The table shows passing rates for the above described event selection applied to on-beam and off-beam data. The numbers in brackets give the passing rate wrt the step before (first percentage) and wrt the generated events (second percentage). Off-beam data has been scaled with a factor 1.23 to normalize to the on-beam data stream.

	on-beam	off-beam	
Triggered	546910	477819	
≥ 1 flash with ≥ 50 PE	135923	(25%/25%)	(20%/20%)
Flash track matching	86425	(64%/16%)	(61%/12%)
Track multiplicity $>=2$ (after Michel removal)	2191	(3%/0.4%)	(0.6%/0.07%)
Track multiplicity $=1$	1037	(1.2%/0.2%)	(0.3%/0.04%)
Final (combined track mult.)	3228	(3.7%/0.59%)	(0.9%/0.09%)

Table 11: **Selection II** Signal and background event numbers at final selection level estimated from a BNB+Cosmic sample and Cosmic only sample normalized to 5×10^{19} PoT. The last column gives the fraction of this signal or background type to the total selected events. The row labeled noise contains events with undefined track origin (neither neutrino nor cosmic).

Signal	ν_μ CC events with true vertex in FV	#Events	Fraction
		2472	64.5%
Backgrounds	Cosmics only events	853	22.2%
	Cosmics in BNB events	140	3.7%
	NC events	263	6.9%
	ν_e and $\bar{\nu}_e$ events	9	0.2%
	$\bar{\nu}_\mu$ events	14	0.4%
	ν_μ CC events with true vertex outside FV	59	1.5%
	Noise	24	0.6%

1173 **6.2.4 Truth distributions for selection II**

1174 To better understand the sensitivity of the selection algorithm in different phase
1175 space, we show here plots of the truth ν_μ CC lepton momentum, $\cos\theta$ and ϕ
1176 distributions before and after the selection. The curves before and the selection
1177 show only true ν_μ CC events with true interaction vertex inside the fiducial
1178 volume. The curves after selection only show true ν_μ CC events with true
1179 interaction vertex inside the fiducial volume where a track originating from the
1180 neutrino interaction was correctly selected. In case the selection selected not
1181 the muon but rather a pion (or any other track) from the neutrino interaction,
1182 we still plot the true muon kinematics in these plots.

1183 We also show selection efficiencies as a function of muon kinematics. Note
1184 that no systematic uncertainties are shown on these.

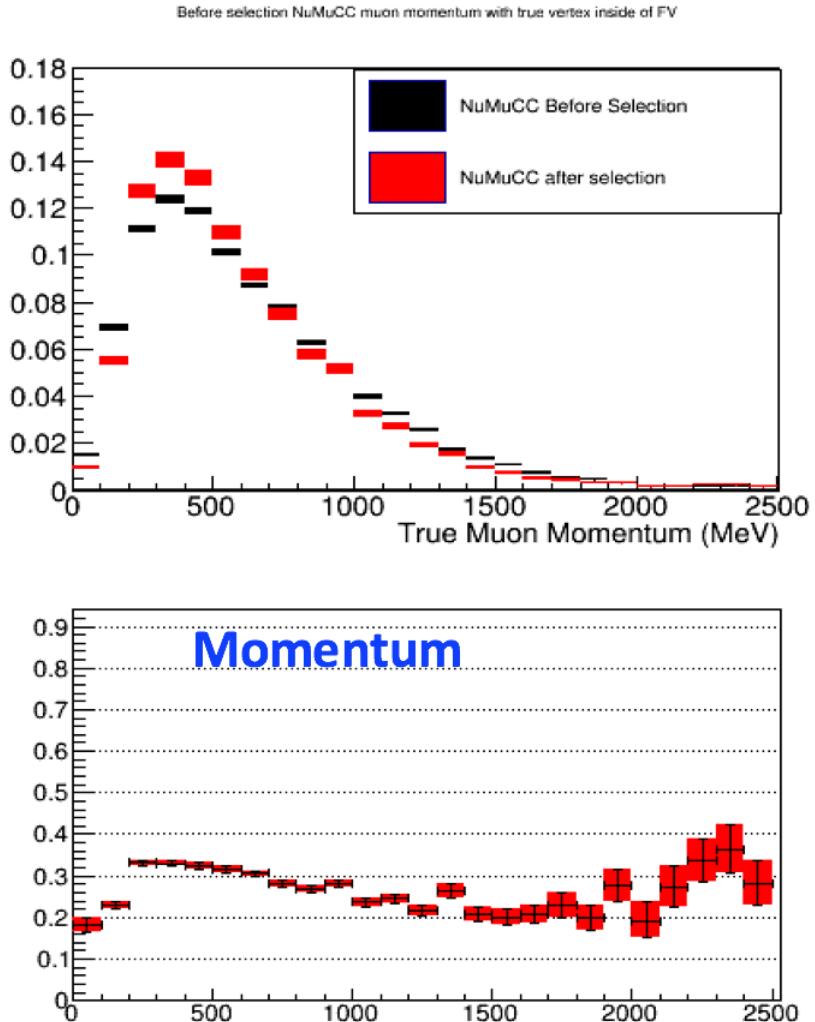


Figure 33: Top: MC truth lepton track momentum distribution before and after the selection. Both are shape normalized. Bottom: the efficiency of the selected true events divided by the all events with NuMuCC true vertex inside of fiducial volume. The division was done on absolute normalized distributions. Black is before the selection showing all the NuMuCC events with the true vertex inside of fiducial volume. Red is after the selection the true lepton information of the survived events.

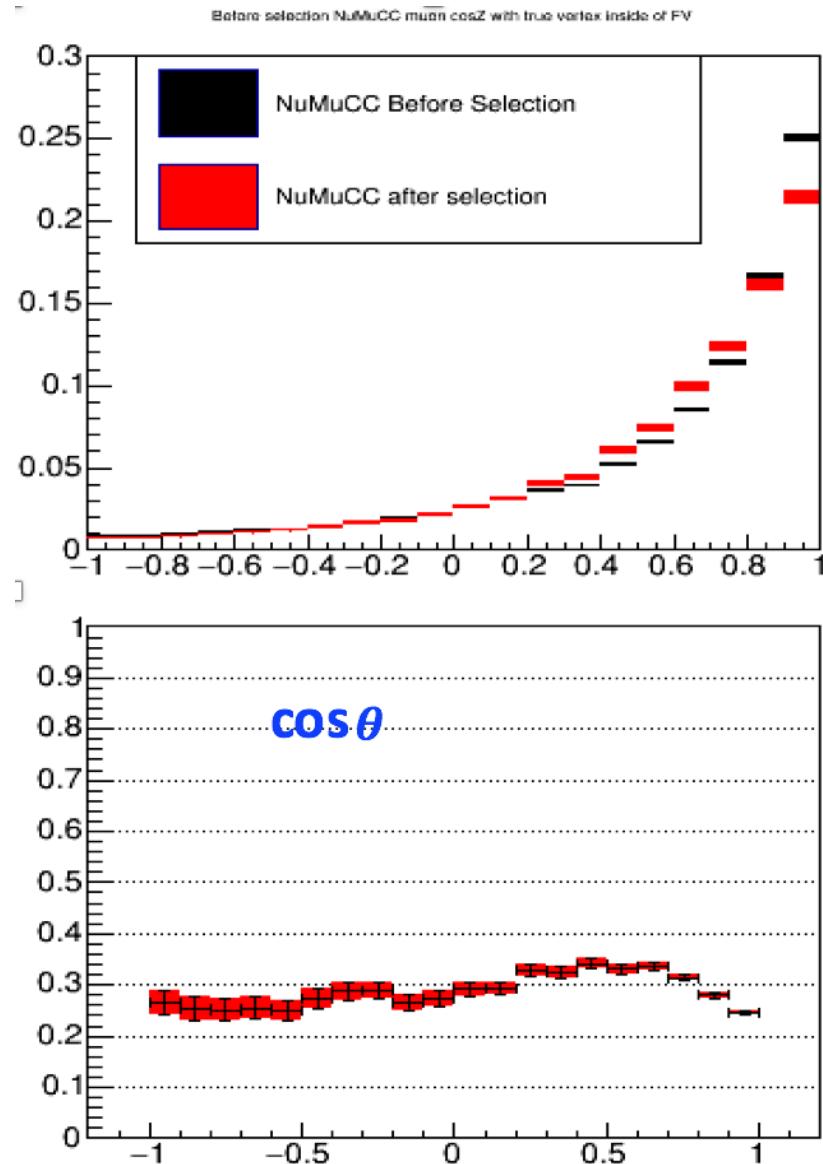


Figure 34: Top: MC truth lepton track $\cos\theta$ distribution before and after the selection. Both are shape normalized. Bottom: the efficiency of the selected true events divided by the all events with NuMuCC true vertex inside of fiducial volume. The division was done on absolute normalized distributions. Black is before the selection showing all the NuMuCC events with the true vertex inside of fiducial volume. Red is after the selection the true lepton information of the survived events.

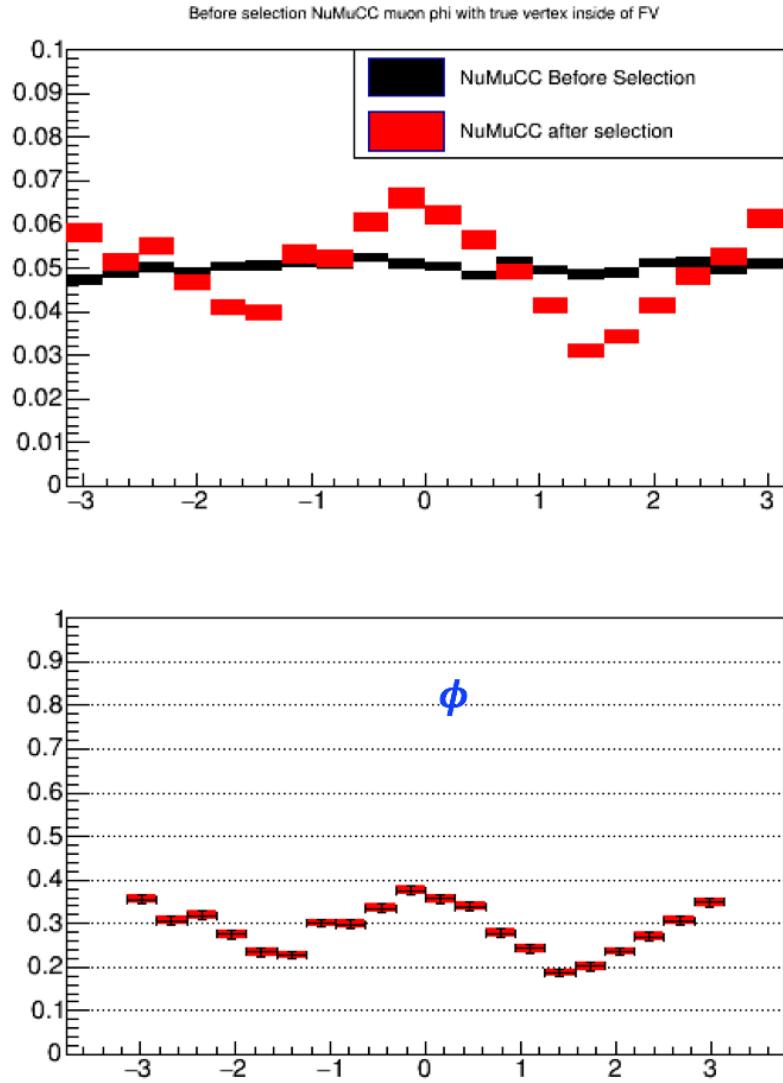


Figure 35: Top: MC truth lepton track ϕ distribution before and after the selection. Both are shape normalized. Bottom: the efficiency of the selected true events divided by the all events with NuMuCC true vertex inside of fiducial volume. The division was done on absolute normalized distributions. Black is before the selection showing all the NuMuCC events with the true vertex inside of fiducial volume. Red is after the selection the true lepton information of the survived events.

1185 **6.2.5 Passing rates for different physics processes**

1186 As stated previously, the selected ν_μ CC signal events compose of 43% quasi-
1187 elastic, 42% resonant, 14% deep inelastic, and 1% coherent. The selection effi-
1188 ciency for all the ν_μ CC events at different physics processes are CCQE: 22.3%,
1189 Resonant: 38%, DIS: 35% and Coherent: 41.1%. Please note that statistics for
1190 coherent events are very low, which is why they are not shown in the plots.
1191 Figure 36, 37, and 38 show the distributions as function of muon kinematics
1192 before and after selection broken up into physics processes, as well as efficiencies.

1193 Selection II has a smaller fraction of QE events, but a larger fraction of DIS
1194 events than selection I. The reason is that selection I requires full containment
1195 and therefore selects more lower energetic events. The lower-energetic events are
1196 typically QE. Selection II also allows exiting muon tracks. These are the high-
1197 energetic events, which are often DIS. The reason that QE events have lower
1198 selection efficiency is that they are often the events with track multiplicity 1.
1199 This is treated differently in the cuts, and one has to cut harder for single
1200 tracks to remove cosmic background. The QE efficiency decreases with muon
1201 momentum because there is a containment requirement for single tracks in both
1202 selections.

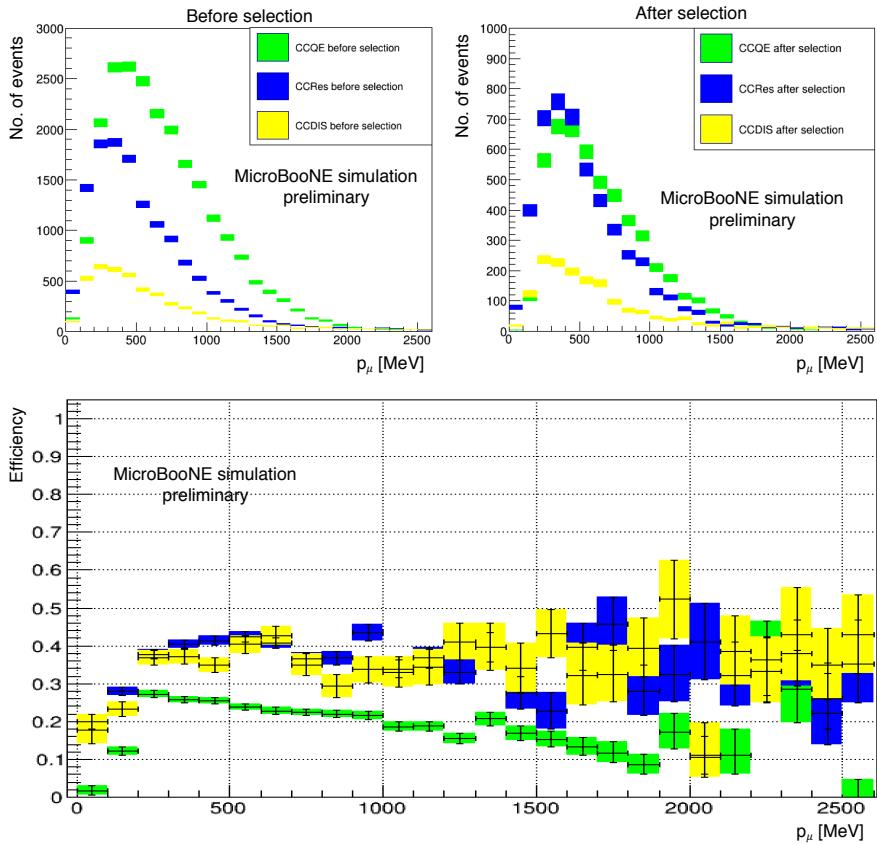


Figure 36: Top Left: MC truth lepton track momentum distribution before the selection. Top right: MC truth lepton track momentum distribution before the selection. Bottom: the efficiency of the selected events at different processes. CCQE(green), Resonant(blue), DIS(yellow). Due to very low statistics, the curves for coherent processes are not shown. Please note that the definition of QE, RES and DIS is according to GENIE and not based on e.g. a cut on the invariant mass W . This varies from generator to generator and explains why MiniBooNE (using NUANCE) reported a smaller DIS fraction ($\sim 1\%$).

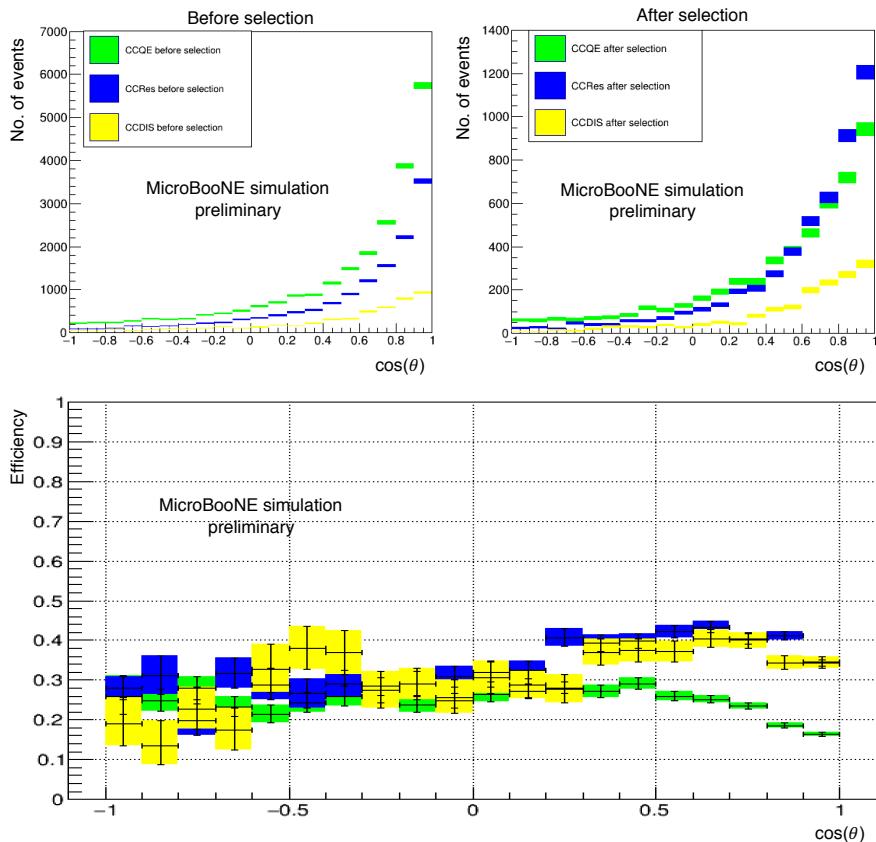


Figure 37: Top Left: MC truth lepton track $\cos\theta$ distribution before the selection. Top right: MC truth lepton track $\cos\theta$ distribution after the selection. Bottom: the efficiency of the selected events at different processes. CCQE(green), Resonant(blue), DIS(yellow). Due to very low statistics, the curves for coherent processes are not shown. Please note that the definition of QE, RES and DIS is according to GENIE and not based on e.g. a cut on the invariant mass W . This varies from generator to generator and explains why MiniBooNE (using NUANCE) reported a smaller DIS fraction ($\sim 1\%$).

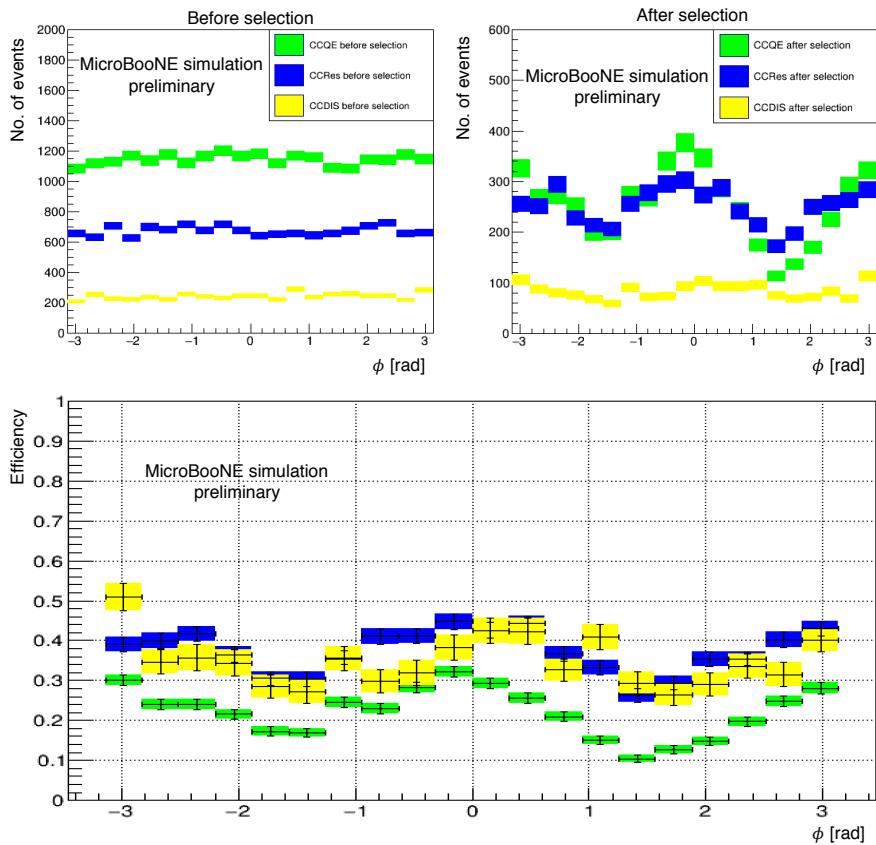


Figure 38: Top Left: MC truth lepton track ϕ distribution before the selection. Top right: MC truth lepton track ϕ distribution after the selection. Bottom: the efficiency of the selected events at different processes. CCQE(green), Resonant(blue), DIS(yellow). Due to very low statistics, the curves for coherent processes are not shown. Please note that the definition of QE, RES and DIS is according to GENIE and not based on e.g. a cut on the invariant mass W . This varies from generator to generator and explains why MiniBooNE (using NUANCE) reported a smaller DIS fraction ($\sim 1\%$).

	Sel I	Sel II	Overlap	% of Sel I	% of Sel II
Full Sample	6920	14842	2714	39.2	18.3
Signal	5440	11955	2303	42.3	19.3
Contained	6920	8597	2491	36.0	29.0
Signal+contained	5440	6299	2124	39.0	33.7
Background+contained	1480	2298	367	24.8	16.0

Table 12: Event numbers for MC selected by each selection, demonstrating the overlap between the selections.

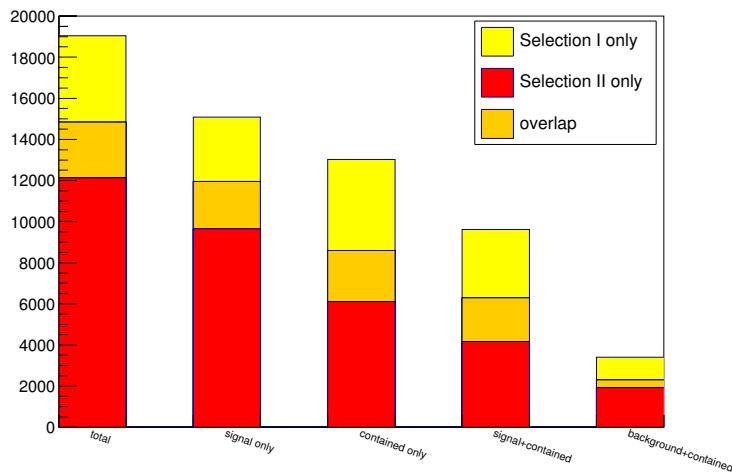


Figure 39: Event numbers exclusive to, and overlapping between, each selection. Shown for all events, signal events, contained events, the overlap of signal and contained events, and the overlap of background and contained events.

1203 6.3 Overlap between selections

1204 The two selections presented are very different, in particular with selection II
 1205 selecting both contained and uncontained events. We performed a short study
 1206 on the overlap between the two selections. Based on MC (BNB+Cosmics), we
 1207 can separate the true signal events and the contained events, and calculate the
 1208 overlap between the two selections for these categories. Table 12 shows the
 1209 numbers of events of each category selected by each event, and the size of the
 1210 overlap. Figure 39 shows the same information in graphical form. Somewhat
 1211 surprisingly, even when considering only contained signal interactions the over-
 1212 lap is still relatively small - 39% of selection I and 34% of selection II.

1213 In addition, for the events which were selected by both, we compare the
 1214 selected vertex positions. The two selections rely on different reconstruction
 1215 algorithms, so vertices are not guaranteed to be in the same place in the two.
 1216 Figure 40 shows the distribution of distances between the two reconstructed
 1217 vertex positions for events that are in both selections. Clearly most events that
 1218 overlap between the selections are passing due to the same tracks and vertices,
 1219 as the selected vertex positions are usually very close. Some of the selected

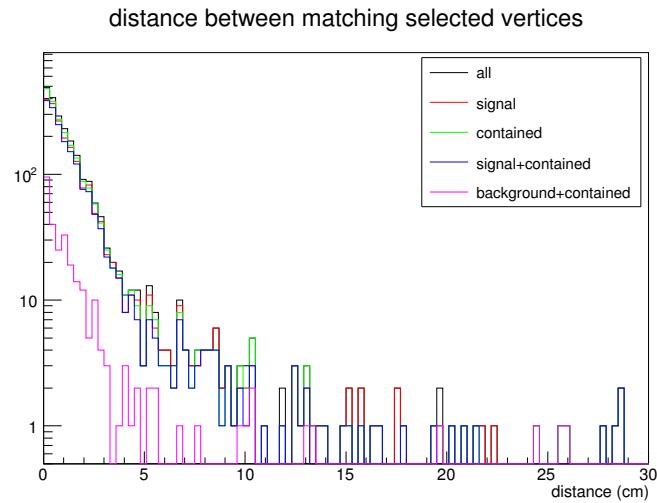


Figure 40: Distance between reconstructed vertex positions for events which were selected by both selection I and selection II. Clearly peaked at small distances indicating the same vertex is being selected.

₁₂₂₀ vertices will be, for example, at the wrong end of a track, which leads to some
₁₂₂₁ larger distances even for signal events.

1222 7 Systematic variations

1223 This chapter discusses some of the systematics that were studies in the context
1224 of this analysis. At this point in time, we do not have a complete treatment
1225 of systematic uncertainties. Therefore, we have picked individual systematic
1226 parameters which we believed to have an impact on the CC inclusive analy-
1227 sis. These parameters are discussed below. Our main interest was to find out
1228 how varying these parameters changes the muon kinematic distributions and
1229 background expectations.

1230 This chapter is discussing the systematics we studied, the technical imple-
1231 mentation, and gives an estimate of the effect on the analysis. Later in this
1232 note, these systematic uncertainties are presented as a variation of the baseline
1233 MC distributions at final cut level. These plots will be added to chapter 8.

1234 7.1 BNB neutrino Beam

1235 Different systematics from the BNB beam are included in this analysis. Such
1236 systematics include the proton delivery, the hadron production, horn current
1237 and skin depth systematical uncertainties.

1238 The beam induced hadron production systematics are obtained from the
1239 previous MiniBooNE measurements on the BNB beam simulation with studies
1240 on external data from the HARP experiment.

1241 A technote to fully describe these uncertainties is in progress [2].

1242 What is being used in this ν_μ CC inclusive study are tabulated fractional flux
1243 uncertainties as a function of the neutrino energy and flavor, which are inherited
1244 from MiniBooNE and have been published previously with MiniBooNE analyses.
1245 These are shown in table 13. The error band on final MC distributions in chapter
1246 8 will be produced by weighting events with their fractional flux uncertainty
1247 when filling the histograms of muon kinetic variables. These plots will be added
1248 in the next version of this note.

Table 13: Fractional beam flux uncertainties per neutrino flavor as a function of the neutrino energy binned in 50MeV steps.

Energy [GeV]	numu	numubar	nue	nuebar
0.0250	0.683	0.434	0.363	0.427
0.0750	0.486	0.173	0.258	0.412
0.1250	0.313	0.129	0.209	0.993
0.1750	0.253	0.096	0.173	0.965
0.2250	0.211	0.084	0.156	0.656
0.2750	0.173	0.084	0.133	0.451
0.3250	0.130	0.081	0.115	0.354
0.3750	0.095	0.078	0.099	0.261
0.4250	0.077	0.078	0.086	0.200
0.4750	0.069	0.076	0.080	0.190
0.5250	0.066	0.077	0.073	0.187
0.5750	0.061	0.076	0.070	0.185
0.6250	0.058	0.074	0.069	0.188
0.6750	0.056	0.073	0.068	0.195
0.7250	0.054	0.073	0.068	0.195
0.7750	0.053	0.074	0.068	0.189
0.8250	0.053	0.079	0.069	0.208
0.8750	0.054	0.081	0.070	0.212
0.9250	0.054	0.082	0.069	0.216
0.9750	0.055	0.085	0.071	0.207
1.0250	0.055	0.089	0.073	0.217
1.0750	0.058	0.089	0.075	0.242
1.1250	0.061	0.085	0.075	0.239
1.1750	0.062	0.083	0.075	0.247
1.2250	0.063	0.087	0.077	0.243
1.2750	0.065	0.094	0.081	0.260
1.3250	0.067	0.092	0.079	0.269
1.3750	0.071	0.094	0.081	0.273
1.4250	0.076	0.103	0.083	0.252
1.4750	0.083	0.106	0.086	0.274
1.5250	0.091	0.112	0.086	0.280
1.5750	0.098	0.107	0.088	0.245
1.6250	0.105	0.099	0.088	0.239
1.6750	0.112	0.101	0.090	0.290
1.7250	0.121	0.104	0.089	0.289
1.7750	0.131	0.121	0.091	0.238
1.8250	0.143	0.121	0.092	0.307
1.8750	0.155	0.130	0.096	0.311
1.9250	0.164	0.132	0.098	0.258
1.9750	0.173	0.125	0.090	0.303
2.0250	0.182	0.125	0.100	0.323
2.0750	0.189	0.116	0.100	0.315
2.1250	0.195	0.133	0.095	0.328
2.1750	0.200	0.147	0.103	0.307
2.2250	0.206	0.142	0.103	0.344

Table 14: Continuation of table 13. Fractional beam flux uncertainties per neutrino flavor as a function of the neutrino energy binned in 50MeV steps.

Energy [GeV]	numu	numubar	nue	nuebar
2.2750	0.197	0.196	0.106	0.315
2.3250	0.192	0.236	0.106	0.353
2.3750	0.187	0.254	0.104	0.367
2.4250	0.176	0.294	0.106	0.357
2.4750	0.166	0.310	0.090	0.359
2.5250	0.159	0.321	0.101	0.344
2.5750	0.152	0.305	0.114	0.378
2.6250	0.148	0.301	0.113	0.385
2.6750	0.143	0.300	0.110	0.383
2.7250	0.142	0.288	0.109	0.372
2.7750	0.140	0.254	0.107	0.393
2.8250	0.139	0.230	0.109	0.389
2.8750	0.141	0.352	0.117	0.411
2.9250	0.139	0.282	0.111	0.410
2.9750	0.139	0.306	0.107	0.388
3.0250	0.139	0.360	0.114	0.328
3.0750	0.138	0.412	0.117	0.410
3.1250	0.136	0.546	0.115	0.416
3.1750	0.133	0.371	0.104	0.400
3.2250	0.134	0.402	0.133	0.416
3.2750	0.133	0.200	0.106	0.400
3.3250	0.131	0.349	0.108	0.397
3.3750	0.131	0.321	0.119	0.459
3.4250	0.129	0.436	0.122	0.436
3.4750	0.126	0.287	0.102	0.431
3.5250	0.127	0.239	0.119	0.453
3.5750	0.124	0.246	0.153	0.435
3.6250	0.124	0.465	0.117	0.467
3.6750	0.123	0.272	0.115	0.129
3.7250	0.125	0.210	0.124	0.461
3.7750	0.122	0.535	0.131	0.427
3.8250	0.121	0.143	0.110	0.467
3.8750	0.122	0.347	0.109	0.433
3.9250	0.121	0.232	0.120	0.437
3.9750	0.125	0.359	0.123	0.423
4.0250	0.118	0.340	0.135	0.488
4.0750	0.119	0.325	0.108	0.454
4.1250	0.118	0.410	0.105	0.413
4.1750	0.119	0.667	0.133	0.660
4.2250	0.118	0.324	0.595	0.307
4.2750	0.117	0.365	0.135	0.545
4.3250	0.115	0.355	0.120	0.467
4.3750	0.116	0.302	0.132	0.546
4.4250	0.118	0.213	0.136	0.492
4.4750	0.115	0.444	0.148	0.329

Table 15: Continuation of table 13 and 14. Fractional beam flux uncertainties per neutrino flavor as a function of the neutrino energy binned in 50MeV steps.

Energy [GeV]	numu	numubar	nue	nuebar
4.5250	0.120	0.494	0.132	0.491
4.5750	0.116	0.271	0.127	0.549
4.6250	0.117	0.337	0.114	0.552
4.6750	0.114	0.357	0.164	0.592
4.7250	0.130	0.279	0.145	0.811
4.7750	0.116	0.091	0.133	0.372
4.8250	0.116	0.478	0.107	0.583
4.8750	0.118	0.227	0.297	0.472
4.9250	0.115	0.619	0.194	0.637
4.9750	0.117	0.194	0.120	0.703

1249 **7.2 Model systematics**

1250 **7.2.1 Model variations through reweighting**

1251 GENIE provides a built-in framework of event re-weighting for evaluating sys-
1252 tematic uncertainties in an analysis [3, 4]. Given a certain physics parameter P
1253 with estimated prior uncertainty δP , we are interested in the effect on the final
1254 cross section if this parameter is changed to $P' = P + x_P \delta P$. x_P is a systematic
1255 parameter. For now, we are only considering $x_P = \pm 1$, i.e. we are changing the
1256 physics quantities by $\pm 1\sigma$ (1σ as defined within GENIE, see table 16).

1257 The GENIE re-weighting framework is pretty straightforward to use. The user
1258 needs to feed GENIE with a GENIE file with nominal values of the parameters,
1259 plus the values chosen for x_P . The output returned by GENIE is a set
1260 of event weights; when applied to each event reproduces the distribution that
1261 would have been obtained had you run GENIE with those parameters to begin
1262 with¹. The MC used has to have sufficient statistics for reweighting to apply,
1263 otherwise larger variations will be seen. A list of the systematic parameters is
1264 given in Table 16.

1265 In LArSoft, we use the `EventWeight` framework to interface with GENIE.
1266 `EventWeight` is a generic platform with tools to create independent algorithms
1267 for different systematics (doc-db 4318). The code is available on the feature
1268 branch `feature/Anne_eventweight`.

1269 The code to analyze the output of the GENIE re-weighting package is avail-
1270 able on github, and the instructions on how to use it are on: <https://github.com/marcodeltutto/EvtWgtAnalysis>.

1272 The LArSoft version used for this analysis is v04_30_03. All the plots shown
1273 in this section were made running over MCC6 Monte Carlo files and correspond
1274 to 1.22×10^{20} simulated POT. This MC is older than the MCC7 MC used
1275 for the rest of this note. However, it is sufficient to illustrate the effect of the
1276 systematics. No quantitative statements derived from MCC6 go into the ν_μ CC
1277 inclusive analysis.

1278 Figure 41 shows how the number of events changes if one parameter is in-
1279 creased or decreased by $\pm 1\sigma$ as a function of the outgoing muon momentum.
1280 This Figure just shows an example for two cases: axial mass for CC quasi elastic
1281 and for CC resonance neutrino interactions. This means that, if we take the plot
1282 41a, the re-weighting happened only if the interaction was a CC quasi-elastic.
1283 All other events have a weight equal to one. A comprehensive list of plots can
1284 be found in doc-db 5668.

1285 As a further step, we also looked at the number of background events and
1286 how this number is affected by the GENIE model parameters. To accomplish
1287 this, we used the same selection code that was used for the “MC performance
1288 study for an early ν_μ CC inclusive analysis with MicroBooNE” (doc-db 4925).
1289 The selection used in this technical note is not the same selection as described
1290 in Chapter 6. The reason for this is that the current event selection wasn’t
1291 ready when these systematics studies were performed. All these studies will be
1292 updated in the future for the final analysis, although no major differences are
1293 expected.

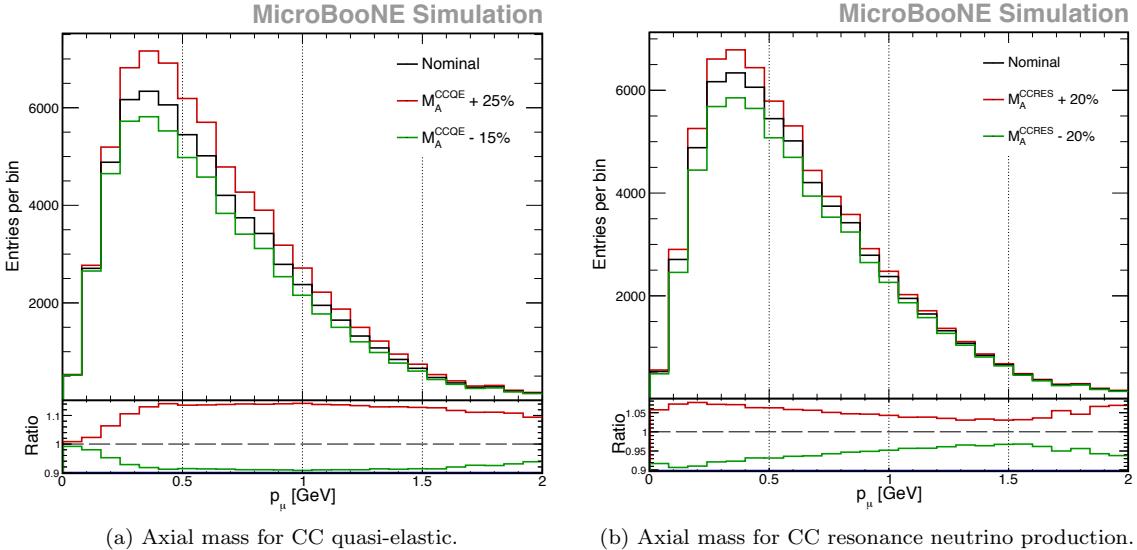
1294 The results are shown in Figure 42. This Figure shows the events that passed

¹Assuming the phase space coverage with the new parameter is the same or smaller. One cannot reweight events into existence

Table 16: Neutrino interaction model parameters and uncertainties. This information is reproduced here from the GENIE User Manual [4] for convenience.

For the non-resonance background, we are considering the following combinations: `NonResRvp1pi` ($\nu + p$ and $\bar{\nu} + n$ (1 π) type interactions), `NonResRvbarp1pi` ($\nu + n$ and $\bar{\nu} + p$ (1 π) type interactions), `NonResRvp2pi` ($\nu + p$ and $\bar{\nu} + n + n$ (2 π) type interactions), `NonResRvbarp2pi` ($\nu + n$ and $\bar{\nu} + p$ (2 π) type interactions).

Parameter P	Description of P	Value	$\delta P/P$
M_A^{NCEL}	Axial mass for NC elastic	0.990 GeV	$\pm 25\%$
η^{NCEL}	Strange axial form factor η for NC elastic	0.120 GeV	$\pm 30\%$
M_A^{CCQE}	Axial mass for CC quasi-elastic	0.990 GeV	$-15\% + 25\%$
M_A^{CCRES}	Axial mass for CC resonance neutrino production	1.120 GeV	$\pm 20\%$
M_V^{CCRES}	Vector mass for CC resonance neutrino production	0.840 GeV	$\pm 10\%$
M_A^{NCRES}	Axial mass for NC resonance neutrino production	1.120 GeV	$\pm 20\%$
M_V^{NCRES}	Vector mass for NC resonance neutrino production	0.840 GeV	$\pm 10\%$
$M_{COH\pi}^A$	Axial mass for CC and NC coherent pion production	1.000 GeV	$\pm 50\%$
$R_0^{COH\pi}$	Nuclear size param. controlling π absorption in RS model	1.000 fm	$\pm 10\%$
CCQE-PauliSup (p)	CCQE Pauli suppression (via changes in Fermi level k_F)	0.242 GeV	$\pm 35\%$
CCQE-PauliSup (n)	CCQE Pauli suppression (via changes in Fermi level k_F)	0.259 GeV	$\pm 35\%$
A_{HT}^{BY}	A_{HT} higher-twist param in BY model scaling variable ξ_w	0.538	$\pm 25\%$
B_{HT}^{BY}	B_{HT} higher-twist param in BY model scaling variable ξ_w	0.305	$\pm 25\%$
C_{V1u}^{BY}	C_{V1u} u valence GRV98 PDF correction param in BY model	0.291	$\pm 30\%$
C_{V2u}^{BY}	C_{V2u} u valence GRV98 PDF correction param in BY model	0.189	$\pm 40\%$
FZ (pion)	Hadron formation zone	0.342 fm	$\pm 50\%$
FZ (nucleon)	Hadron formation zone	2.300 fm	$\pm 50\%$
$BR(\gamma)$	Branching ratio for radiative resonance decays	-	$\pm 50\%$
$BR(\eta)$	Branching ratio for single- η resonance decays	-	$\pm 50\%$
$R_{\nu p}^{CC1\pi}$	Non-resonance bkg in νp CC1 π reactions	-	$\pm 50\%$
$R_{\nu p}^{CC2\pi}$	Non-resonance bkg in νp CC2 π reactions	-	$\pm 50\%$
$R_{\nu n}^{CC1\pi}$	Non-resonance bkg in νn CC1 π reactions	-	$\pm 50\%$
$R_{\nu n}^{CC2\pi}$	Non-resonance bkg in νn CC2 π reactions	-	$\pm 50\%$
$R_{\nu p}^{NC1\pi}$	Non-resonance bkg in νp NC1 π reactions	-	$\pm 50\%$
$R_{\nu p}^{NC2\pi}$	Non-resonance bkg in νp NC2 π reactions	-	$\pm 50\%$
$R_{\nu n}^{NC1\pi}$	Non-resonance bkg in νn NC1 π reactions	-	$\pm 50\%$
$R_{\nu n}^{NC2\pi}$	Non-resonance bkg in νn NC2 π reactions	-	$\pm 50\%$
x_{abs}^N	Nucleon mean free path (total rescattering probability)	-	$\pm 20\%$
x_{ce}^N	Nucleon charge exchange probability	-	$\pm 50\%$
x_{el}^N	Nucleon elastic reaction probability	-	$\pm 30\%$
x_{inel}^N	Nucleon inelastic reaction probability	-	$\pm 40\%$
x_{mfp}^N	Nucleon absorption probability	-	$\pm 20\%$
x_{pi}^N	Nucleon π -production probability	-	$\pm 20\%$
x_{PI}^{abs}	π mean free path (total rescattering probability)	-	$\pm 20\%$
x_{PI}^{ce}	π charge exchange probability	-	$\pm 50\%$
x_{PI}^{el}	π elastic reaction probability	-	$\pm 10\%$
x_{PI}^{inel}	π inelastic reaction probability	-	$\pm 40\%$
x_{PI}^{mfp}	π absorption probability	-	$\pm 20\%$
x_{PI}^{pi}	π π -production probability	-	$\pm 20\%$



(a) Axial mass for CC quasi-elastic.

(b) Axial mass for CC resonance neutrino production.

Figure 41: The black distributions show the number of events as a function of muon momentum for nominal values of GENIE parameters. The red and green distributions show the number of events with the axial mass for CC quasi-elastic, 41a, and CC resonance interaction, 41b, changed by $\pm 1\sigma$ respectively.

1295 the selection cuts. The solid line shows the distribution for the nominal value
 1296 of the GENIE parameters, the dashed line if the parameter is increased by 1σ
 1297 and the dotted line if decreased by 1σ .

1298 Other than the number of events, we also want to understand how the effi-
 1299 ciency is affected by the uncertainties on the GENIE model parameters. Using
 1300 the same selection code, the efficiency is shown in Figure 43 as a function of
 1301 true muon momentum. Again, a comprehensive list of plots can be found in
 1302 doc-db 5668.

1303

1304 The final goal is to understand the effect that these uncertainties have on the
 1305 final cross-section. For now, we are only interested in understanding how the
 1306 cross-section changes if we change the parameters by $\pm 1\sigma$. Such cross-sections
 1307 will be:

$$\sigma^\pm \propto \frac{N - B^\pm}{\epsilon^\pm}$$

1308 where N is the number of selected events, B^\pm is the number of background
 1309 events (form Fig. 42, considering $\pm 1\sigma$ tweaked parameters), and ϵ^\pm is the
 1310 efficiency (form Fig. 43, again considering $\pm 1\sigma$ tweaked parameters). We are
 1311 not varying the number of events since this will be a measured quantity in the
 1312 final analysis. We now have all the ingredients to calculate the variation of the
 1313 cross-section w.r.t. to the nominal values of the parameters:

$$\text{Percental difference} = \left| \frac{\sigma - \sigma^\pm}{\sigma} \right|$$

1314 This percental difference can be calculated for each systematic parameter. This

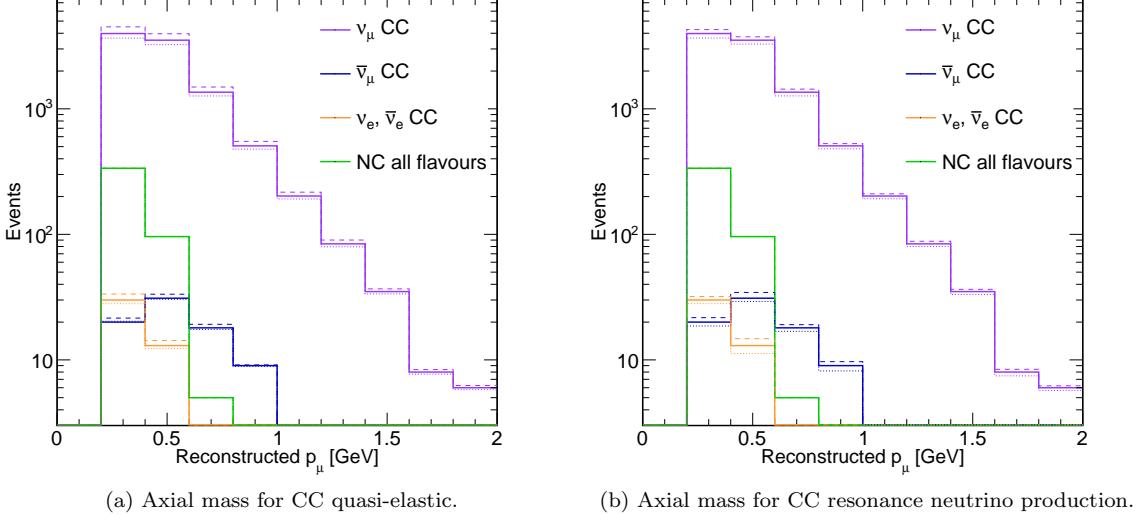


Figure 42: Distribution of events that pass the selection cuts. Events are shown as a function of the reconstructed muon momentum. For background events, the length of the selected candidate track is used to calculate the momentum under the assumption that it is a muon. Solid line shows the distribution for the nominal value of the GENIE parameters. Dashed line if the parameter is increased by 1σ and dotted line if decreased by 1σ . Color shows different types of events. The ν_μ CC signal events are shown in violet. The background events are: blue for $\bar{\nu}_\mu$ CC events, yellow for ν_e and $\bar{\nu}_e$ CC events and green for NC events. Only two examples are reported here: axial mass for CC quasi-elastic (left), and for CC resonance interaction (right).

1315 allows to understand which parameter has the biggest impact on the final sys-
 1316 tematic uncertainty and which one can be neglected.

1317 We also want to show the comparison of this percental difference with the
 1318 statistical uncertainty for an exposure of 0.5×10^{20} and 6.6×10^{20} POT. The
 1319 statistical uncertainty shown in the following is calculated as:

$$\text{Fractional stat. uncert.} = \frac{\sqrt{S+B}}{S}$$

1320 where S and B are the number of signal and background selected events respec-
 1321 tively. Cosmic background is not taken into account at this point. If included,
 1322 the fractional statistical uncertainty can only increase. Our goal is to show
 1323 that GENIE model uncertainties are well below the statistical one for 0.5×10^{20}
 1324 POT, and this is true even if cosmic background is neglected (see following).

1325 The cross-section variation can be seen in Figure 44. The two black dis-
 1326 tributions show the statistical uncertainty as a function of muon momentum
 1327 (solid for 6.6×10^{20} POT and dashed for 0.5×10^{20} POT). Colored curves show
 1328 the cross-section percental difference as a function of muon momentum for six
 1329 different systematic parameters. Instead of plotting two curves, one for $+1\sigma$ and
 1330 the other for -1σ reweighting, the maximum between these two is shown bin by
 1331 bin. Although the total number of systematic parameters is much higher than

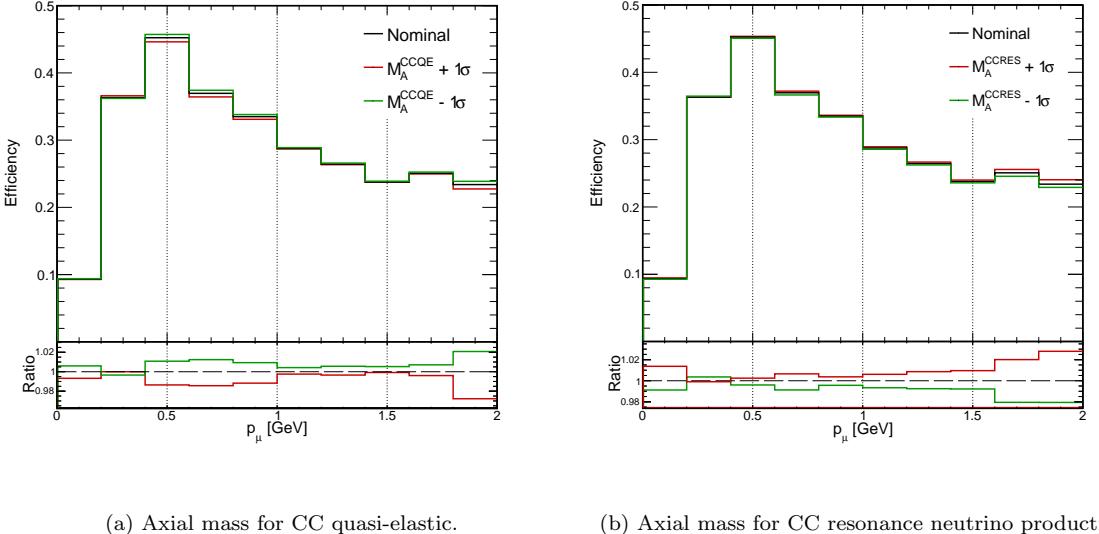


Figure 43: Efficiency of the event selection as a function of the true muon momentum. Red and green distributions show number of events with the axial mass for CC quasi-elastic (left), and CC resonance interaction (right), changed by $+1\sigma$ and -1σ respectively.

1332 six, we chose to show only these because the others are negligible. The two red
 1333 distributions show the effect of the axial mass uncertainty for CC interactions:
 1334 solid for quasi-elastic and dashed for resonance neutrino production. The two
 1335 violet distributions show the effect of the axial mass uncertainty for NC interac-
 1336 tions: solid for elastic and dashed for resonance neutrino production. The green
 1337 distribution is the effect of the Pauli suppression (with changes in the Fermi
 1338 motion). Finally, the blue distribution is the effect of the uncertainty on the
 1339 hadron formation zone.

1340 All the cross-section percental differences are well below the statistical un-
 1341 certainty at 0.5×10^{20} POT and are comparable with the one at 6.6×10^{20} POT
 1342 only in the low momentum bins.

1343 Figure 44 only shows the effect of some GENIE parameters. The plots of
 1344 the cross-section percental difference for all parameters, as a function of muon
 1345 momentum, can be found in Appendix A. A summary is shown in Table 17.
 1346 This table shows the maximum value of the cross-section percental difference
 1347 among the p_μ bins. It also shows the value of the statistical uncertainty in that
 1348 p_μ bin for an exposure of 0.5×10^{20} POT.

1349 Modified Nucleon Axial Mass

1350 GENIE uses a value of $M_A = 0.99$ GeV for the axial vector mass for QE pro-
 1351 cesses. By re-weighting the BNB+Cosmic MC sample one can study the effects
 1352 of a different axial mass. Here, $M_A = 1.35$ GeV (motivated by initial Mini-
 1353 BooNE measurements [5]) was chosen. Figure 46 shows the baseline MC in red
 1354 and the one with the modified M_A in green. These distributions are shown

Table 17: Values of the cross-section percental difference for the different GENIE parameters. First column shows the cross-section percental difference in the p_μ bin where such difference is maximum. Second column where the cross-section is maximum (that corresponds to the [0.2,0.4] GeV bin in Fig. 42). Last two lines show the value of the fractional statistical uncertainty for an exposure of 0.5 and 6.6×10^{20} POT. Fractional statistical uncertainty calculated as $\sqrt{S+B}/S$. The percental difference for $M_A^{COH\pi}$ and $R_0^{COH\pi}$ doesn't really reflect the real effect on the cross-section: the high value that they have is probably due to the few coherent events in our simulation; the re-weighting procedure only works if we have a large enough statistical sample.

	Cross-section percental difference at maximum point	Cross-section percental difference at cross-section peak
M_A^{NCEL}	1.1707	1.1707
η^{NCEL}	0.0683733	0.0683733
M_A^{CCQE}	1.37874	0.858075
M_A^{CCRES}	1.00678	0.226237
M_V^{CCRES}	0.586791	0.142185
M_A^{NCRES}	1.2378	1.2378
M_V^{NCRES}	0.366465	0.366465
$M_A^{COH\pi}$	(52.1012)	(1.94777)
$R_0^{COH\pi}$	(54.1199)	(1.97408)
CCQE-PauliSup	1.04368	0.721536
A_{HT}^{BY}	0.0316585	0.0316585
B_{HT}^{BY}	0.0480709	0.0480709
C_{V1u}^{BY}	0.0246201	0.0185337
C_{V2u}^{BY}	0.0214913	0.0161913
FZ	1.11864	0.661157
$BR(\eta)$	0.213923	0.191805
$BR(\gamma)$	0.0226982	0.0170411
NonResRvbarp1pi	1.22076	0.400095
NonResRvbarp2pi	0.275822	0.275822
NonResRvp1pi	0.557264	0.0760578
NonResRvp2pi	0.506743	0.261474
x_{abs}^N	0.469734	0.469734
x_{cex}^N	0.863909	0.183535
x_{el}^N	1.37074	0.542205
x_{inel}^N	0.934414	0.11268
x_{mfp}^N	0.690747	0.403551
x_{pi}^N	0.0977966	0.0960583
x_{abs}^{PI}	0.367807	0.367807
x_{cex}^{PI}	0.454639	0.14336
x_{el}^{PI}	0.156806	0.156806
x_{inel}^{PI}	0.781218	0.0466197
x_{mfp}^{PI}	0.90829	0.280258
x_{pi}^{PI}	0.0599161	0.0326048
Stat. @ 0.5e20 POT	-	2.57952
Stat. @ 6.6e20 POT	-	0.70999

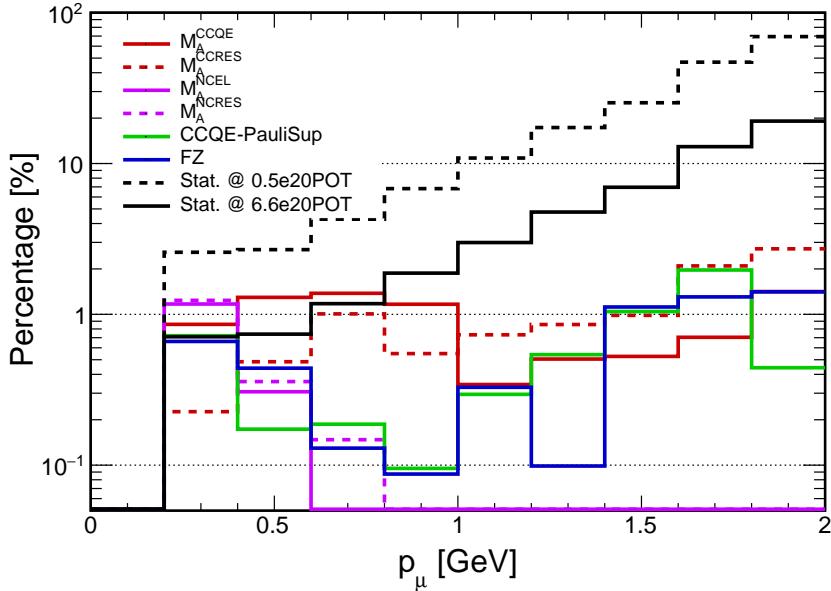


Figure 44: Cross-section variation as a function of true muon momentum. Also statistical uncertainty is shown in solid black for 6.6×10^{20} POT and in dashed black for 0.5×10^{20} POT. This plot does not show the systematic uncertainty on the final cross-section, but how the measured cross-section changes if one systematic parameter is changed by $\pm 1\sigma$.

for muon track length, $\cos\theta$ and ϕ angle. They are shown for both selection I and II. All the distributions were scaled to the same number of simulated POT: 2.2×10^{20} .

7.2.2 Model changes through re-generation

There are model changes which cannot be done through re-weighting. In particular if changing between two models which have different phase-space coverage.

Recent MiniBooNE CCQE measurements [5] showed an enhancement over a simple Fermi gas model of the nucleus. One explanation of this excess is an additional process, where a neutrino interacts with a correlated pair of nucleons (which may be in a quasi-deuteron state). As the correlation between nucleons is mediated by the exchange of a meson, this process is known as MEC (Meson Exchange Current). It is also often referred to as 2p2h, or multinucleon interactions. In MiniBooNE, which was unable to detect most protons from CCQE events, this would look identical to a CCQE event.

GENIE has two models for this process, a microscopic model for MEC by Nieves *et al.*, and the Transverse-Enhancement Model (TEM) which comes paired with an Effective Spectral Function (ESF) model for nuclear motion and binding energies. The TEM model simply enhances the CCQE interaction strength in various regions (the transverse form factors) according to electron scattering data fits. There are no additional nucleons ejected. These additional models were introduced in GENIE v2.10.0, so new MC productions were made

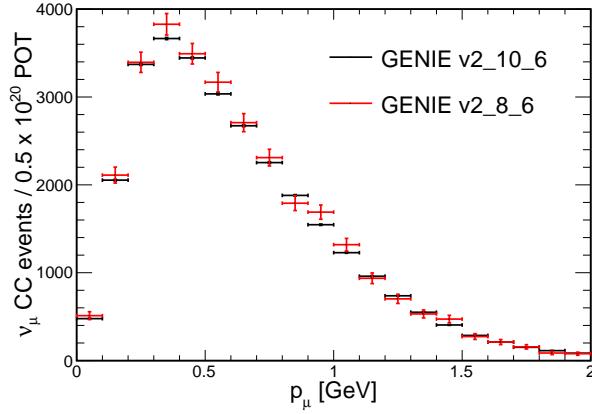


Figure 45: Muon momentum distribution for ν_μ CC events as simulated by GENIE with the parameters set to their nominal values. The black distribution was made with GENIE v2.10.6 while the red one with v2.8.6. GENIE v2.8.6 was used throughout this technical note, but v2.10.6 is going to be used to generate new samples with MEC and ESF (see text).

1376 with this version.

1377 To check that there were no other changes between versions, a set of events
1378 were produced with the default settings to compare to nominal MC, see Figure
1379 45. We are currently running two full productions with 200k neutrino interac-
1380 tions, one with MEC included, and one with the TEM+ESF model, both used
1381 GENIE 2.10.6.

1382 These samples were ran through the event selections and compared to final
1383 kinematic distributions with the baseline MC. These comparison plots can be
1384 seen in Figure 47. The red distribution shows the baseline MC, the green one is
1385 the baseline with the addition of MEC events and the blue one is with ESF and
1386 TEM. These distributions are shown for muon track length, $\cos \theta$ and ϕ angle.
1387 They are shown for both selection I and II. All the distributions were scaled to
1388 the same number of simulated POT: 2.2×10^{20} .

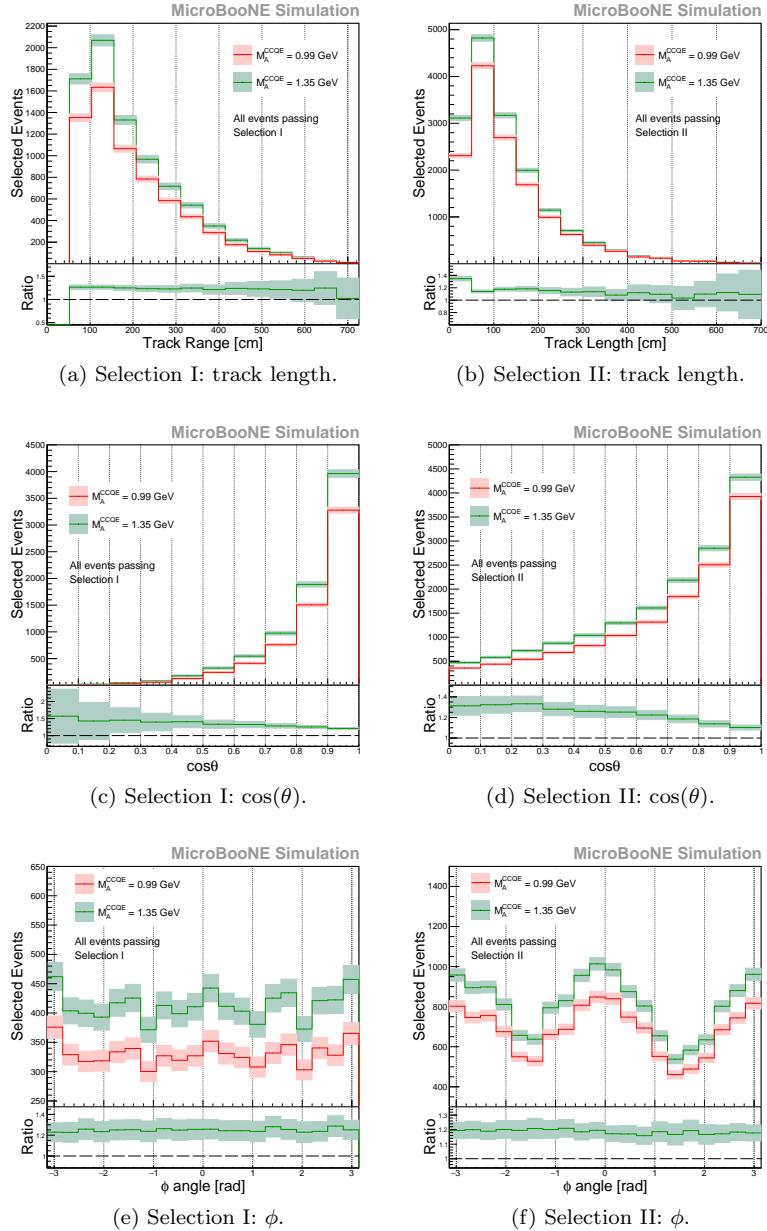


Figure 46: The effect of re-weighting the axial vector mass for quasi-elastic processes on the expected muon kinematic distributions for selection I (left) and selection II (right). The red distribution shows the baseline simulation of BNB+Cosmic events, using $M_A = 0.99$ GeV. The green curve shows the expected distribution using $M_A = 1.35$ GeV, as motivated by the MiniBooNE measurements [5]. The shaded areas show the statistical uncertainty only. The plots are scaled to 2.2×10^{20} . The track range is defined as the 3D distance between the start and end point of the muon candidate track. The angle θ is defined as the angle in beam direction (with $\theta = 0$ beam direction). ϕ is the angle around beam direction with $-\pi/2$ being vertically downward going and $\pi/2$ being vertically upward going tracks.

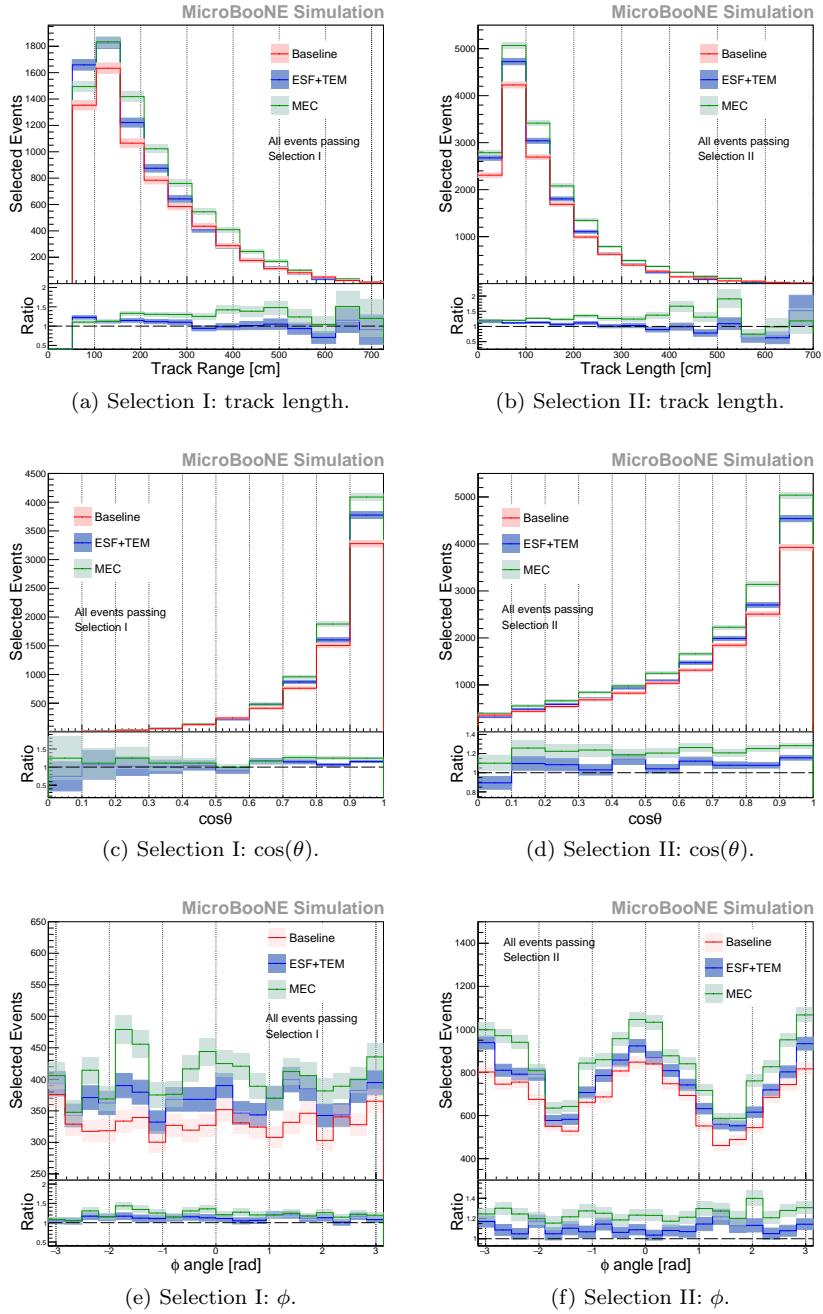


Figure 47: The effect of using different nuclear models in the neutrino simulation on the expected muon kinematic distributions for selection I (left) and selection II (right). Green distribution shows the effect of simulating Meson Exchange Current (MEC) interactions in addition to the GENIE nominal ones (CC, RES, DIS, etc.). The blue distribution shows the effect of using the Effective Spectral Function model (in place of the GENIE nominal Bodek-Ritchie Fermi Gas model) and the Transverse Enhancement model (which is a simplified version of MEC). The shaded areas show the statistical uncertainty only. The plots are scaled to 2.2×10^{20} . The track range is defined as the 3D distance between the start and end point of the muon candidate track. The angle θ is defined as the angle in beam direction (with $\theta = 0$ beam direction). ϕ is the angle around beam direction with $-\pi/2$ being vertically downward going and $\pi/2$ being vertically upward going tracks.

1389 **7.3 Detector systematics**

1390 The drift of free electrons in liquid argon is affected by many detector parameters such as non-uniform electric field, electron lifetime and diffusion. For this
1391 analysis, the following two quantities were studied as part of detector system-
1392 atics:

- 1394 • Electron lifetime: Electro-negative impurities such as O₂ and H₂O present
1395 in trace-levels in argon can absorb ionization electrons causing signal at-
1396 tenuation. The electron lifetime which represents the half life of electrons
1397 in argon is inversely proportional to the impurity concentration in the liq-
1398 uid argon. In this note we study the impact of argon purity on the analysis
1399 by varying the electron lifetime around the baseline value of 8 ms.
- 1400 • Space charge: This represents the accumulation of slow-moving positive
1401 ions in the detector due to ionizing tracks crossing the detector from var-
1402 ious sources. Since MicroBooNE is located close to the surface without
1403 any significant overburden, the high cosmic muon event rate is expected
1404 to create a significant amount of space charge that could cause distortions
1405 in the electric field. This results in distortions in the reconstructed trajec-
1406 tories of particles that travel through the active volume of the TPC. We
1407 turn on the space charge effect in MC samples to study the impact of space
1408 charge on muon tracking efficiency and true muon kinematic distributions.

1409 A set of six BNB plus Cosmic MC samples with \sim 20,000 events were generated
1410 in order to study the impact of the above two parameters:

- 1411 • Electron lifetime of 3 ms, space charge off
1412 • Electron lifetime of 6.4 ms, space charge off
1413 • Electron lifetime of 10 ms, space charge off
1414 • Electron lifetime of 8 ms, space charge off (baseline)
1415 • Electron lifetime of 8 ms, space charge on
1416 • Electron lifetime of 6.4 ms, space charge on
1417 • Electron lifetime of 10 ms, space charge on

1418 The samples that do not include space charge simulation are produced using
1419 larsoft version v04_36_00_03 from generation through detector simulation steps
1420 and larsoft version v05_08_00 for reconstruction and analysistree steps. The
1421 samples that include space charge simulation use larsoft version v05_08_00 for all
1422 production steps. The 8 ms electron lifetime is considered the nominal lifetime.
1423 The lifetime variation (6.4 ms and 10 ms) used to represent the low and high
1424 limits is based on the fluctuations observed in the purity monitor data. The 3 ms
1425 lifetime sample was generated mainly to understand whether we can include data
1426 with 3 ms lifetime (as indicated by the purity monitor) in our final good runs
1427 list.

1428 In order to verify that the space charge and lifetime effects were successfully
1429 implemented in the simulation samples, we produced some plots as a sanity
1430 check that demonstrate that the effects were indeed simulated. One of the

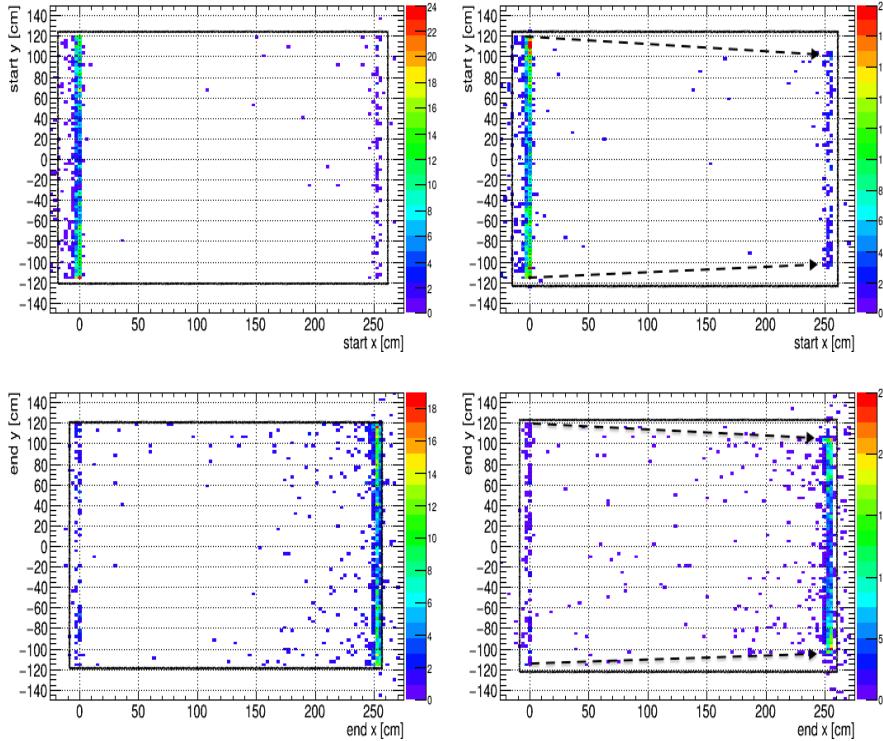


Figure 48: Distribution of track start (top row) and end (bottom row) points in Y as a function of track start/end points in X with (right) and with out (left) space charge effect. The plots are made using tracks that cross both anode and cathode planes as t_0 for these is known. The electron lifetime value in these samples corresponds to 10 ms.

effects of space charge is displacement of track ends from the edges of the TPC in Y and Z, particularly noticeable for high values of X. Figure 48 shows the distribution of track end points in Y as a function of X distance for tracks that cross both cathode and anode planes. On the right hand side plots, one can clearly see the effect of space charge at high values of X. These plots indicate that the space charge effect was simulated in our samples. The effect of lifetime can be seen by plotting the charge per track pitch (dQ/ds) as a function of drift time. For infinite lifetime, this distribution will be flat indicating no charge is lost during drift. For smaller lifetimes and larger drift distances, the attenuation will be large. Figure 49 shows how the dQ/ds vs drift time distribution varies for different lifetime samples using tracks that cross both anode and cathode.

7.3.1 Impact on muon tracking efficiency

Both Electron lifetime and space charge directly impact the signal reconstruction. Low electron lifetime results in signal attenuation and space charge can result in distortions in the reconstructed position of ionization electrons recorded by the anode wire planes. One way to understand the impact of lifetime and

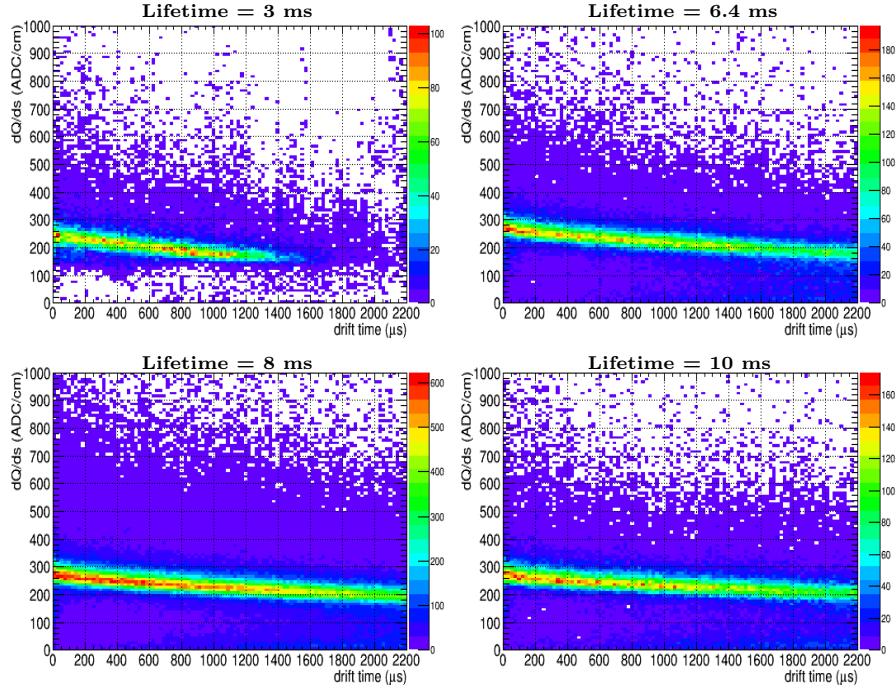


Figure 49: Charge per track pitch distribution as a function of drift time for samples corresponding to different electron lifetime values. The plots are made using tracks that cross both anode and cathode planes as t_0 of these tracks is known. The space charge effect was turned off in all of these samples.

space charge is to look at muon tracking efficiency. Figure 50 shows the muon tracking efficiency as a function of reconstructed muon momentum for track-kalmanhit (left column) and pandoraNu (right column) tracks. The plots in the top row show how the variation in lifetime effects the muon tracking efficiency. The plots in the middle row show the effect of space charge (with a fixed electron lifetime of 8 ms) on muon tracking efficiency. The plots in the bottom row show the effect of both lifetime and space charge on muon tracking efficiency. One can conclude from these plots that the muon tracking efficiency is not significantly affected by turning on the space charge effects or varying the lifetime from 8 ms to 6.4 ms or 10 ms. The impact of 3 ms lifetime on the tracking efficiency and an extended discussion on how this impacts the CC inclusive analysis is presented in sub-section 7.3.2.

It is important to note that the detector physics and calibration parameters (and their implementation) are not fully developed in MicroBooNE at the time of the writing of this note. Since the detector physics parameters implemented in MC do not accurately represent data, the studies presented in this section are to be considered preliminary. More sophisticated studies to estimate detector systematics will be performed in the near future as the detector physics and calibration systems mature in MicroBooNE. Based on the preliminary studies presented in this section and given the current implementation of the MicroBooNE calibration parameters, we conclude that when compared to

1468 the nominal simulation, the space charge or the 6.4 ms and 10 ms variation in
1469 lifetime do not have significant impact on the muon tracking efficiency in the
1470 way we are currently modeling these effects in our simulation. More detailed
1471 studies are ongoing.

1472 **7.3.2 Impact of 3 ms lifetime sample on data**

1473 Based on MC studies, the plots in Fig. 50 (top row) show that the muon tracking
1474 efficiency reduces significantly with 3 ms lifetime. Fig. 49 (top left) also indicates
1475 that the charge attenuation is larger for the 3 ms lifetime sample compared to
1476 other relatively higher lifetime samples. From Fig. 51 (left), one can see that
1477 the average number of tracks per event for the 3 ms lifetime sample is reduced
1478 by about $\sim 25\%$ compared to the 8 ms sample.

1479 As a next step, we looked at similar quantities in data to understand if
1480 we see similar deviations in data during the lower purity runs as indicated by
1481 the purity monitor. The top (bottom) plot in Figure 52 shows the average hit
1482 area charge (hit peak charge) per run as a function of the run number and one
1483 can clearly see the range of runs (between runs 5650 and 5750) where both
1484 quantities undergo attenuation due to lower lifetime. The lifetime measurement
1485 by the purity monitor for these set of runs are shown in Fig. 53. While the data
1486 shows deviations due to lower purity in the hit charge distributions, the average
1487 number of tracks per event distribution remained stable during this period as
1488 indicated in Figure 54 (top). This contradicts what we saw in Monte Carlo
1489 studies, see Fig. 51 (left) and Fig. 50 (top row). This indicates that there are
1490 non-understood differences between data and MC (for example, different hit
1491 thresholds, noise simulation etc.) and one cannot directly translate the effects
1492 of detector parameters from MC to data and make decisions based on MC.

1493 Another reason that could be causing these contradictory observations in
1494 MC and data is not considering the purity monitor measurement fluctuations.
1495 Figure 55 shows the QA/QC measurement from the purity monitor as a function
1496 of time [6]. The statistical and systematic uncertainty on the measurement
1497 is shown by the red band which indicates that the purity measurement can
1498 fluctuate roughly by 1 to 2 ms. This indicates that although the purity monitor
1499 shows 3 ms lifetime for runs between 5650 and 5750, in principle considering
1500 the measurement fluctuations, the electron lifetime might be higher by about 1
1501 to 2 ms than what was measured. This can explain why the average number of
1502 tracks per event stayed stable during this low purity runs.

1503 To decide whether to include the seemingly low-purity runs in our final
1504 data sample, a set of data-driven studies were performed. We divided the on-
1505 beam and off-beam data into three categories: low purity sample (runs 5650
1506 to 5750), high (normal) purity sample and full (all runs) sample. The CC
1507 inclusive event selection (as detailed in the previous section) passing rates and
1508 kinematic distributions (as will be shown in the next section) from both Selection
1509 I (for original and modified versions) and Selection II are evaluated from the
1510 three sample categories. Selection II also looked at start, stop and vertex (X,
1511 Y, Z) distributions for both on beam and off beam data. For detailed plots
1512 and distributions, refer to doc-db. 5930 (for Selection I) and doc-db. 5943
1513 (for Selection II). As demonstrated by these studies, there were no abnormal
1514 features observed due to inclusion of low purity runs in both selections. This
1515 was deduced by comparing the high purity sample to the full sample. The

1516 shape normalized distributions were found to be basically consistent. However,
1517 since high purity sample is a subset of the full sample, the errors are correlated,
1518 therefore, the absolutely normalized bin center differences reflect the effect of
1519 the longer lifetime. Also, since the low purity sample only forms 6% – 8% of the
1520 full (on-beam and off-beam) sample, statistical errors were at a negligible-level.
1521 Based on the studies briefly discussed above, the low purity runs were included
1522 in the final sample.

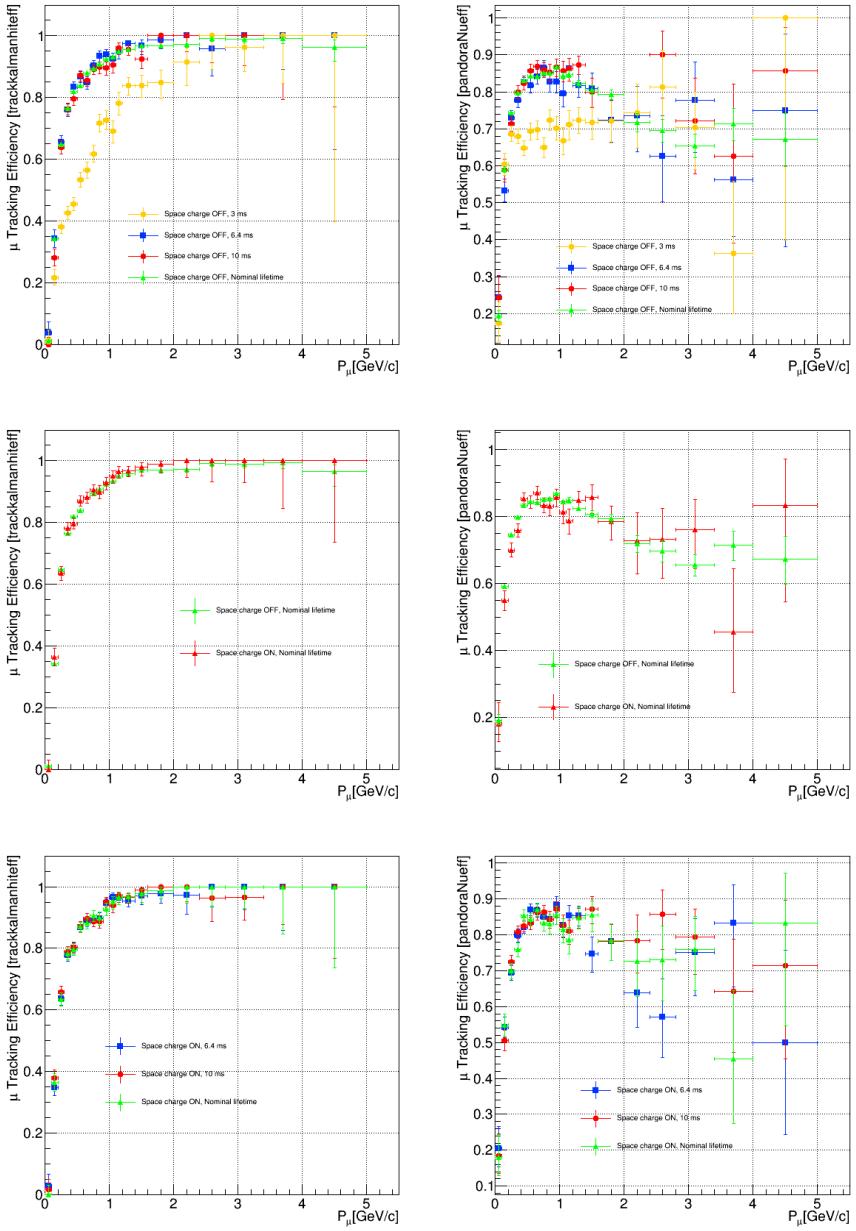


Figure 50: Muon tracking efficiency as a function of reconstructed muon momentum using TrackKalmanhit tracks (left column) and pandoraNu tracks (right column). Plots show the effect of lifetime (top row) and space charge (middle row) separately and combined (bottom row) on muon tracking efficiency.

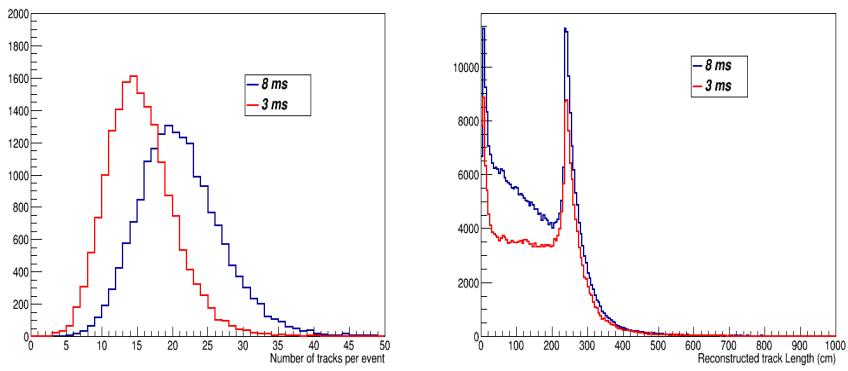


Figure 51: Average number of tracks per event (left) and average track length per event (right) distributions for 3 ms and 8 ms electron lifetime simulation samples. The peak at ~ 250 cm in the track length distribution is caused by the TPC geometry. It is about the height of the TPC and all through-going cosmic tracks end up being reconstructed with this length.

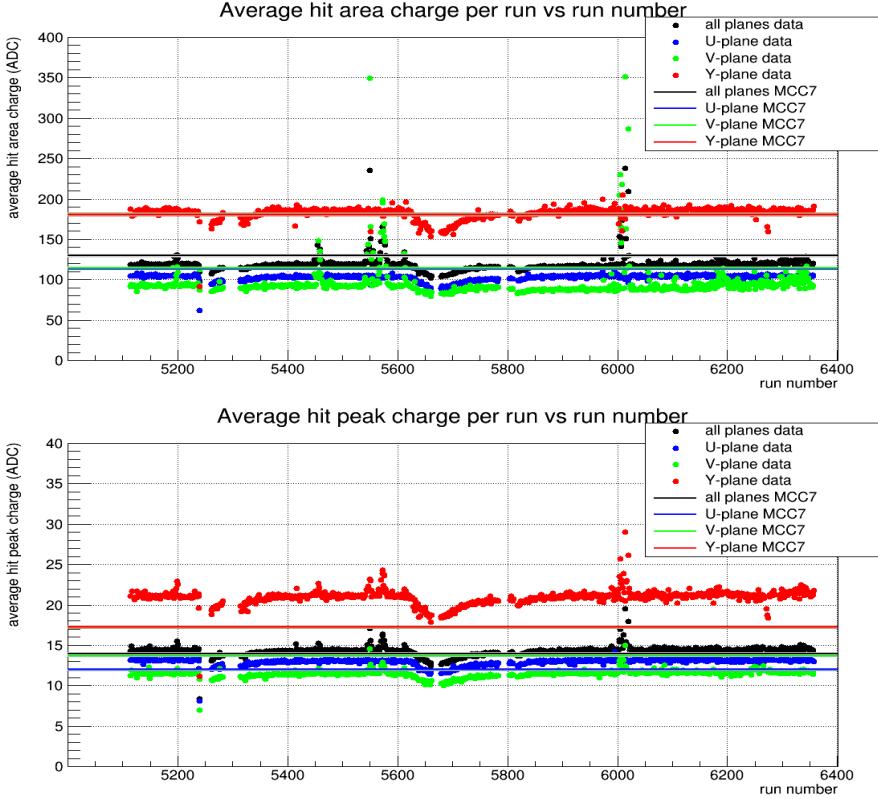


Figure 52: Average hit area charge per run (left) and average hit peak charge per run (right) vs run number distributions. Plots are made using MCC7 CORSIKA in-time cosmic sample and external BNB triggered data over period starting from 02.22.2016 to 05.22.2016 covering run range (5114 – 6356) since the beginning of the official data production with LArSoft production release v05_08_00.

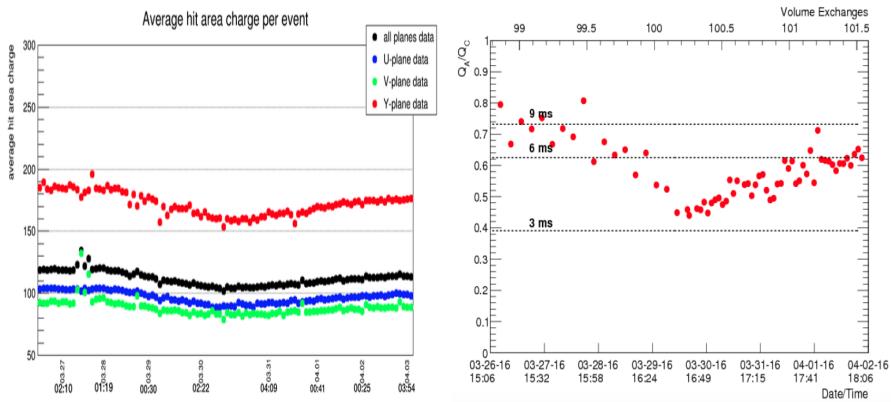


Figure 53: Average hit area charge plot vs date/time (left) and purity monitor plot from the slow monitoring database (right) for the run range 5650 – 5750. These runs correspond to the low purity runs as indicated by the purity monitor.

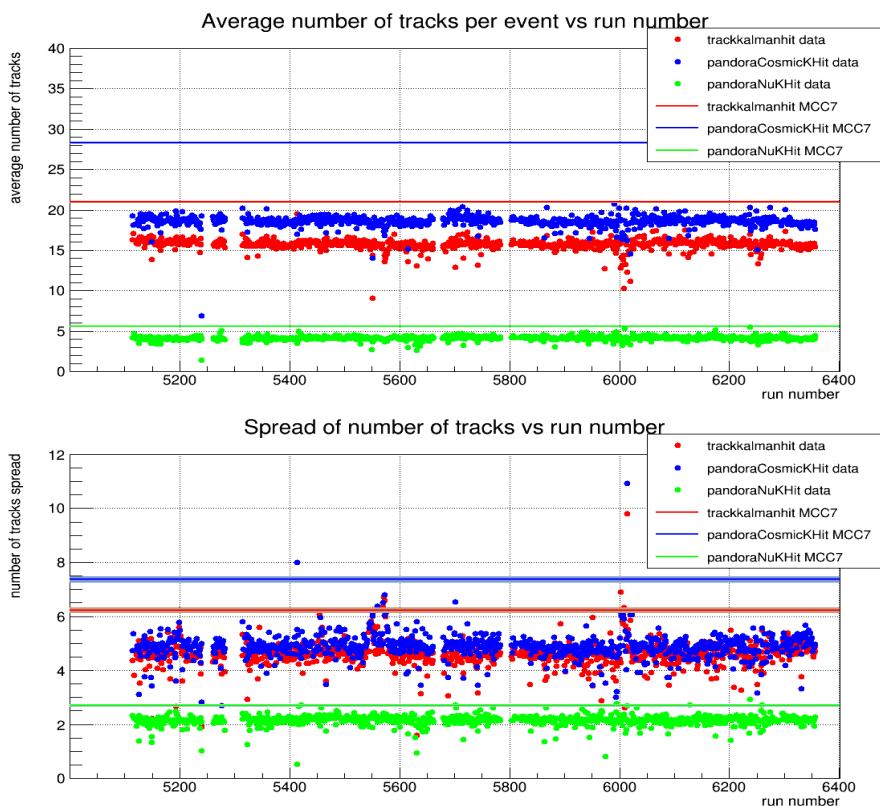


Figure 54: Average number of tracks per run vs run number (left) and spread of number of tracks per run vs run number (right). Plots are made using MCC7 CORSIKA in-time cosmic sample and external BNB triggered data over period starting from 02.22.2016 to 05.22.2016 covering run range (5114 – 6356) since the beginning of the official data production with LArSoft production release v05_08_00.

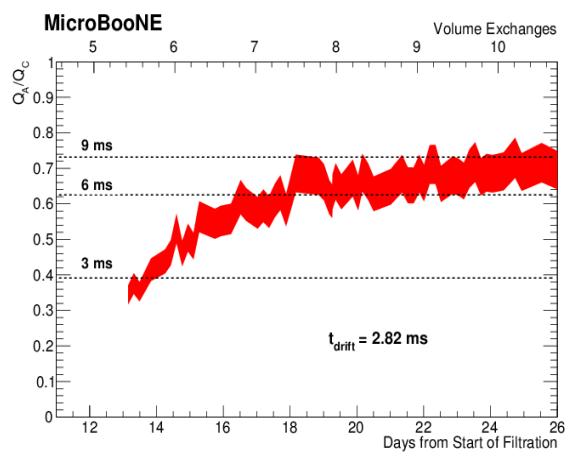


Figure 55: QA/QC measurement using purity monitor data [6]. The red band shows the statistical and systematic uncertainty on the measurement.

1523 8 Final event Distributions

1524 This chapter summarizes the final event distributions for the two event selec-
1525 tions. We are intending to show some of these plots at the NEUTRINO confer-
1526 ence. For both selections, the standard set of plots we are showing is:

- 1527 • On-beam minus off-beam data compared to baseline BNB+Cosmic MC

1528 The subtraction of off-beam data subtracts the background of events of
1529 purely cosmic events. Beam related backgrounds (NC, $\bar{\nu}_\mu$, ν_e) as well as mis-
1530 identified events where a neutrino induced flash in the beam window is present
1531 but a cosmic track is selected as the m.uThese events are andidate are still in-
1532 cluded and are different from the purely cosmic events, which are not related
1533 to any neutrino interactionn thelled cosmics in BNB events). Therefore, this
1534 on-beam minus off-beam data sample can be compared to a BNB+Cosmic MC
1535 sample. A BNB+Cosmic MC sample does contain a neutrino interaction in
1536 every event, and cosmics from CORSIKA are overlaid. [Note: an overlay of
1537 off-beam data with BNB interactions would be preferable, but is not available
1538 at this point.] The error bars in these plots symbolize statistical uncertainties
1539 only.

1540 In lack of a full systematic treatment, we are presenting the effect of system-
1541 atic uncertainties by adding the following set of plots:

- 1542 • On-beam minus off-beam data compared to baseline BNB+Cosmic MC
1543 with beam uncertainty band added
- 1544 • On-beam minus off-beam data compared to BNB+Cosmic MC with M_A^{QE}
1545 scaled
- 1546 • On-beam minus off-beam data compared to BNB+Cosmic MC with meson
1547 exchange currents (MEC) turned on
- 1548 • On-beam minus off-beam data compared to BNB+Cosmic MC with effec-
1549 tive spectral function with transverse enhancement as nuclear model

1550 These systematics are described in more detail in section 7.

1551 Figure 56 illustrate how the angles θ and ϕ , shown in the following plots,
1552 are defined in the MicroBooNE coordinate system.

1553 **All plots in this chapter show area
1554 normalized data to MC comparisons.**

1555 8.1 CC inclusive event distributions for Selection I (orig- 1556 inal)

1557 Using the above described normalization we are now able to subtract the cosmic
1558 only background (off-beam BNB EXT sample) from our on-beam BNB data
1559 sample. This is done for every kinematic distribution separately. As an example
1560 of the normalized on-beam and off-beam distribution is given in Figure 57 and
1561 more can be found in appendix B. The difference of the two curves is then shape
1562 normalized and compared to the BNB + Cosmic distribution, see Figures 58
1563 through 62 for the original selection.

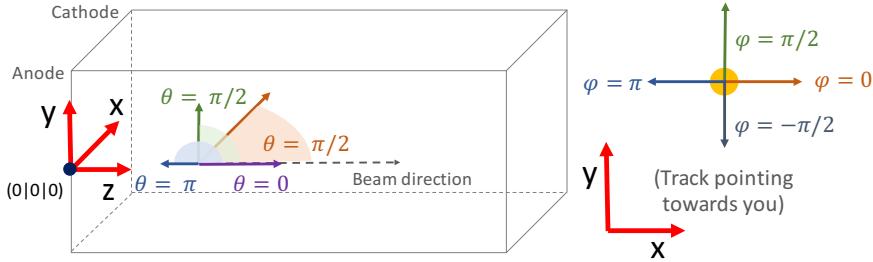


Figure 56: Definitions of angles θ and ϕ in the MicroBooNE coordinate system. The angle θ is the angle of the track wrt the beam axis. ϕ is rotating around the beam axis/track.

1564 8.1.1 Thoughts on the observed excess in θ and ϕ

1565 Figures 59 and 60 show an excess in the on-beam minus off-beam sample wrt
 1566 MC at $1 \leq \theta \leq 1.5$ and $-1.8 \leq \phi \leq -0.8$ in the original event selection scheme.
 1567 This excess is believed to be caused by cosmics, since these directions are exactly
 1568 downward going events. A series of tests have been performed in order to better
 1569 understand or even remove the event excess, which are summarized in the list
 1570 below.

- 1571 1. The 2D θ - ϕ distribution showed the excess in the given θ - ϕ region, see
 1572 Figure 63a. This showed that the excess events in both distributions are
 1573 the same.
- 1574 2. Since it was believed that the excess is caused by cosmic events leaking in,
 1575 the y-distribution of track starts was plotted specifically for events in this
 1576 angular region. However, there was no clear entering point of events. The
 1577 fiducial volume was then reduced in y to a 30 cm border. This should be
 1578 sufficient to exclude space charge effects, which have been observed up to
 1579 a depth of ~ 13 cm. However, the excess remains.
- 1580 3. All vertices and track start/end points in or within ~ 5 cm of the known
 1581 large regions of dead wires have been excluded from the sample. The
 1582 excess remained.
- 1583 4. The flash matching distance cut has been changed from 80cm to 14cm in
 1584 order to prevent against cosmic background being matched to a neutrino
 1585 flash. The excess remained.
- 1586 5. We introduced a cut on the distance between the track range (3D distance
 1587 between start and end point) and the track length (following the track

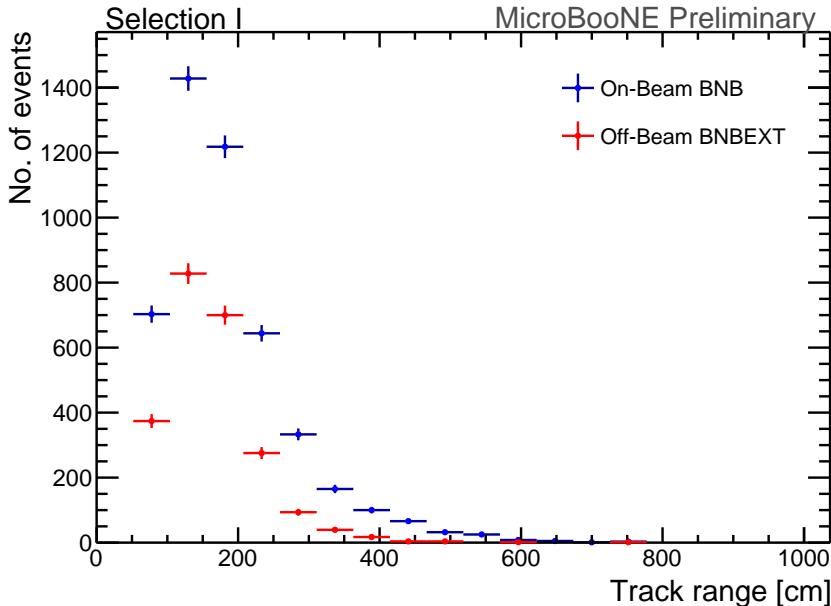


Figure 57: Track range distributions of the normalized on-beam data sample in blue and off-beam data sample in red. By subtracting the off-beam from the on-beam distribution we get our kinematic distributions in Figures 58 through 62.

1588 trajectory) to be ≤ 10 cm. This was done in order to remove tracks that
 1589 might be distorted in the reconstruction through mis-reconstruction or
 1590 space charge effects. The excess remained.

- 1591 6. Added the requirement that there is only exactly one flash ≥ 50 PE in the
 1592 beam window. This is to prevent us from picking the wrong flash to match
 1593 with tracks. The excess remained.
- 1594 7. We ran the event selection over a sample with space charge ON. The excess
 1595 seemed to be slightly less. However, the statistics of this test sample were
 1596 small.

1597 We currently think that the excess is caused by a mis-modeling of our MC.
 1598 Since we do subtract off-beam data, we assume that all entirely cosmic back-
 1599 ground events have been removed from the sample. The events remaining are
 1600 expected to have a good flash caused by a neutrino interaction in or outside the
 1601 FV. We assume that in the following a cosmic track is matched to the flash,
 1602 and passes the selection. Since we know that our cosmic modeling is not per-
 1603 fect in the MC, it could well be that various reasons cause the cosmics in the
 1604 data to more often be contained (through e.g. misreconstruction) than in the
 1605 CORSIKA sample. This could explain the remaining event excess.

1606 The modified selection scheme doesn't show the same magnitude of the above
 1607 discussed excess. The contamination of cosmic background was reduced by
 1608 $\sim 50\%$ in the modified selection (see Table 8). If the assumption is true that

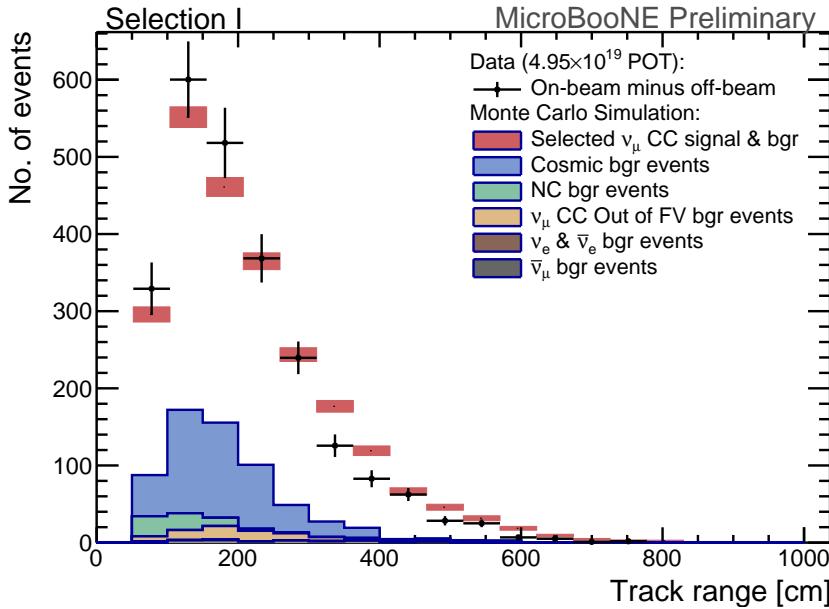


Figure 58: **Selection I: Original** Track range Distribution. The track range is defined as the 3D distance between the start and end point of the muon candidate track. The black data points symbolize on-beam minus off-beam data with statistical error bars. The red shaded histogram represents the BNB+Cosmic MC, with the bands representing the statistical uncertainty on the MC. The backgrounds contained in the red are additionally shown in colors. Backgrounds are stacked.

1609 the excess events are due to cosmics, then this behavior is exactly what we
 1610 expected.

1611 8.1.2 Cut on Vertical Quadrants of ϕ

1612 We also performed a cut on the vertical quadrants of the ϕ -angle, i.e. $-\pi <$
 1613 $\phi < -3/4\pi$, $-\pi/4 < \phi < \pi/4$, and $3/4\pi < \phi < \pi$. The changes in shape are
 1614 marginal but a notable reduction of cosmic background can be seen (figure 64).

1615 Further studies that we have planned are:

- 1616 • Modify the cuts and move the flash matching up in the chain. This is ex-
 1617 pected to reduce the background of cosmics in BNB events, and therefore
 1618 might reduce the excess.
- 1619 • Several ideas to split the sample in two (or more) parts, e.g. first half and
 1620 second half of runs, upstream and downstream side of the detector,... in
 1621 order to see if the excess remains in both subsamples.
- 1622 • In the longterm: produce a sample with BNB simulation and off-beam
 1623 overlay. However, this won't be ready for NEUTRINO2016.

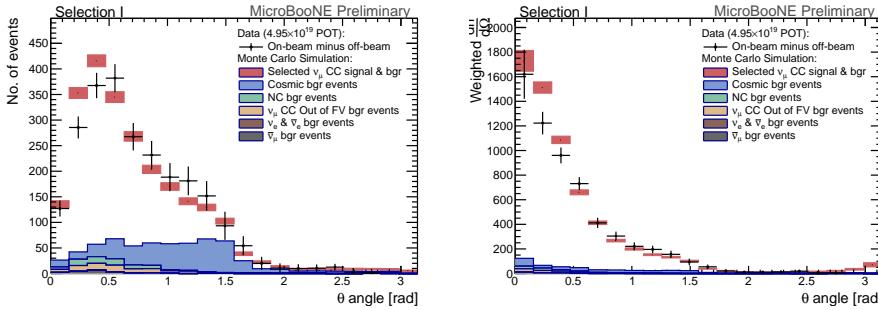


Figure 59: **Selection I: Original θ angle of track distribution.** In the right-hand plot, the distribution has been normalized per solid angle. The angle θ is defined as the angle in beam direction (with $\theta = 0$ beam direction). The black data points symbolize on-beam minus off-beam data with statistical error bars. The red shaded histogram represents the BNB+Cosmic MC, with the bands representing the statistical uncertainty on the MC. The backgrounds contained in the red are additionally shown in colors. Backgrounds are stacked.

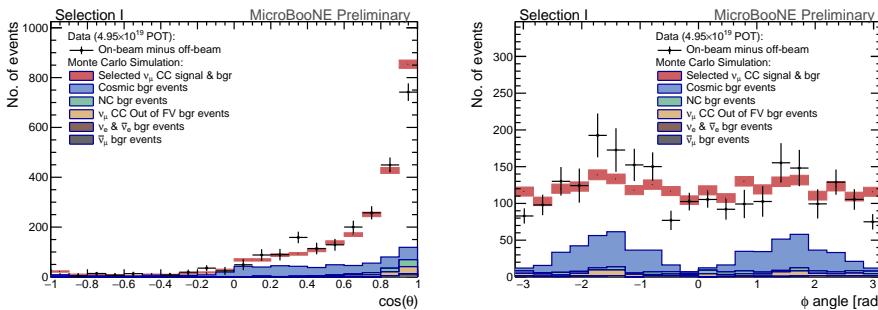


Figure 60: **Selection I: Original $\cos \theta$ and ϕ angle of track distribution.** The angle θ is defined as the angle in beam direction (with $\theta = 0$ beam direction). ϕ is the angle around beam direction with $-\pi/2$ being vertically downward going and $\pi/2$ being vertically upward going tracks. The black data points symbolize on-beam minus off-beam data with statistical error bars. The red shaded histogram represents the BNB+Cosmic MC, with the bands representing the statistical uncertainty on the MC. The backgrounds contained in the red are additionally shown in colors. Backgrounds are stacked.

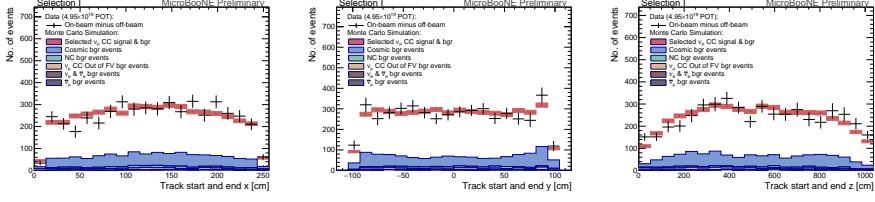


Figure 61: **Selection I: Original** Track start and end point distribution in the x, y, and z-coordinate. The black data points symbolize on-beam minus off-beam data with statistical error bars. The red shaded histogram represents the BNB+Cosmic MC, with the bands representing the statistical uncertainty on the MC. The backgrounds contained in the red are additionally shown in colors. Backgrounds are stacked.

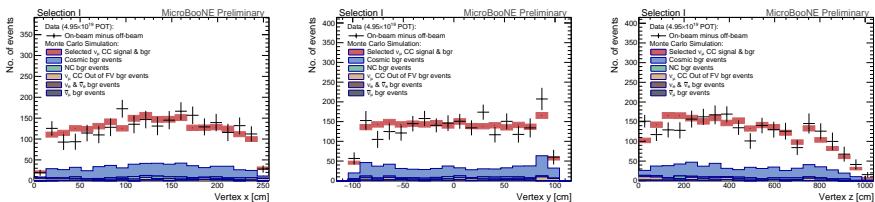


Figure 62: **Selection I: Original** Neutrino interaction candidate vertex distribution in the x, y, and z-coordinate. The black data points symbolize on-beam minus off-beam data with statistical error bars. The red shaded histogram represents the BNB+Cosmic MC, with the bands representing the statistical uncertainty on the MC. The backgrounds contained in the red are additionally shown in colors. Backgrounds are stacked.

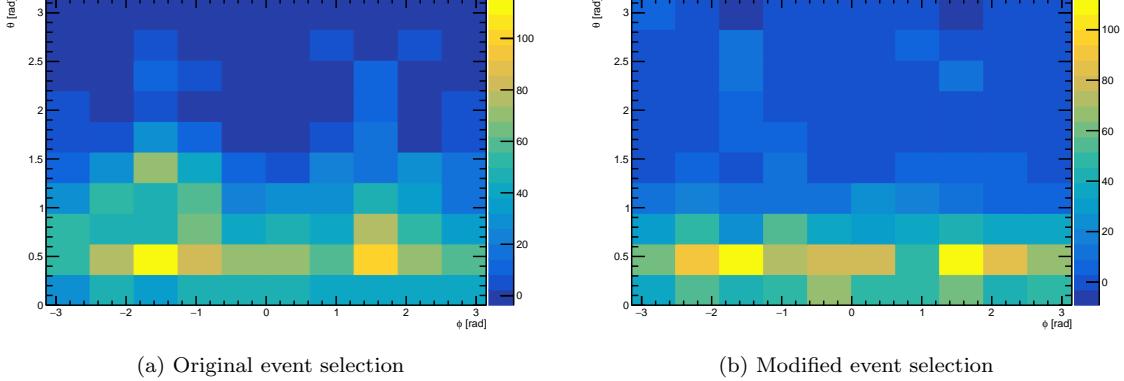


Figure 63: θ - ϕ -distribution of both selection schemes. There is a notable difference in the intervals $1 \leq \theta \leq 1.5$ and $-1.8 \leq \phi \leq -0.8$ between the two selection schemes. While the original event selection shows a clear correlated excess in the given intervals, this feature seems to be cut out effectively. There are still excesses around $\theta \approx 0.5$ and $\phi \approx \pm\pi/2$.

- 1624 • Another in the longterm: we hope that a proper flash matching algorithm
- 1625 will achieve much better removal of cosmics in BNB events. The flash
- 1626 matching module is being worked on, but while first tests on perfect reco
- 1627 seemed promising (see DocDB 5469), the current performance on actual
- 1628 reconstruction suggests that it is not yet ready to be used in analyses (See
- 1629 DocDB 5713, 5777).

1630 **8.2 CC inclusive event distributions for Selection I (modified)**

1632 The same procedure is used to generate the kinematic distributions of the modified selection. They are shown in the figures 65 through 69.

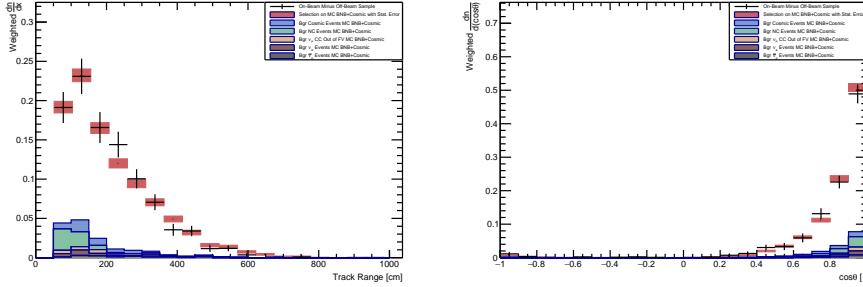


Figure 64: Selection I: Modified Shown on the left is the track range and on the right the $\cos \theta$ distribution after the cut on the ϕ -angle: $-\pi < \phi < -3/4\pi$, $-\pi/4 < \phi < \pi/4$, and $3/4\pi < \phi < \pi$. The black data points symbolize on-beam minus off-beam data with statistical error bars. The red shaded histogram represents the BNB+Cosmic MC, with the bands representing the statistical uncertainty on the MC. The backgrounds contained in the red are additionally shown in colors. Backgrounds are stacked.

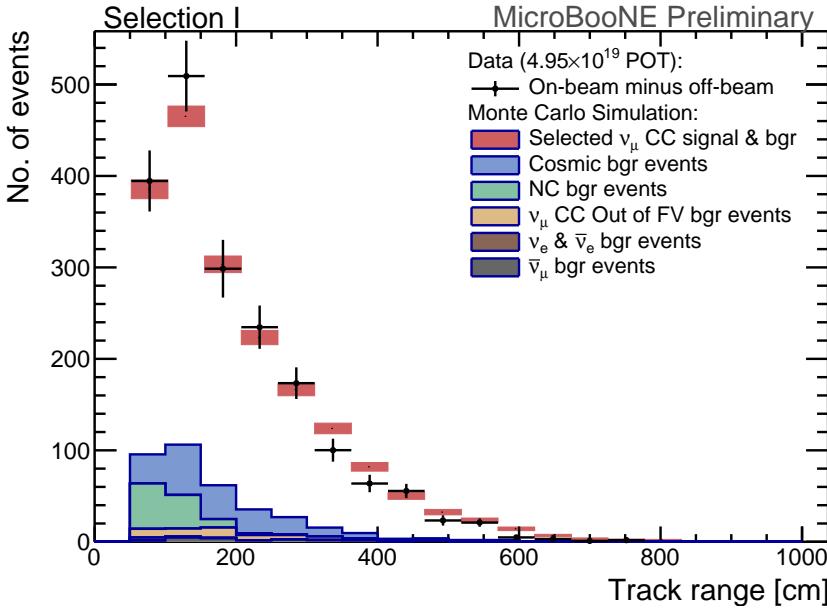


Figure 65: Selection I: Modified Track range Distribution. The track range is defined as the 3D distance between the start and end point of the muon candidate track. The black data points symbolize on-beam minus off-beam data with statistical error bars. The red shaded histogram represents the BNB+Cosmic MC, with the bands representing the statistical uncertainty on the MC. The backgrounds contained in the red are additionally shown in colors. Backgrounds are stacked.

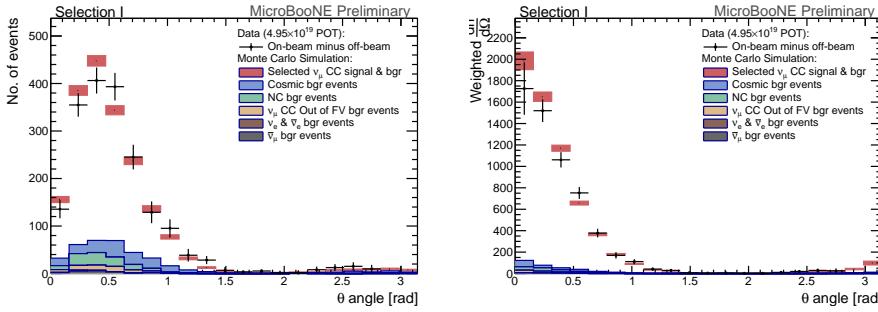


Figure 66: **Selection I: Modified θ** angle of track distribution. In the right-hand plot, the distribution has been normalized per solid angle. The angle θ is defined as the angle in beam direction (with $\theta = 0$ beam direction). The black data points symbolize on-beam minus off-beam data with statistical error bars. The red shaded histogram represents the BNB+Cosmic MC, with the bands representing the statistical uncertainty on the MC. The backgrounds contained in the red are additionally shown in colors. Backgrounds are stacked.

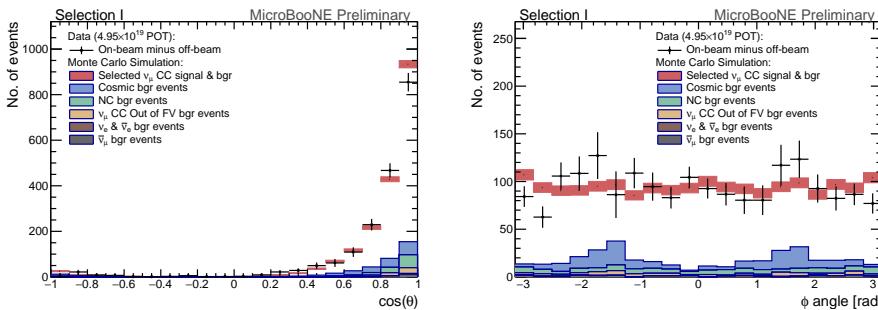


Figure 67: **Selection I: Modified $\cos\theta$ and ϕ** angle of track distribution. The angle θ is defined as the angle in beam direction (with $\theta = 0$ beam direction). ϕ is the angle around beam direction with $-\pi/2$ being vertically downward going and $\pi/2$ being vertically upward going tracks. The black data points symbolize on-beam minus off-beam data with statistical error bars. The red shaded histogram represents the BNB+Cosmic MC, with the bands representing the statistical uncertainty on the MC. The backgrounds contained in the red are additionally shown in colors. Backgrounds are stacked.

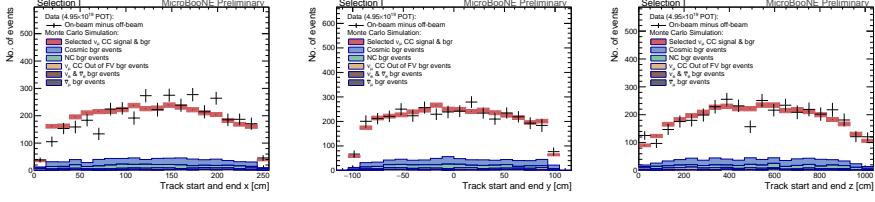


Figure 68: **Selection I: Modified** Track start and end point distribution in the x, y, and z-coordinate. The black data points symbolize on-beam minus off-beam data with statistical error bars. The red shaded histogram represents the BNB+Cosmic MC, with the bands representing the statistical uncertainty on the MC. The backgrounds contained in the red are additionally shown in colors. Backgrounds are stacked.

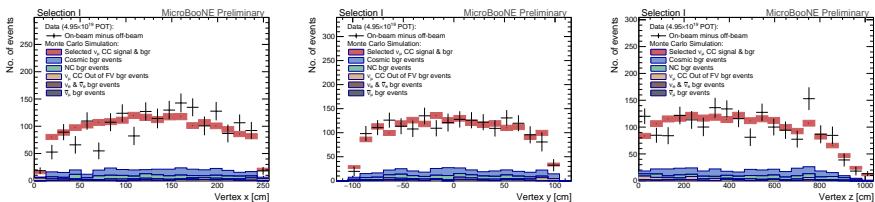


Figure 69: **Selection I: Modified** Neutrino interaction candidate vertex distribution in the x, y, and z-coordinate. The black data points symbolize on-beam minus off-beam data with statistical error bars. The red shaded histogram represents the BNB+Cosmic MC, with the bands representing the statistical uncertainty on the MC. The backgrounds contained in the red are additionally shown in colors. Backgrounds are stacked.

1634 8.3 BNB Systematic Uncertainties for Selection I

1635 These are the same distributions as above with an additional beam systematic
 1636 uncertainty illustrated as an additional band on the BNB+Cosmics histogram
 1637 points. Errors are shown for both selections.

1638 There is a caveat on the errors as they are currently presented: The error
 1639 bars included total flux errors, which isn't a really fair comparison when showing
 1640 area normalized data to MC comparisons. The normalization part would have
 1641 be separated from the shape part of the beam uncertainties. We are currently
 1642 working on this.

1643 An additional caveat is that in the following plots statistical and beam sys-
 1644 tematic error are linearly added. This will get fixed asap.

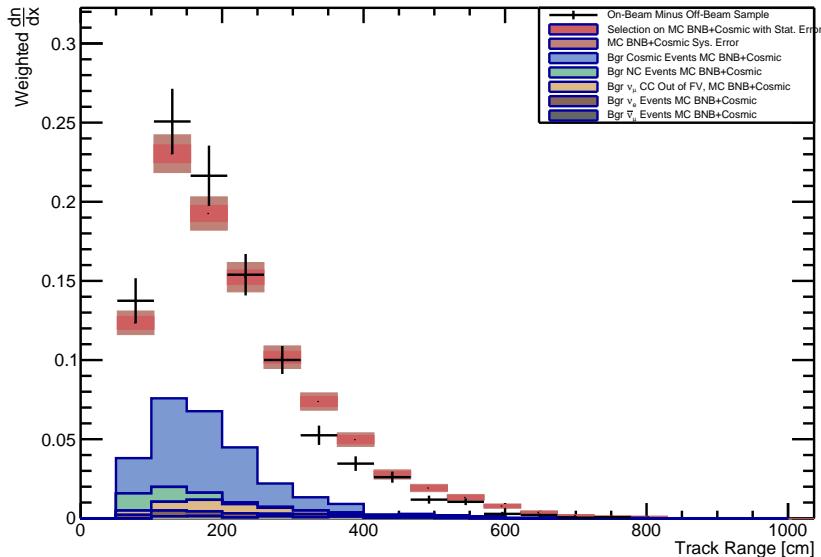


Figure 70: **Selection I: Original** Track range Distribution. The track range is defined as the 3D distance between the start and end point of the muon candidate track. The black data points symbolize on-beam minus off-beam data with statistical error bars. The red shaded histogram represents the BNB+Cosmic MC, with two error bands representing the statistical and the beam systematic uncertainties on the MC. The backgrounds contained in the red are additionally shown in colors. Backgrounds are stacked.

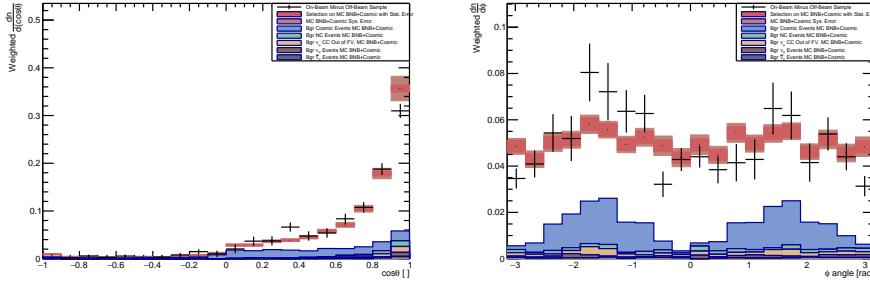


Figure 71: Selection I: Original $\cos \theta$ and ϕ angle of track distribution. The angle θ is defined as the angle in beam direction (with $\theta = 0$ beam direction). ϕ is the angle around beam direction with $-\pi/2$ being vertically downward going and $\pi/2$ being vertically upward going tracks. The black data points symbolize on-beam minus off-beam data with statistical error bars. The red shaded histogram represents the BNB+Cosmic MC, with two error bands representing the statistical and the beam systematic uncertainties on the MC. The backgrounds contained in the red are additionally shown in colors. Backgrounds are stacked.

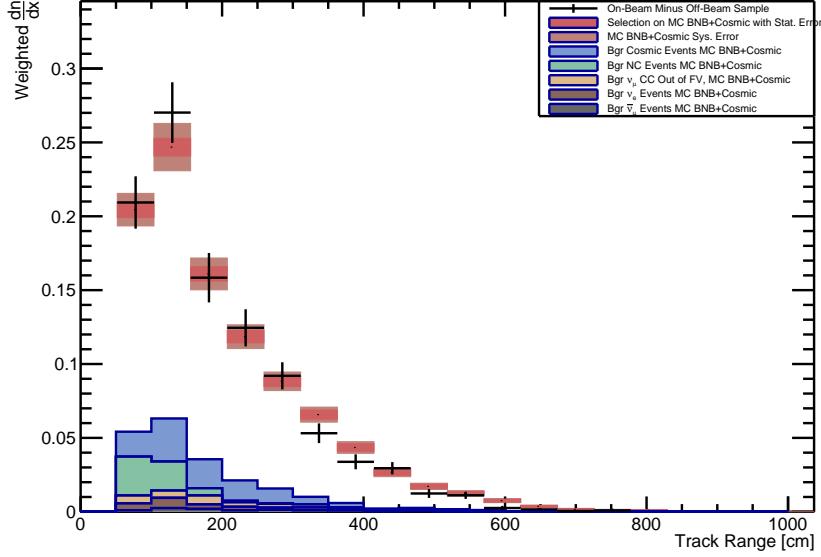


Figure 72: Selection I: Modified Track range Distribution. The track range is defined as the 3D distance between the start and end point of the muon candidate track. The black data points symbolize on-beam minus off-beam data with statistical error bars. The red shaded histogram represents the BNB+Cosmic MC, with two error bands representing the statistical and the beam systematic uncertainties on the MC. The backgrounds contained in the red are additionally shown in colors. Backgrounds are stacked.

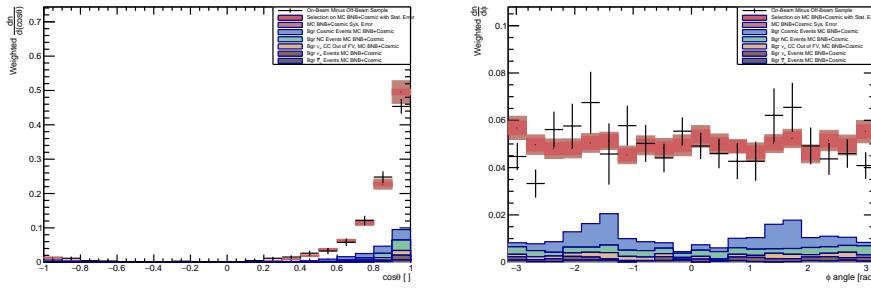


Figure 73: **Selection I: Modified** $\cos\theta$ and ϕ angle of track distribution. The angle θ is defined as the angle in beam direction (with $\theta = 0$ beam direction). ϕ is the angle around beam direction with $-\pi/2$ being vertically downward going and $\pi/2$ being vertically upward going tracks. The black data points symbolize on-beam minus off-beam data with statistical error bars. The red shaded histogram represents the BNB+Cosmic MC, with two error bands representing the statistical and the beam systematic uncertainties on the MC. The backgrounds contained in the red are additionally shown in colors. Backgrounds are stacked.

1645 **8.4 CC inclusive event distributions for Selection II**

1646 This section contains a sets of the plots showing MC and data comparison of
1647 muon candidate distributions from selection II. Selection IIA and IIB have been
1648 added together. Data points of each distribution are from on-beam minus off-
1649 beam data samples. The MC is stacked histogram from BNB+Cosmic sample.
1650 Update: The plots below include the final 15 cm minimum track length cut that
1651 further remove the noise background. Also similar as the public note, the noise
1652 and cosmic background merger as cosmic background, ν_e and $\bar{\nu}_e$ also merged
1653 into one background channel.

- 1654 • **Track length:** Figure 74 shows the track length distribution of the
1655 selected longest track (if multiplicity ≥ 2)
- 1656 • **Track angle $\cos \theta$:** Figure 75 shows the cosine of the beam angle. The
1657 plot shows good agreement between MC and data.
- 1658 • **Track angle ϕ :** Figure 76 shows the ϕ distribution. The selected events
1659 shows the opposite trend as expected by cosmic rays (mostly vertical).
1660 This shows small cosmic contamination in the final sample.
- 1661 • **Reconstructed vertex position in X, Y and Z:** Figure 77 shows
1662 less vertices at the boundary of the fiducial volume in X than MC. A
1663 reconstructed vertex deficiency can be observed in Y direction at around
1664 30cm to 60cm. And obvious drop at 700 cm in Z corresponds to the dead
1665 wire region.

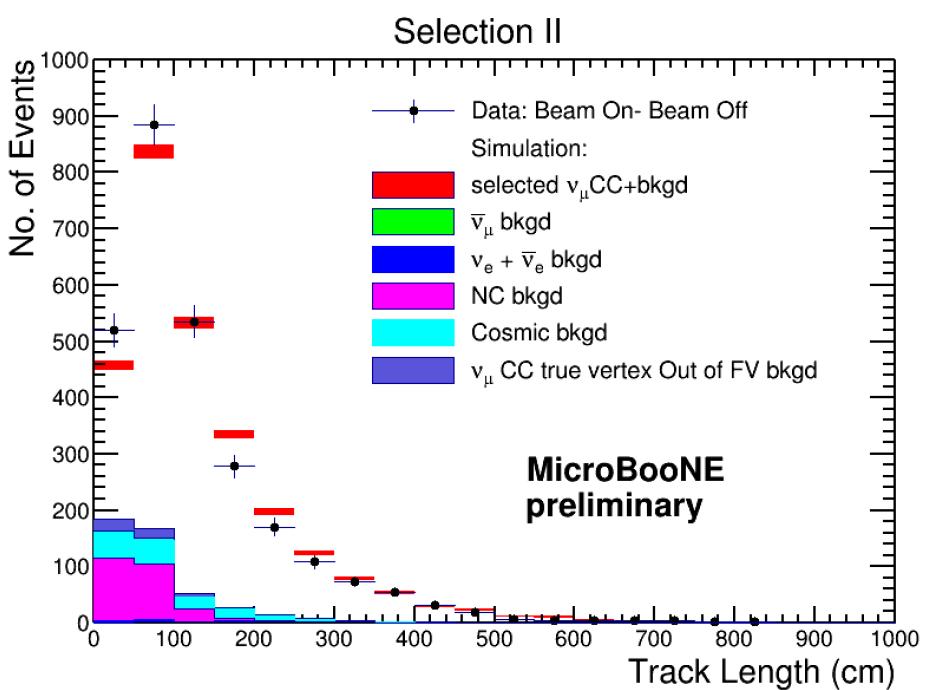


Figure 74: Data (on-beam minus off-beam) vs BNB+Cosmic MC. The red histogram shows the selected events from the BNB+Cosmic MC. Background contributions are shown in the bottom in color. The histograms are stacked. MC and data are area normalized wrt each other.

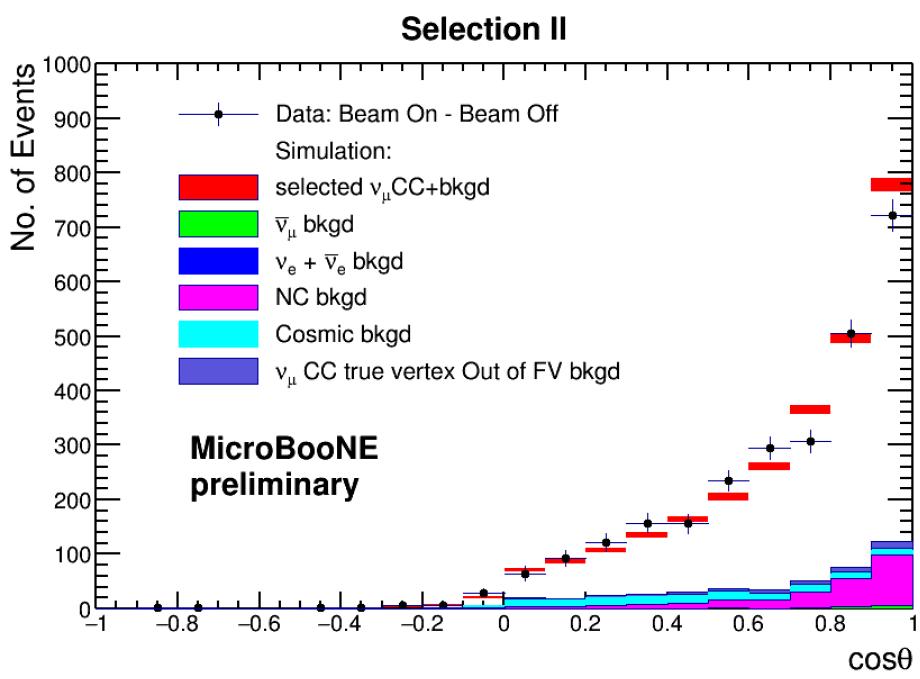


Figure 75: Cos theta distribution of Data (on-beam minus off-beam) vs BNB+Cosmic MC. The red histogram shows the selected events from the BNB+Cosmic MC. Background contributions are shown in the bottom in color. The histograms are stacked. MC and data are area normalized wrt each other.

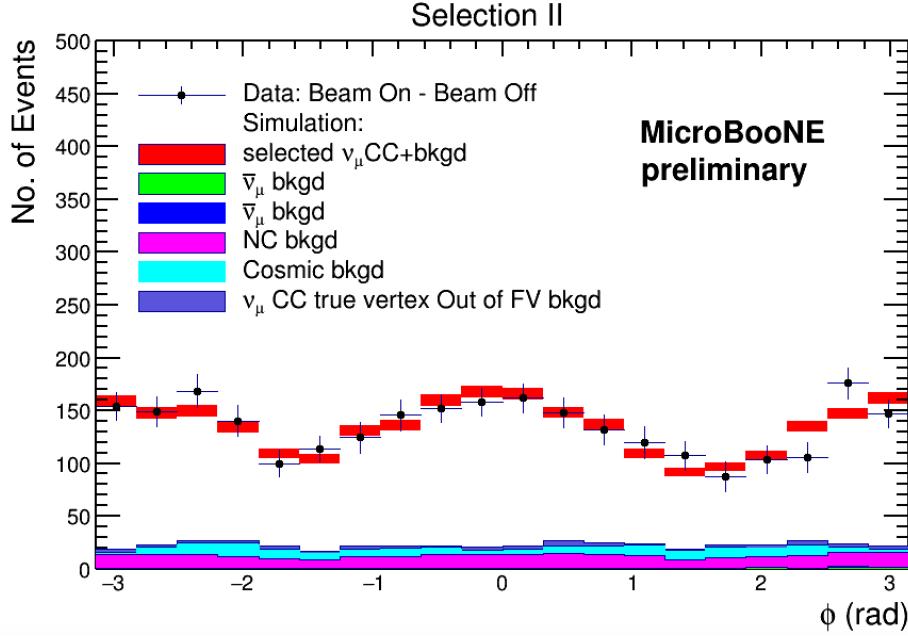


Figure 76: ϕ distribution of Data (on-beam minus off-beam) vs BNB+Cosmic MC. The red histogram shows the selected events from the BNB+Cosmic MC. Background contributions are shown in the bottom in color. The histograms are stacked. MC and data are area normalized wrt each other.

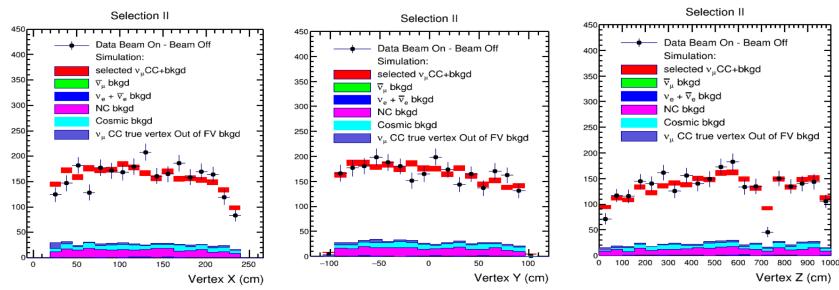


Figure 77: Vertex X,Y,Z (from left to right) distributions of data (on-beam minus off-beam) vs BNB+Cosmic MC. The red histogram shows the selected events from the BNB+Cosmic MC. Background contributions are shown in the bottom in color. The histograms are stacked. MC and data are area normalized wrt each other.

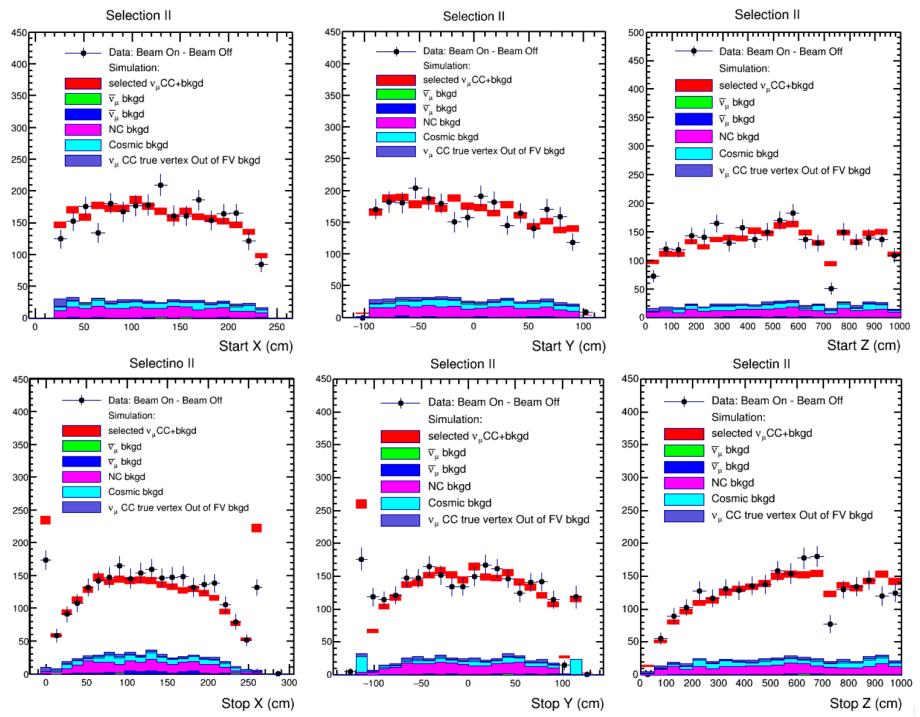
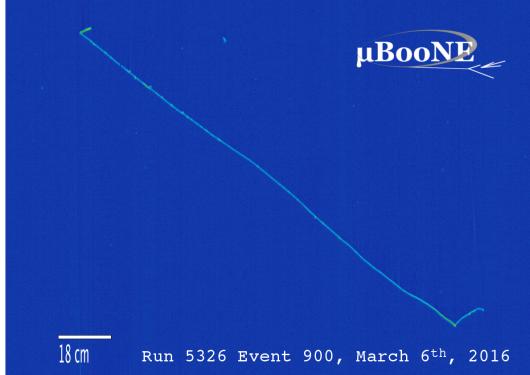


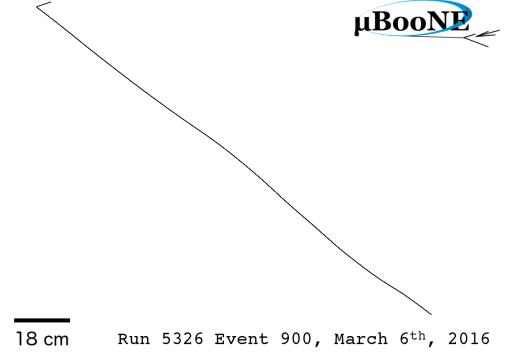
Figure 78: Track start and end positions in X, Y and Z (from left to right) distributions of data (on-beam minus off-beam) vs BNB+Cosmic MC. The red histogram shows the selected events from the BNB+Cosmic MC. Background contributions are shown in the bottom in color. The histograms are stacked. MC and data are area normalized wrt each other.

1666 **8.5 Event views**

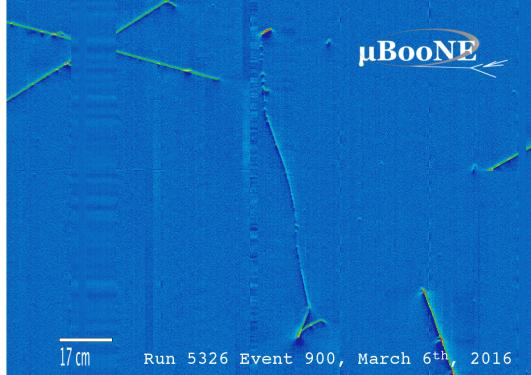
1667 This section shows some event view from selected ν_μ CC candidate events from
1668 the on-beam stream. Figures 79 to 82 show events from selection I with vari-
1669 ous multiplicity, and Figures 83 to ?? show event views from selection II. The
1670 candidate interaction is shown in all three event views, and additional in a 3D
1671 event display showing the reconstructed tracks from pandoraNu (for selection
1672 I) and pandoraNuPMA (for selection II). These events were cherry-picked to
1673 include examples for various track multiplicities.



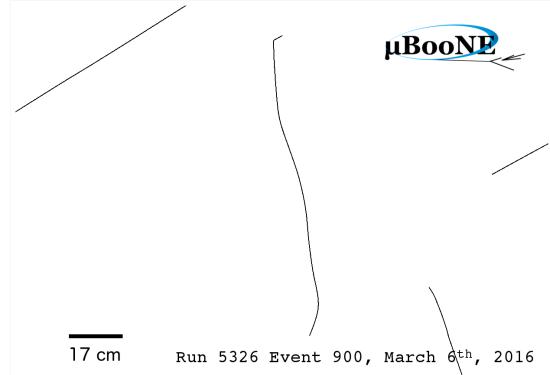
(a) Collection plane (Y)



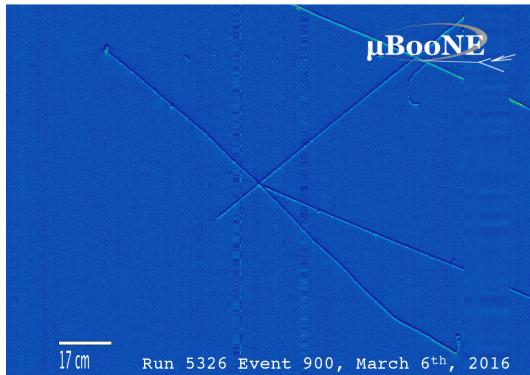
(b) Reconstructed 3D image (Y plane projection)



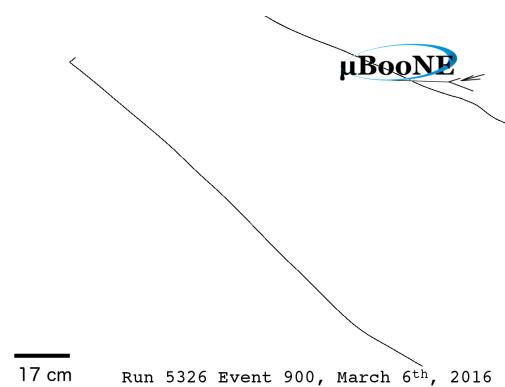
(c) Induction plane (U)



(d) Reconstructed 3D image (U plane projection)



(e) Induction plane (V)



(f) Reconstructed 3D image (V plane projection)

Figure 79: Event view for run 5326, event 900, selected by selection I. The plots on the left show the event view in all three wire planes. The induction planes show features of gaps of unresponsive wires and noise (see e.g. left hand side of 79e). The plots on the right show the three-dimensional reconstructed image projected onto each wire plane. The track reconstruction algorithm shown is pandoraNu. All figures have the same scale (indicated by the bar in the bottom left) and aspect ratio.

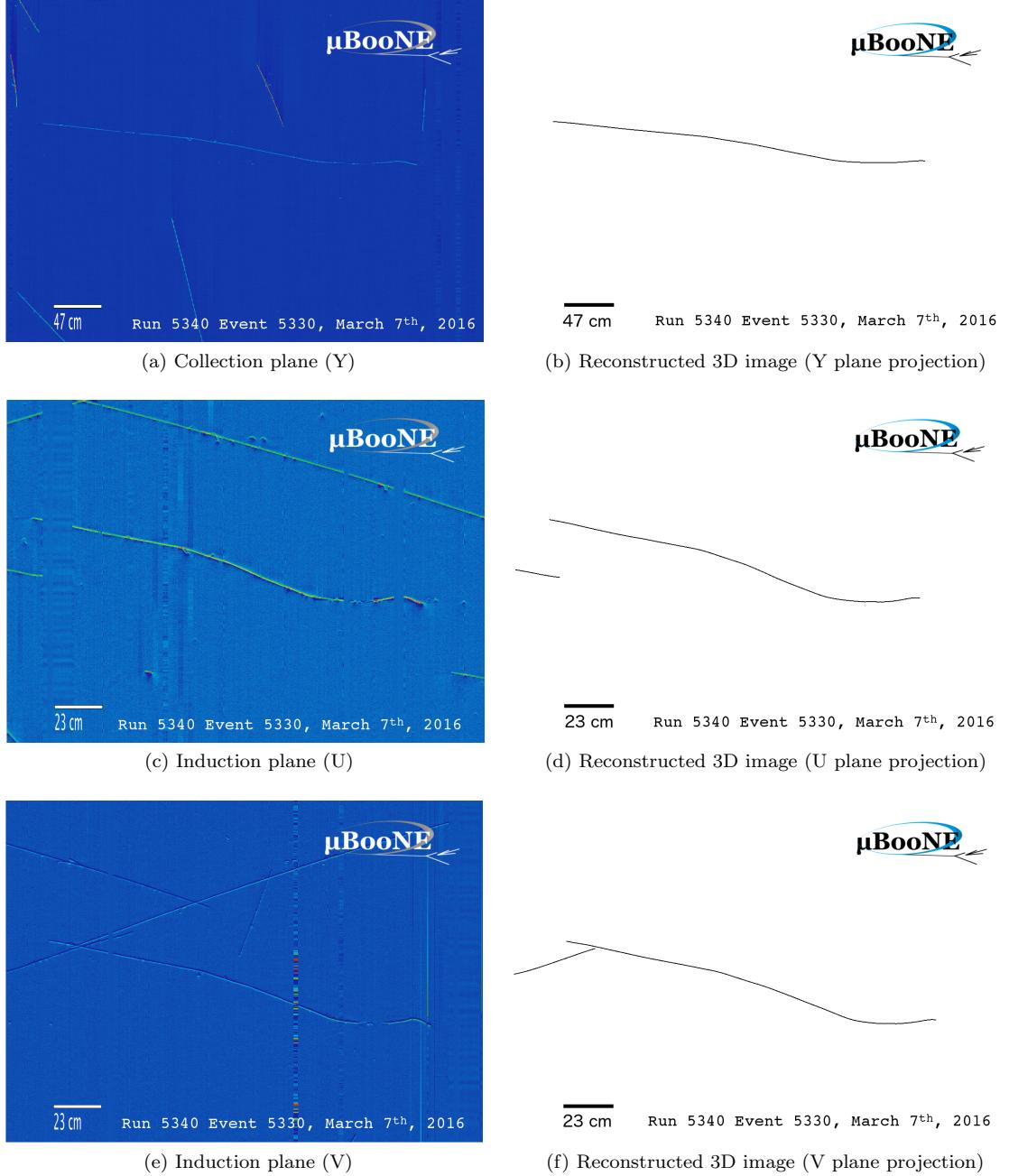


Figure 80: Event view for run 5340, event 5330, selected by selection I. The plots on the left show the event view in all three wire planes. Both induction planes show noise (vertical stripes in Figures 80c and 80e). The plots on the right show the three-dimensional reconstructed image projected onto each wire plane. The track reconstruction algorithm shown is pandoraNu. Note that some tracks visible on the left hand plots are missing on the right because they have been rejected in the cosmic removal pass. All figures have the same scale (indicated by the bar in the bottom left) and aspect ratio.

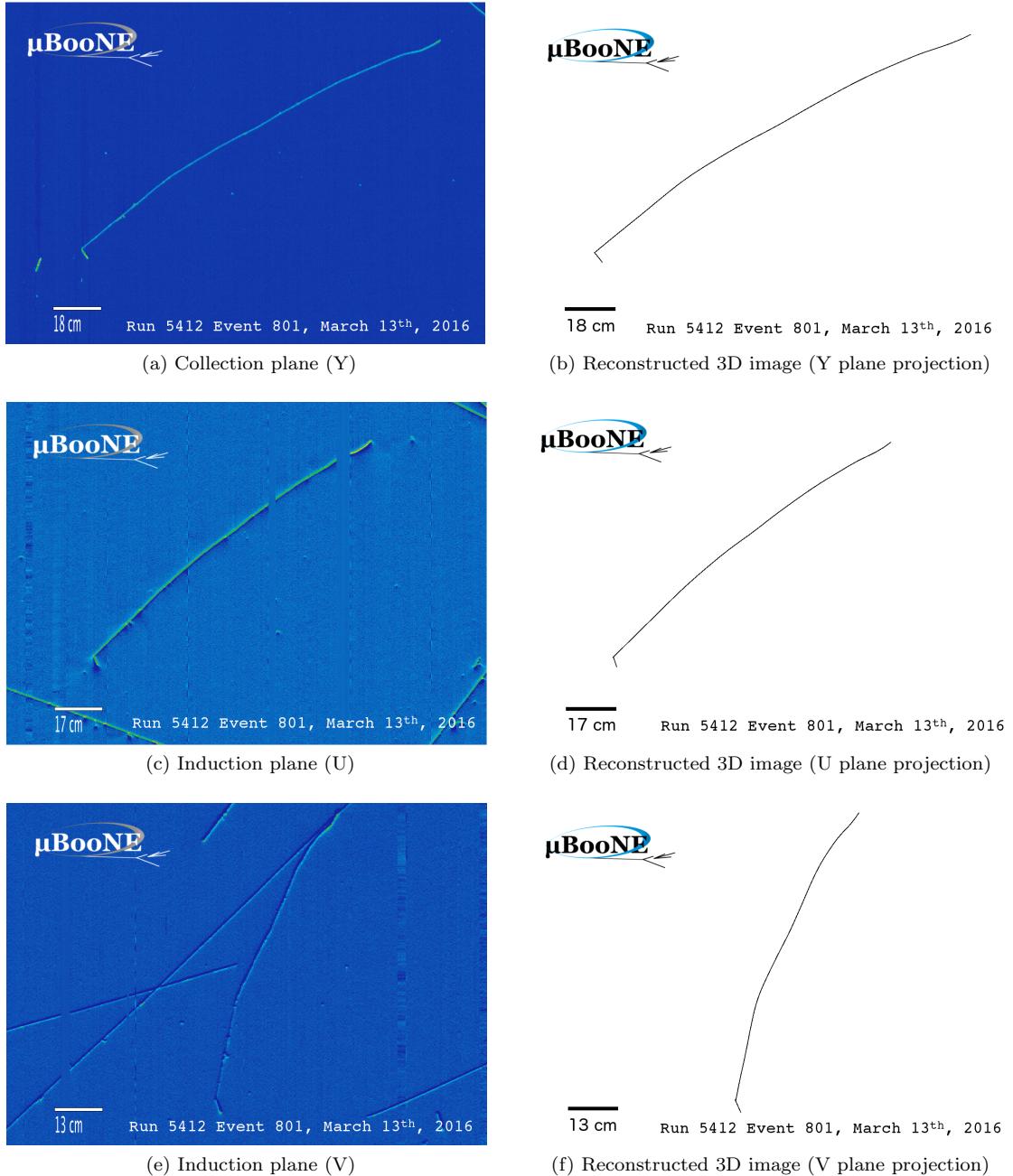


Figure 81: Event view for run 5412, event 801, selected by selection I. The plots on the left show the event view in all three wire planes. The plots on the right show the three-dimensional reconstructed image projected onto each wire plane. The track reconstruction algorithm shown is pandoraNu. Note that some tracks visible on the left hand plots are missing on the right because they have been rejected in the cosmic removal pass. All figures have the same scale (indicated by the bar in the bottom left) and aspect ratio.

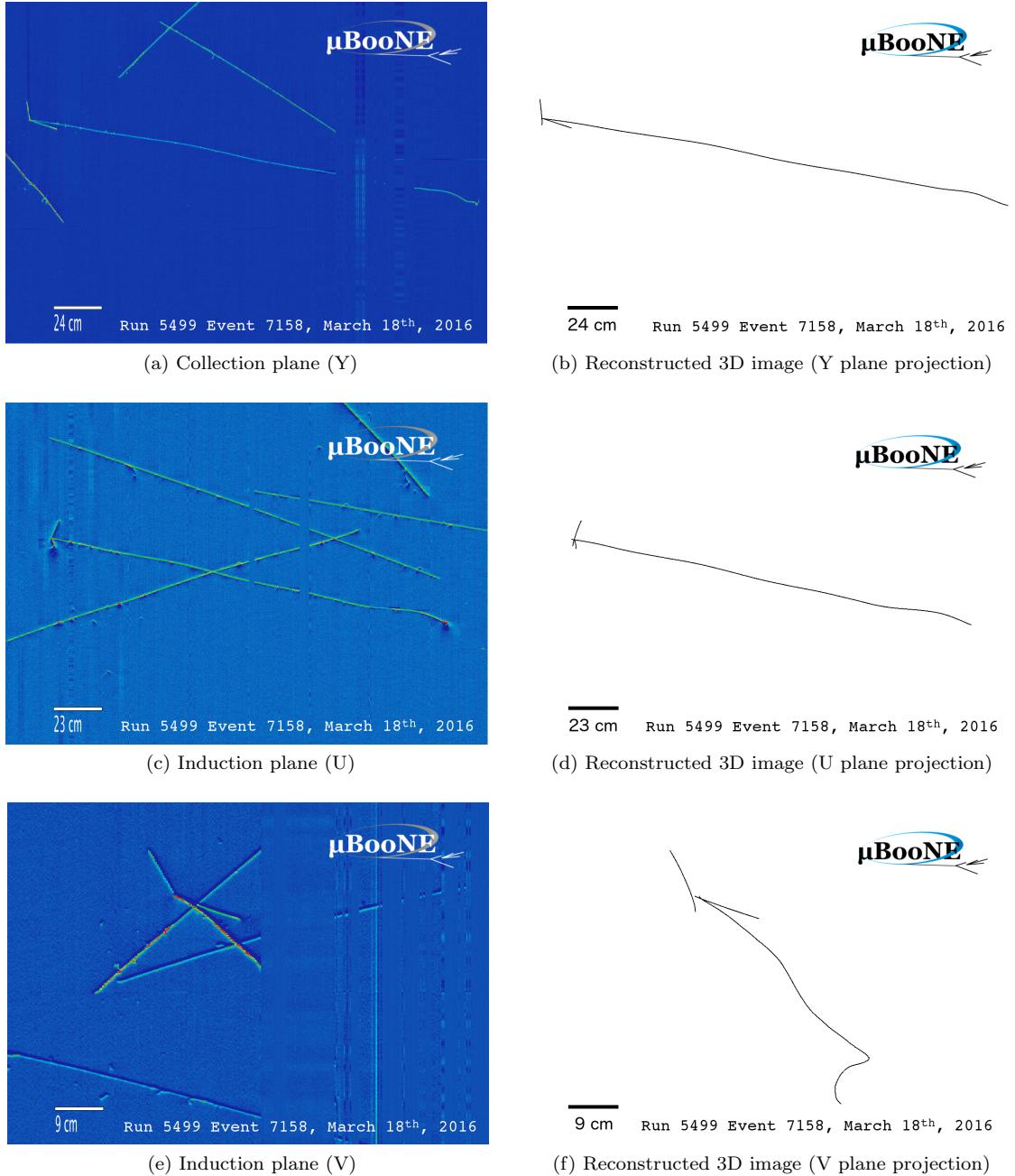


Figure 82: Event view for run 5499, event 7158, selected by selection I. The plots on the left show the event view in all three wire planes. The Y and V plane show gaps of unresponsive wires and noise (see Figures 82a and 82e). The reconstruction successfully bridges these gaps. The plots on the right show the three-dimensional reconstructed image projected onto each wire plane. The track reconstruction algorithm shown is pandoraNu. Note that some tracks visible on the left hand plots are missing on the right because they have been rejected in the cosmic removal pass. All figures have the same scale (indicated by the white bar in the bottom left) and aspect ratio.

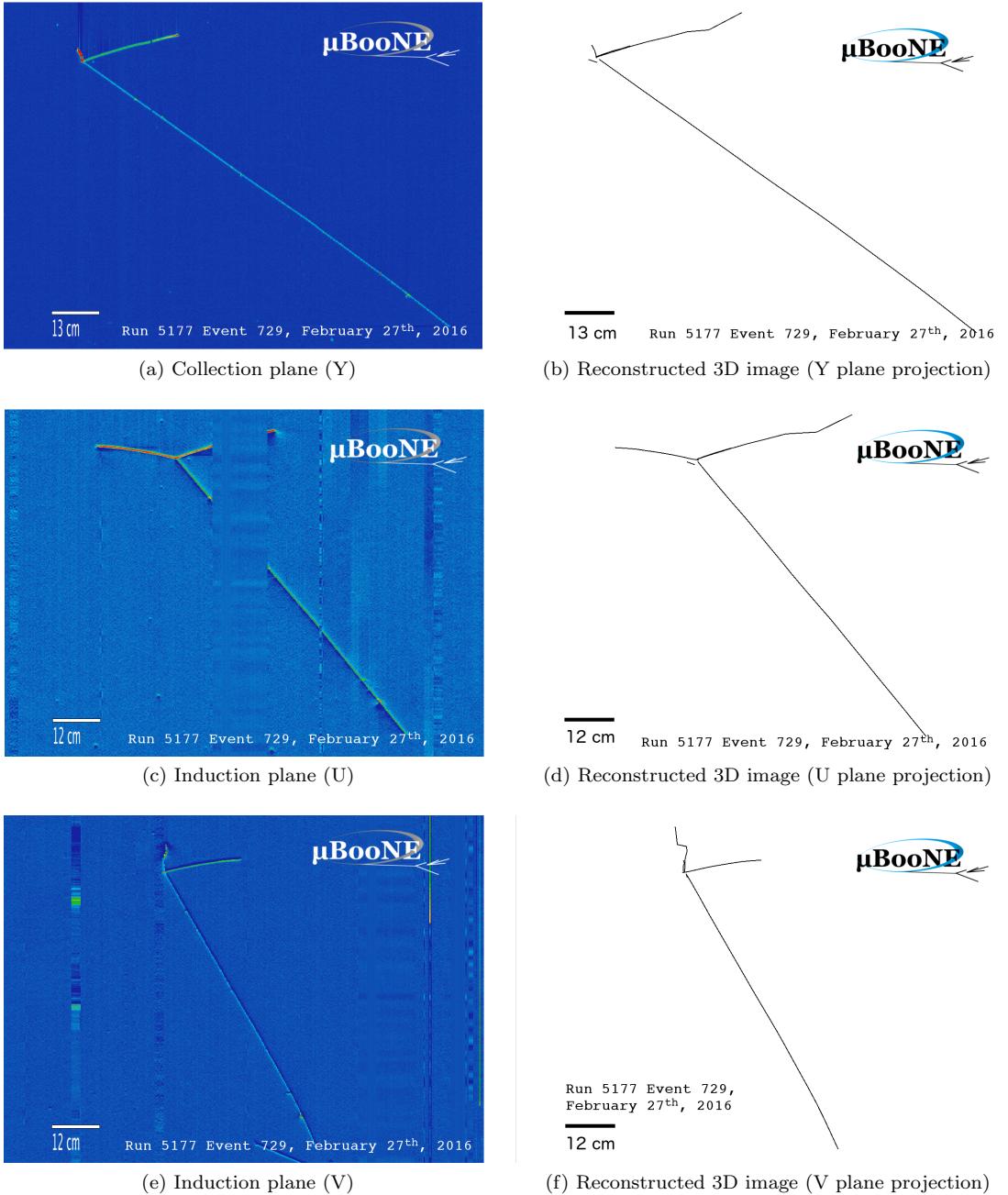


Figure 83: Event view for run 5177, event 729, selected by selection II. The plots on the left show the event view in all three wire planes. A large gap of unresponsive wires is present in the U plane (see 83c). The plots on the right show the three-dimensional reconstructed image projected onto each wire plane. Comparing the reconstruction successfully bridges gaps of unresponsive wires. The track reconstruction algorithm shown is pandoraNuPMA. All figures have the same scale (indicated by the white bar in the bottom left) and aspect ratio.

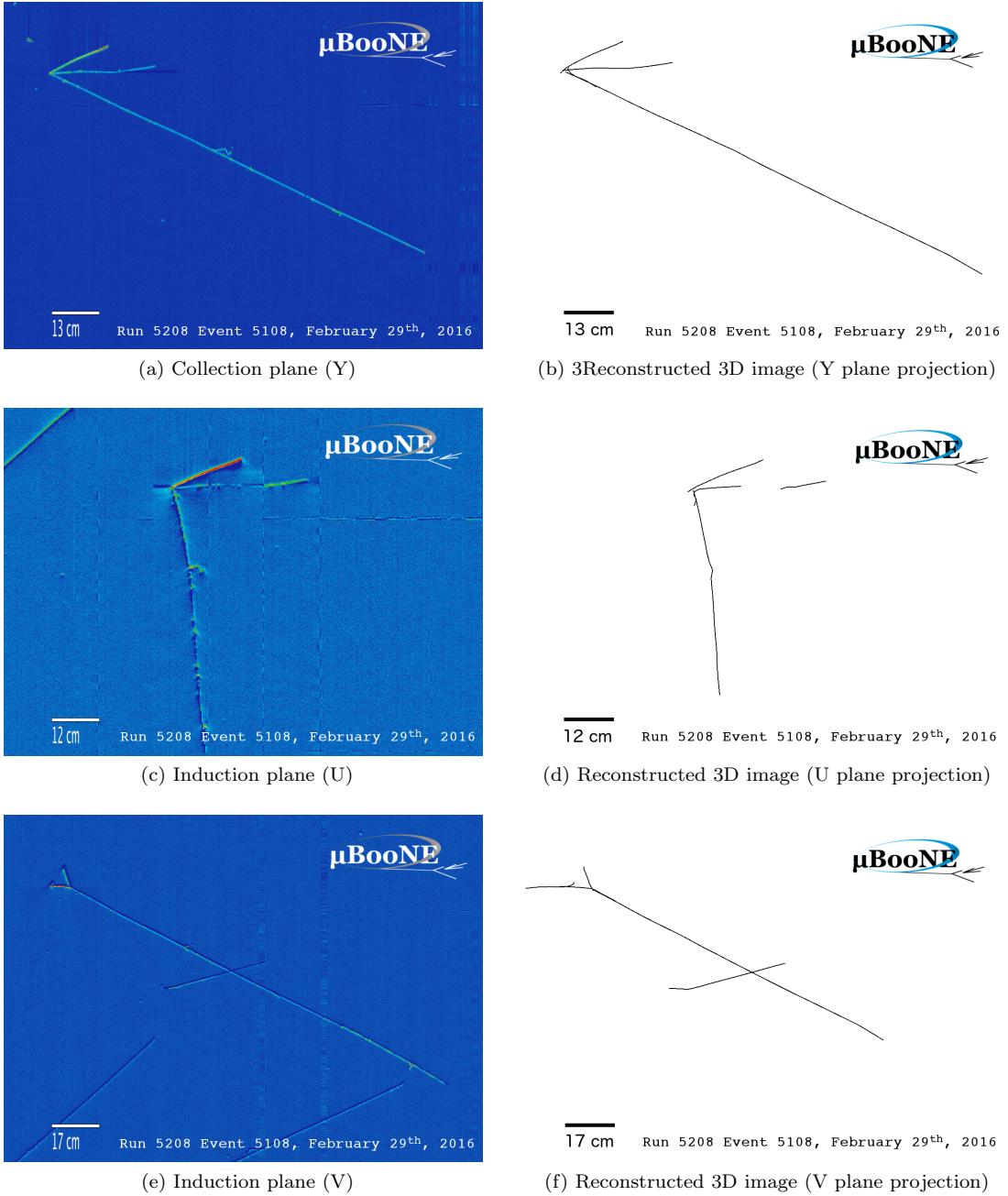


Figure 84: Event view for run 5208, event 5108, selected by selection II. The plots on the left show the event view in all three wire planes. The plots on the right show the three-dimensional reconstructed image projected onto each wire plane. The track reconstruction algorithm shown is pandoraNuPMA. Note that some tracks visible on the left hand plots are missing on the right because they have been rejected in the cosmic removal pass. All figures have the same scale (indicated by the white bar in the bottom left) and aspect ratio.

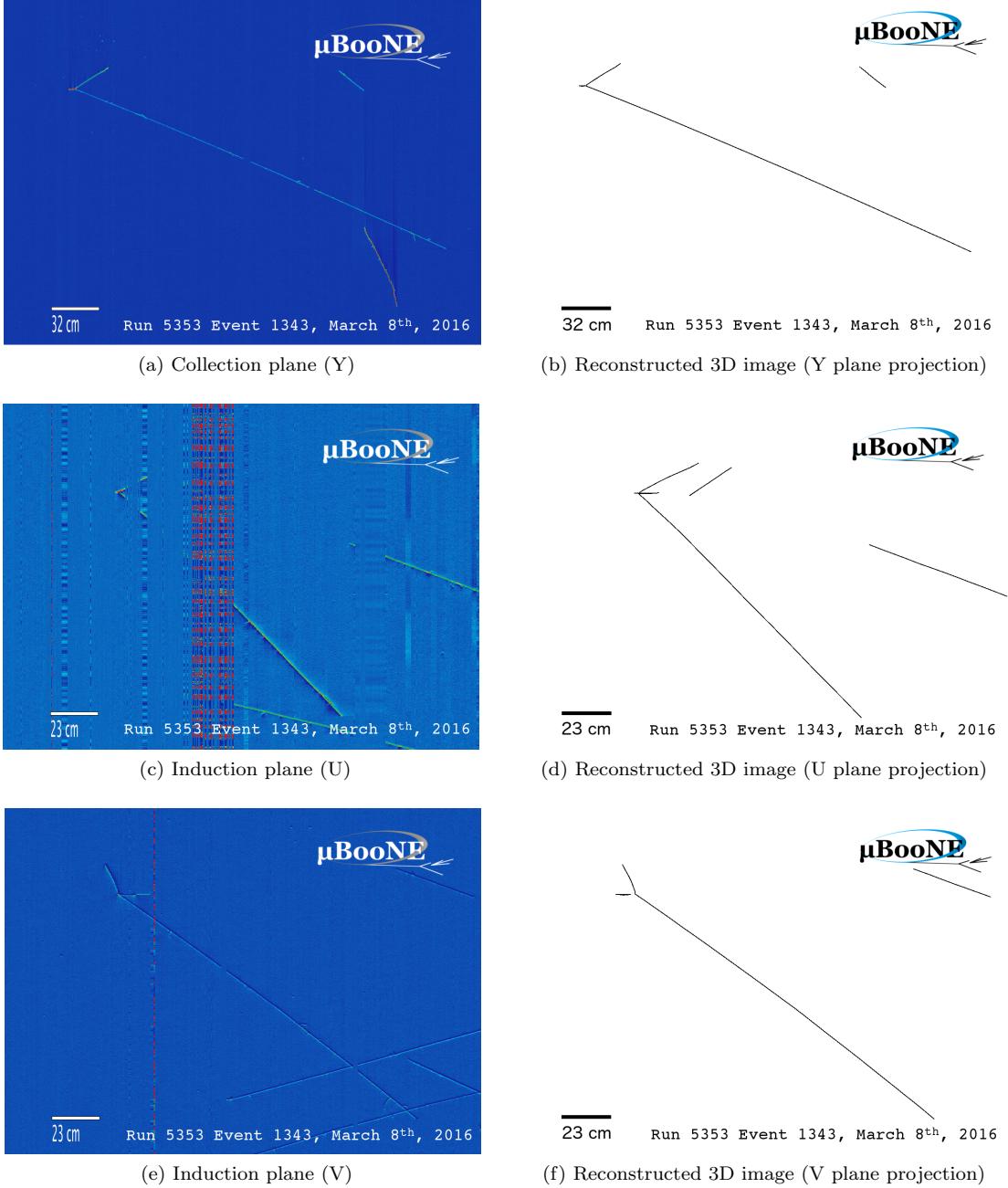


Figure 85: Event view for run 5353, event 1343, selected by selection II. The plots on the left show the event view in all three wire planes. The U plane shows features of gaps of unresponsive wires and noise (vertical red stripes, see 85c). The plots on the right show the three-dimensional reconstructed image projected onto each wire plane. Comparing the reconstructed tracks to the wire plane views on the left it can be seen that the reconstruction successfully bridges gaps of unresponsive wires. The track reconstruction algorithm shown is pandoraNuPMA. Note that some tracks visible on the left hand plots are missing on the right because they have been rejected in the cosmic removal pass. All figures have the same scale (indicated by the white bar in the bottom left) and aspect ratio.

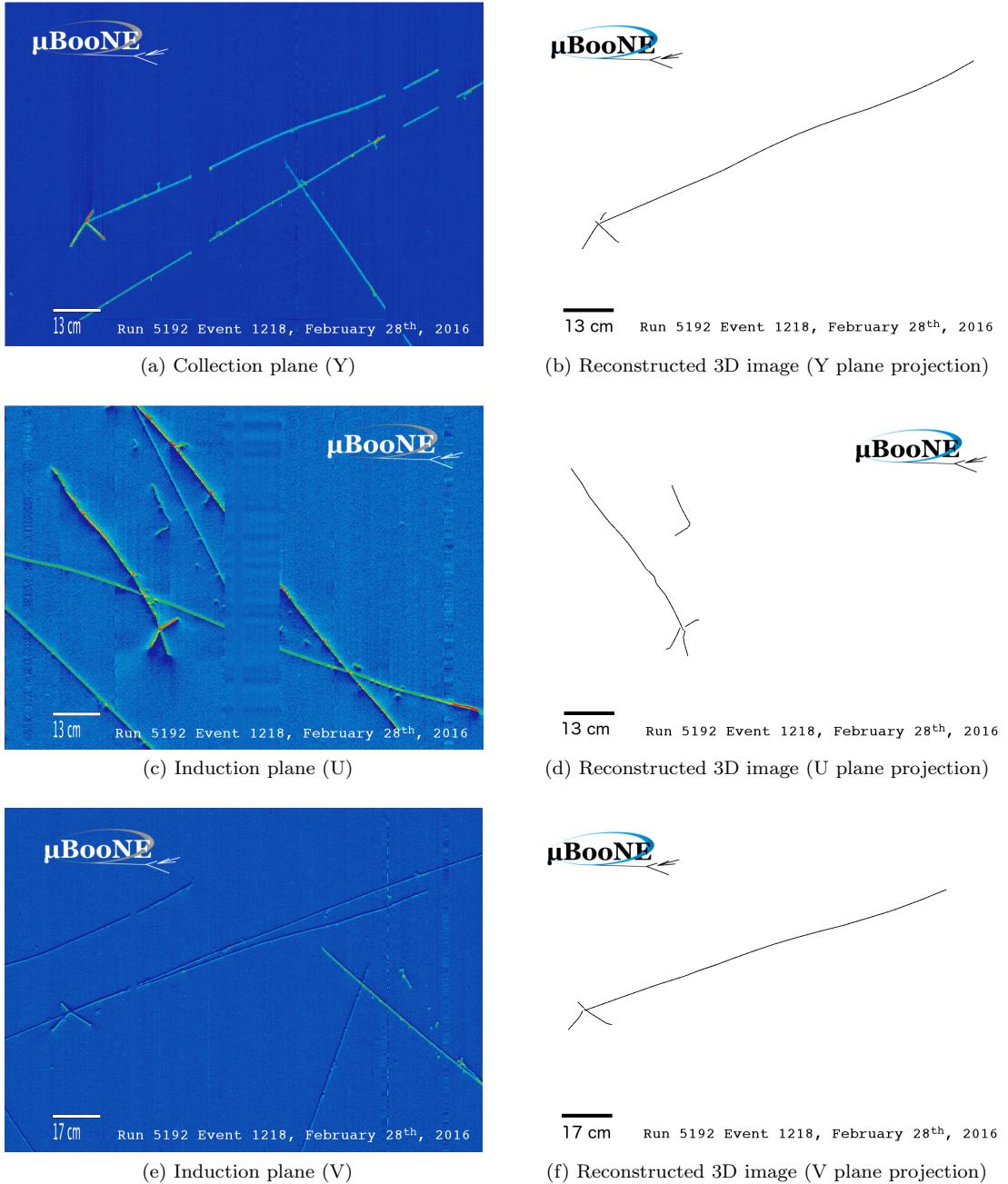
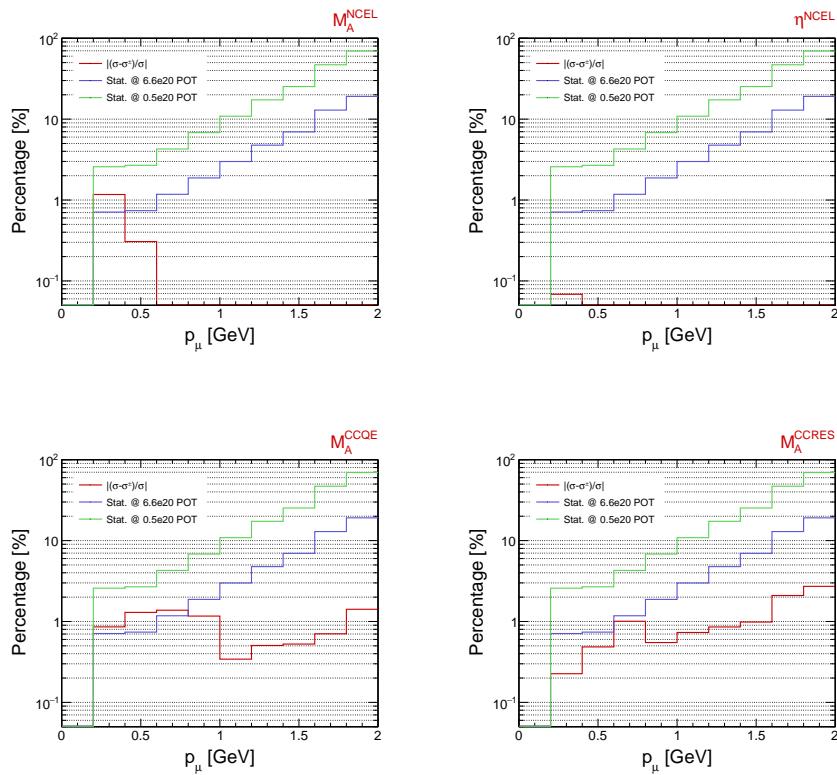
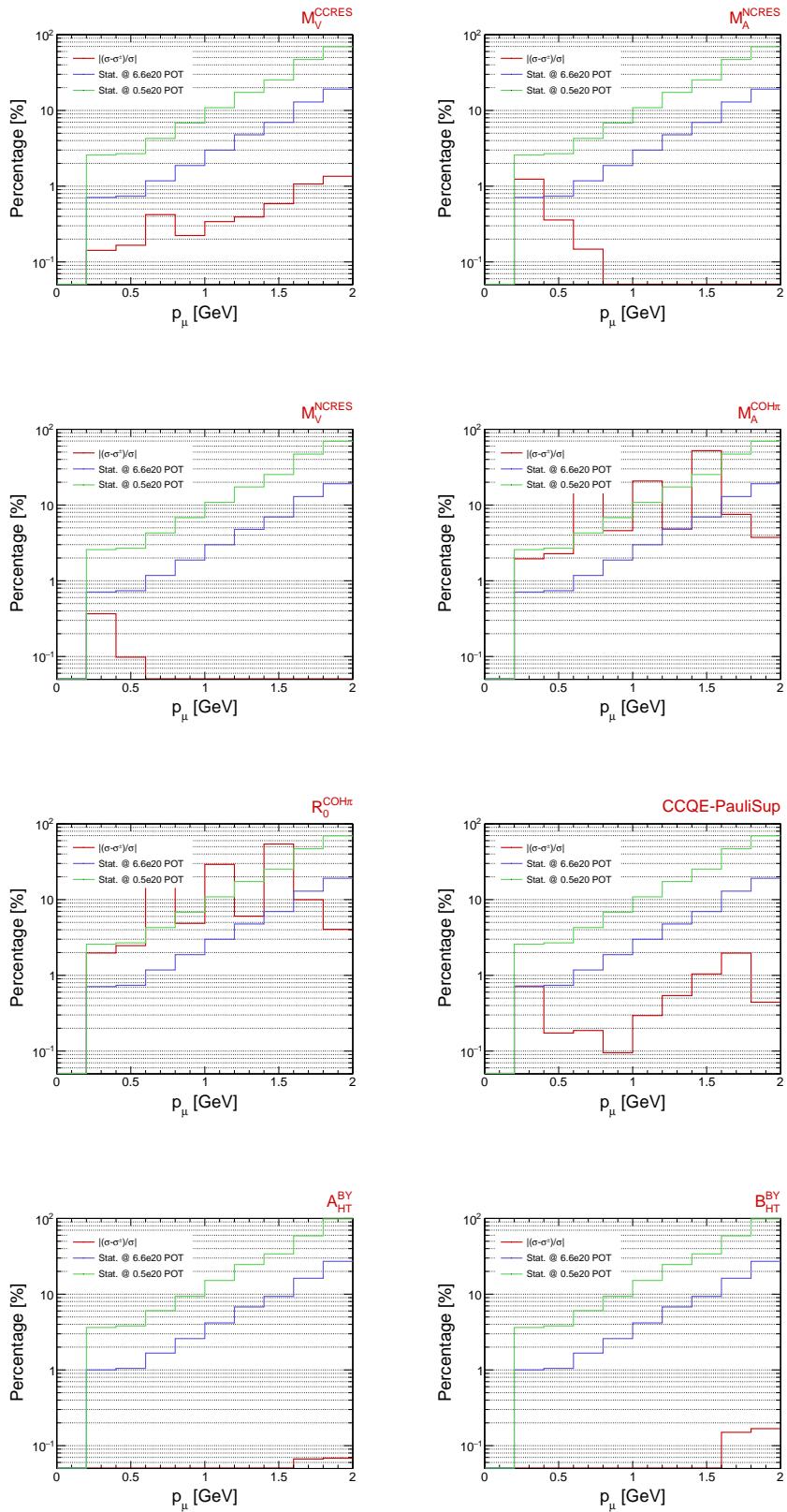


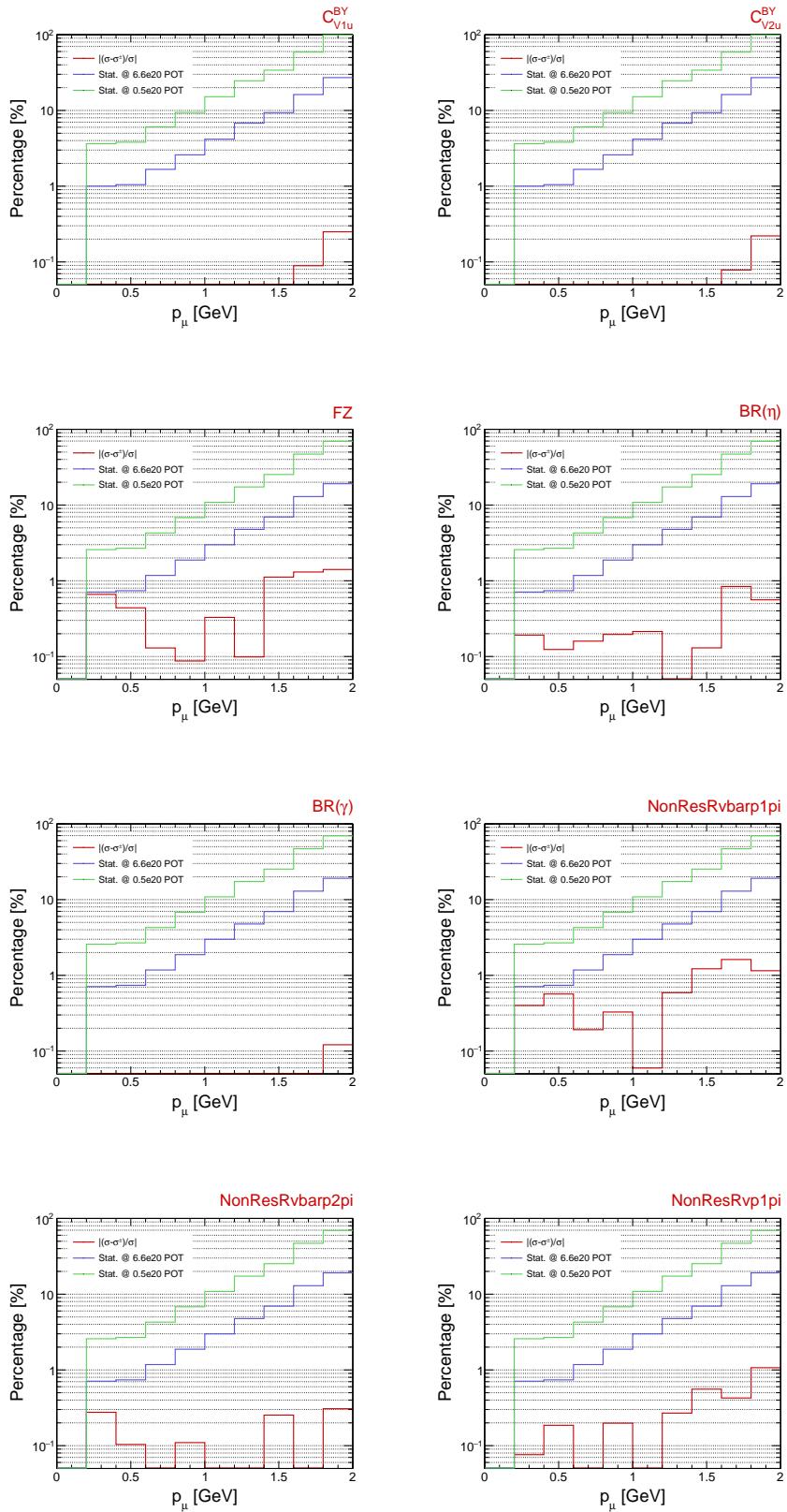
Figure 86: Event view for run 5192, event 1218, selected by selection II. The plots on the left show the event view in all three wire planes. The U plane as well as the Y plane show gaps of unresponsive wires (see 86a and 86c). The plots on the right show the three-dimensional reconstructed image projected onto each wire plane. The track reconstruction algorithm shown is pandoraNuPMA. Note that some tracks visible on the left hand plots are missing on the right because they have been rejected in the cosmic removal pass. All figures have the same scale (indicated by the white bar in the bottom left) and aspect ratio.

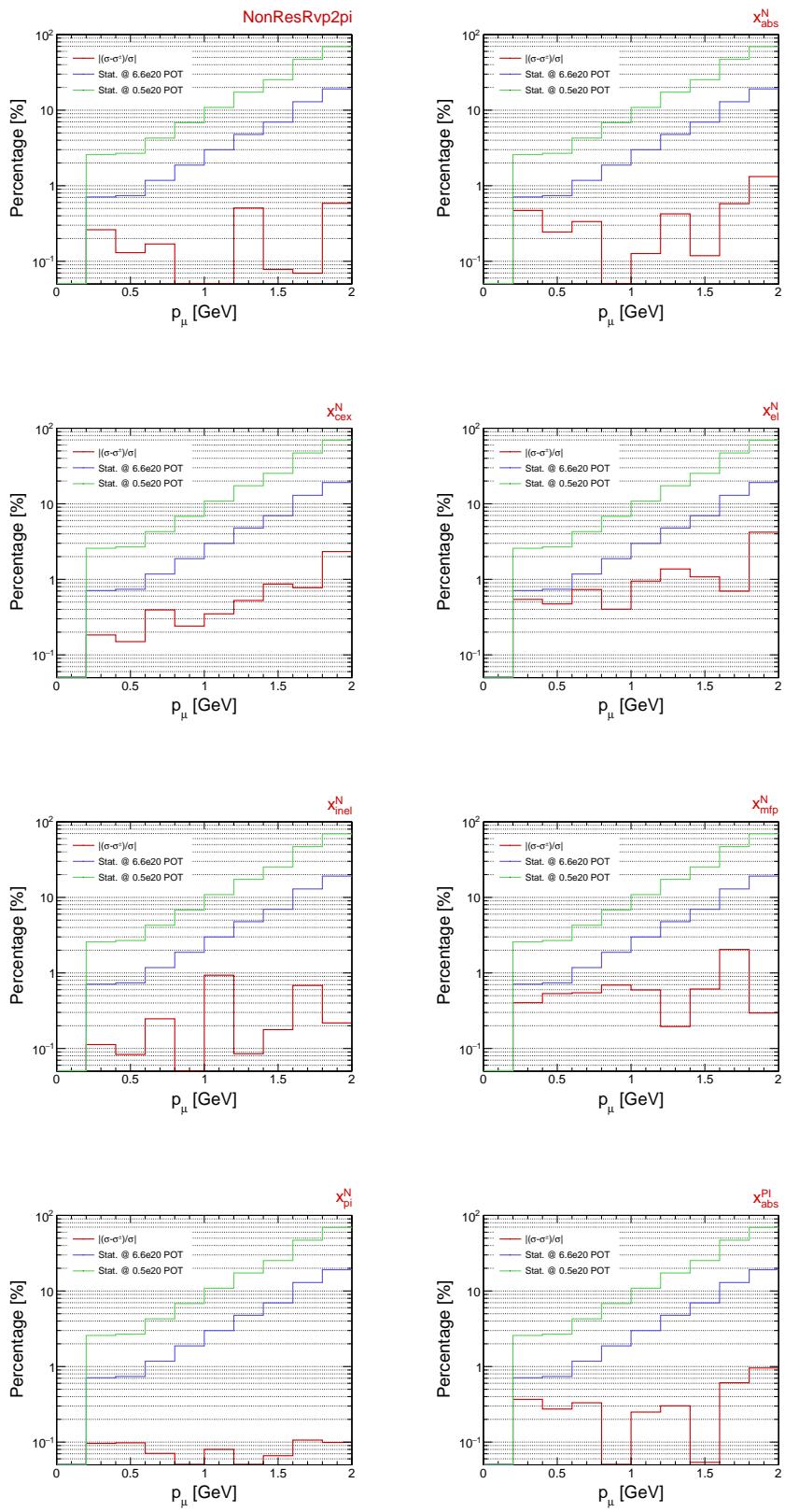
1674 **A GENIE Model Uncertainties - Cross section
1675 difference plots**

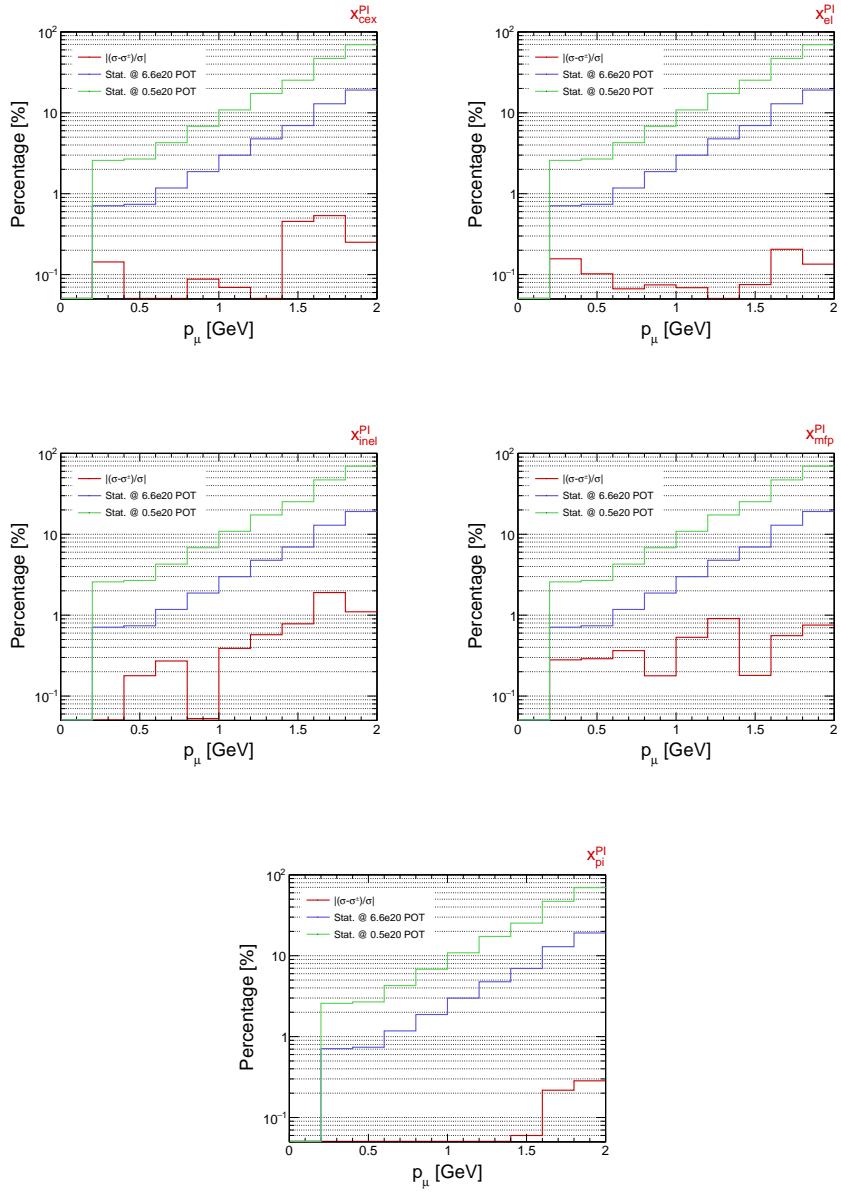
1676 This appendix contains plots showing the cross-section variation as a function
1677 of true muon momentum. One plot for each systematic parameter. See Section
1678 7.2 for details. For each plot, the red line shows the cross percental difference
1679 as one GENIE parameter is increased/decreased by $\pm 1\sigma$ (bin by bin, the bigger
1680 value between these two is chosen). For comparison, green and blue lines show
1681 the value of the statistical uncertainty at 0.5 and 6.6×10^{20} POT respectively.











₁₆₈₂ B Normalized On-Beam and Off-Beam Distributions for selection I
₁₆₈₃

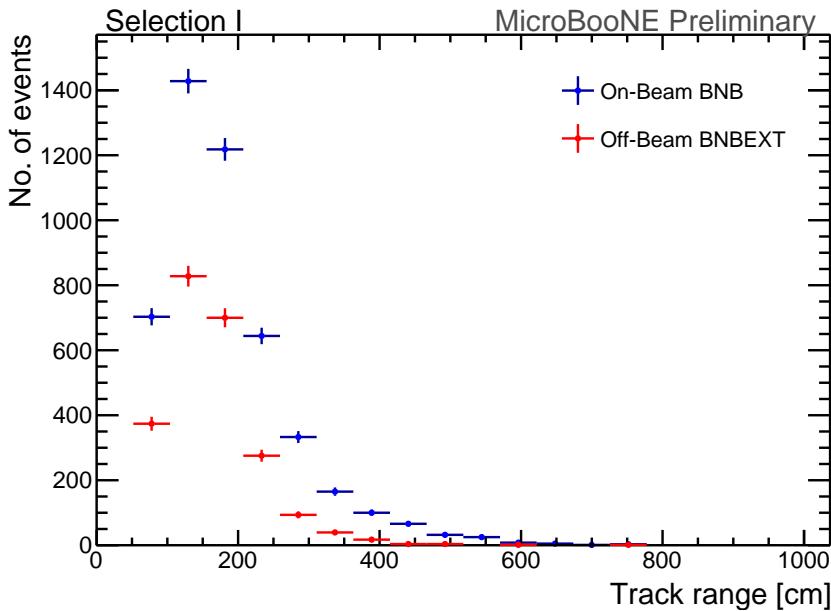


Figure 87: **Selection I: Original** Track range Distribution. The track range is defined as the 3D distance between the start and end point of the muon candidate track. The blue points are the normalized on-beam and the red points the normalized off-beam distribution.

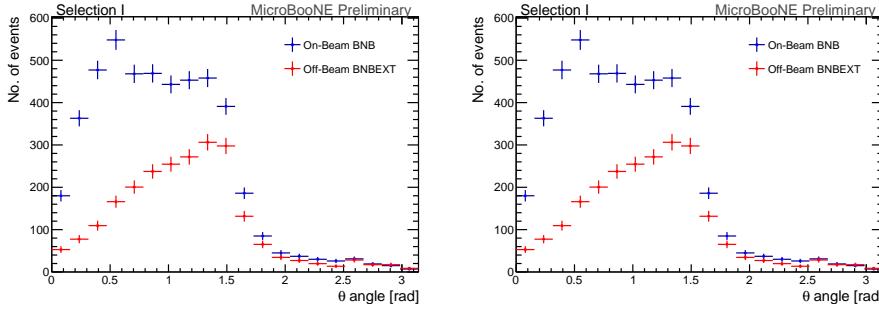


Figure 88: **Selection I: Original** θ angle of track distribution. In the right-hand plot, the distribution has been normalized per solid angle. The angle θ is defined as the angle in beam direction (with $\theta = 0$ beam direction). The blue points are the normalized on-beam and the red points the normalized off-beam distribution.

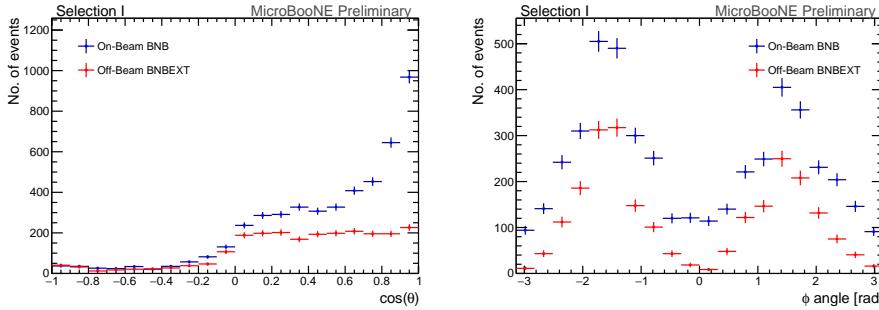


Figure 89: **Selection I: Original** $\cos \theta$ and ϕ angle of track distribution. The angle θ is defined as the angle in beam direction (with $\theta = 0$ beam direction). ϕ is the angle around beam direction with $-\pi/2$ being vertically downward going and $\pi/2$ being vertically upward going tracks. The blue points are the normalized on-beam and the red points the normalized off-beam distribution.

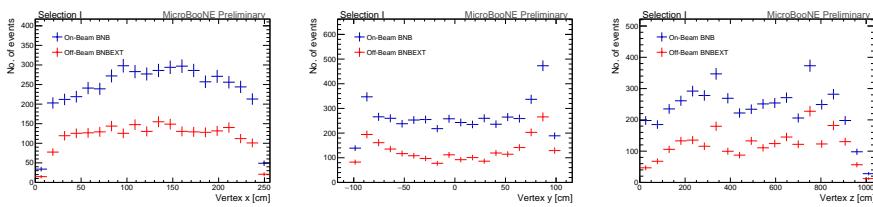


Figure 90: **Selection I: Original** Track start and end point distribution in the x, y, and z-coordinate. The blue points are the normalized on-beam and the red points the normalized off-beam distribution.

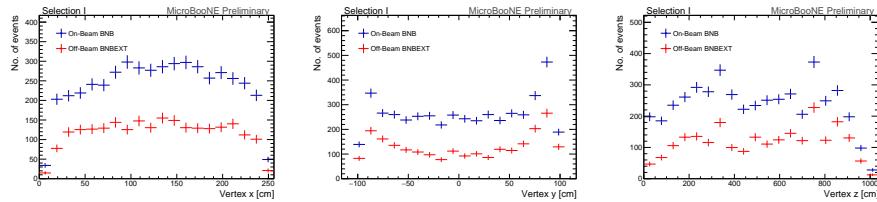


Figure 91: **Selection I: Original** Neutrino interaction candidate vertex distribution in the x, y, and z-coordinate. The blue points are the normalized on-beam and the red points the normalized off-beam distribution.

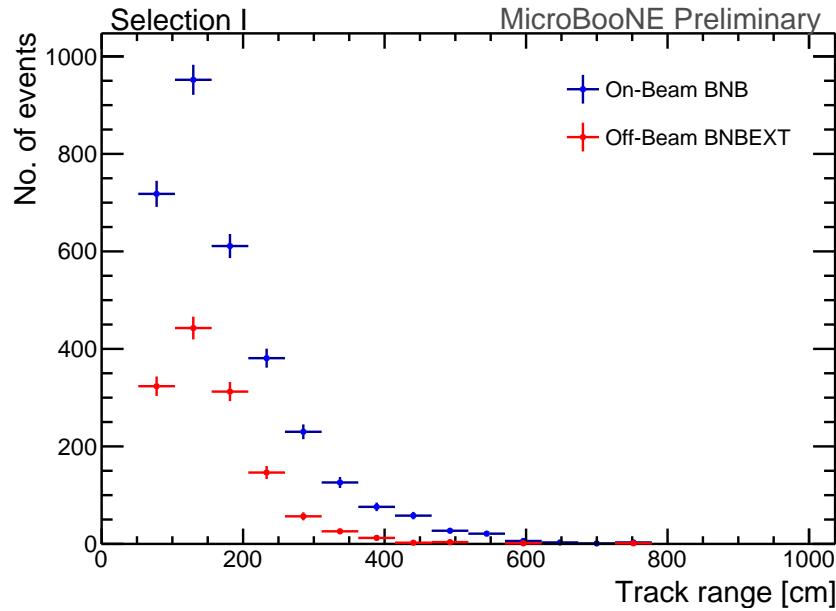


Figure 92: **Selection I: Modified** Track range Distribution. The track range is defined as the 3D distance between the start and end point of the muon candidate track. The blue points are the normalized on-beam and the red points the normalized off-beam distribution.

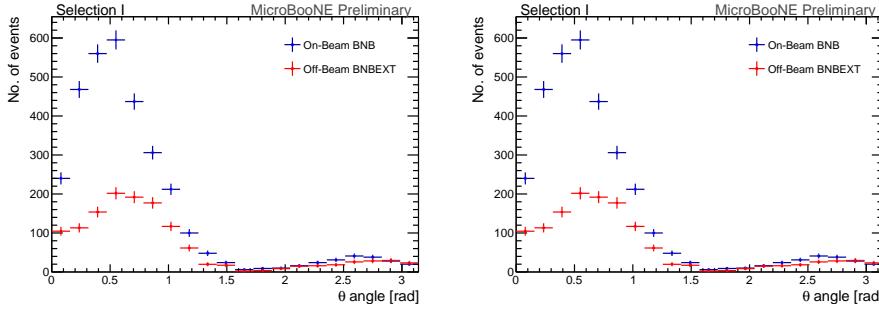


Figure 93: **Selection I: Modified** θ angle of track distribution. In the right-hand plot, the distribution has been normalized per solid angle. The angle θ is defined as the angle in beam direction (with $\theta = 0$ beam direction). The blue points are the normalized on-beam and the red points the normalized off-beam distribution.

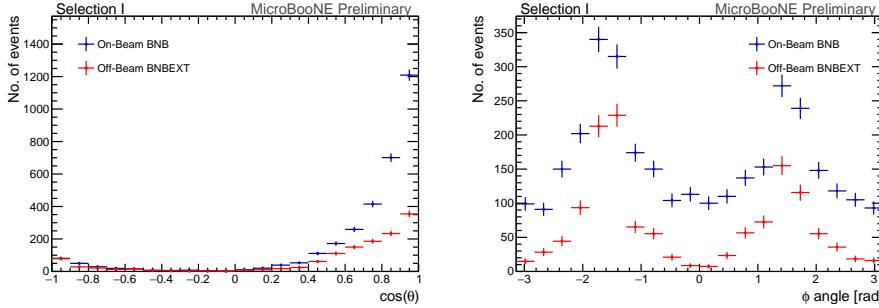


Figure 94: **Selection I: Modified** $\cos\theta$ and ϕ angle of track distribution. The angle θ is defined as the angle in beam direction (with $\theta = 0$ beam direction). ϕ is the angle around beam direction with $-\pi/2$ being vertically downward going and $\pi/2$ being vertically upward going tracks. The blue points are the normalized on-beam and the red points the normalized off-beam distribution.

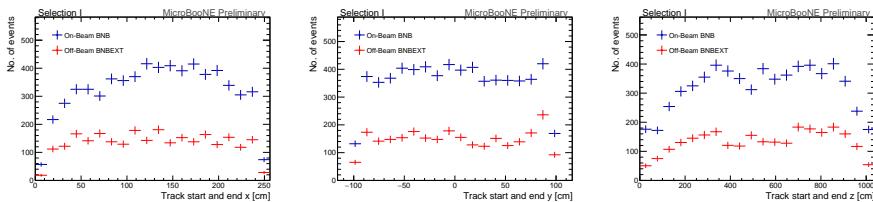


Figure 95: **Selection I: Modified** Track start and end point distribution in the x, y, and z-coordinate. The blue points are the normalized on-beam and the red points the normalized off-beam distribution.

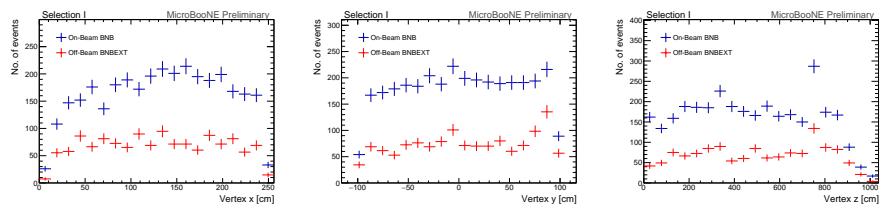


Figure 96: **Selection I:** Modified Neutrino interaction candidate vertex distribution in the x, y, and z-coordinate. The blue points are the normalized on-beam and the red points the normalized off-beam distribution.

1684 **C Normalized On-Beam and Off-Beam Distributions for selection II**

1685

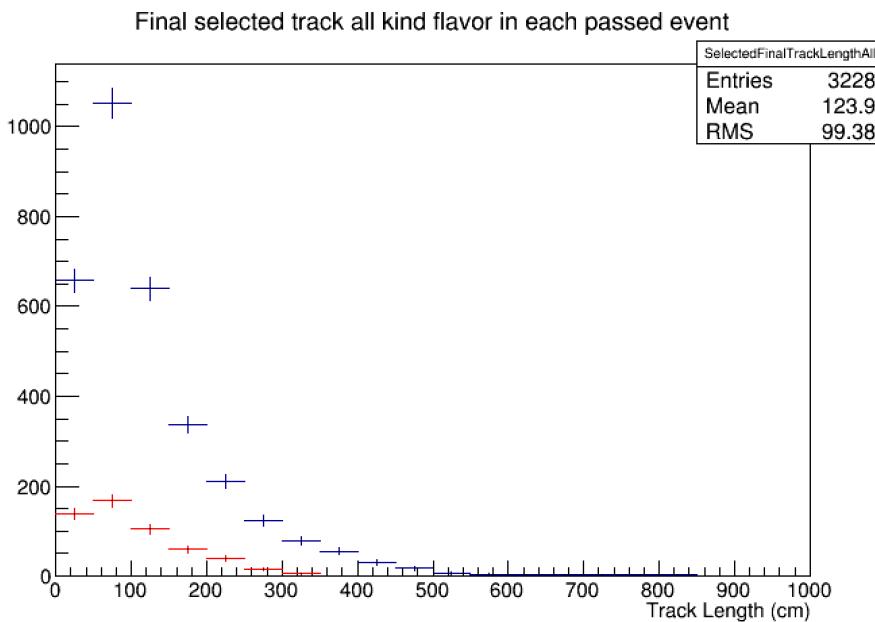


Figure 97: **Selection II: Track length distribution.** Data Beam On (blue) Vs Beam Off(red). Beam off is scaled up by 1.23.

Final selected track all kind flavor in each passed event

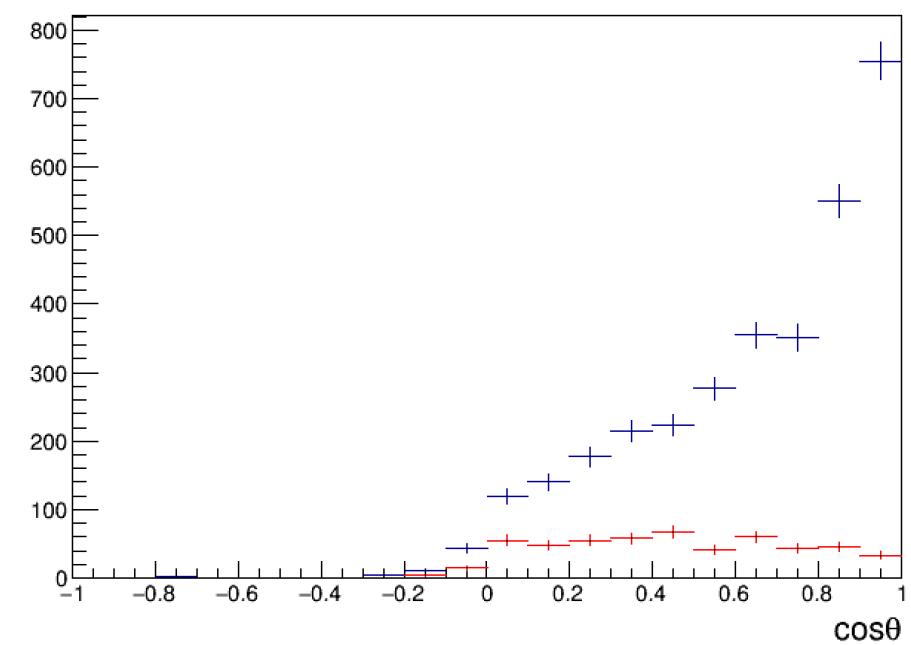


Figure 98: **Selection II: Track $\cos\theta$ distribution.** Data Beam On (blue) Vs Beam Off(red). Beam off is scaled up by 1.23.

Final selected track all kind flavor in each passed event

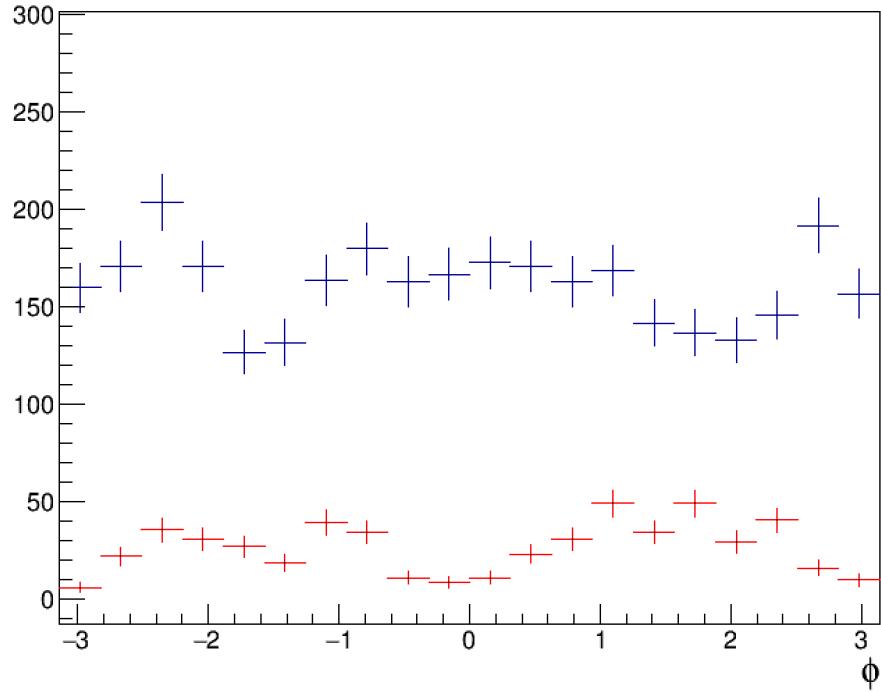


Figure 99: **Selection II: Track ϕ distribution.** Data Beam On (blue) Vs Beam Off(red). Beam off is scaled up by 1.23.

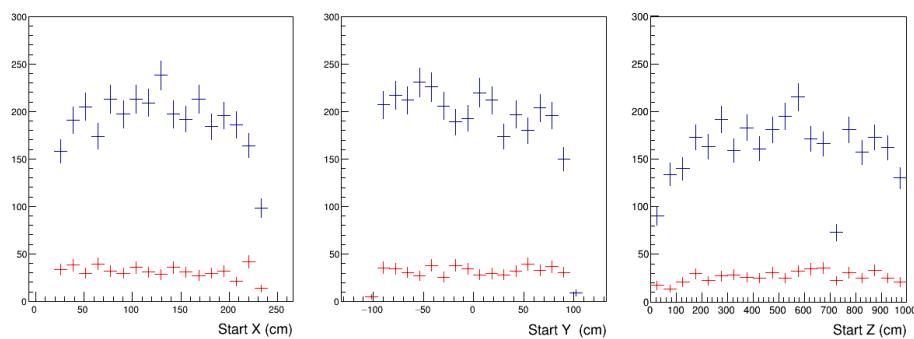


Figure 100: **Selection II: Track Start XYZ** Data Beam On (blue) Vs Beam Off(red). From left to right is start X, Y, and Z. Beam off is scaled up by 1.23.

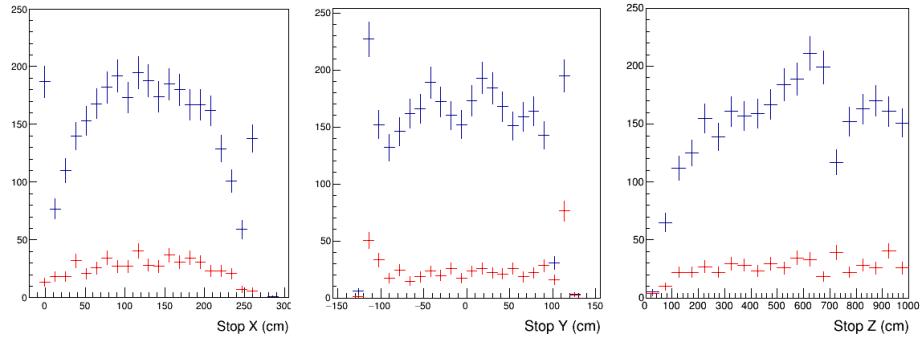


Figure 101: **Selection II: Track Stop XYZ** Data Beam On (blue) Vs Beam Off(red). From left to right is start X, Y, and Z. Beam off is scaled up by 1.23.

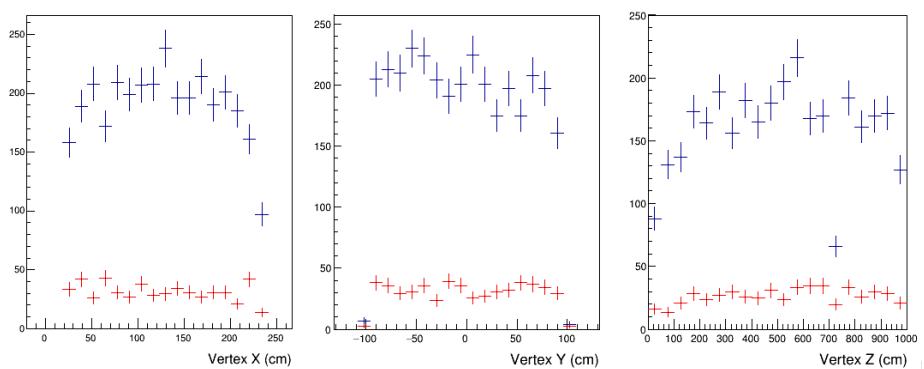


Figure 102: **Selection II: Track Vertex XYZ** Data Beam On (blue) Vs Beam Off(red). From left to right is start X, Y, and Z. Beam off is scaled up by 1.23.

1686 D Selection studies with various reco algorithms

1687 This appendix summaries results from a study presented in DocDB 5778. A
 1688 basic selection very similar to the Selection I Original was performed with dif-
 1689 ferent combinations of track and vertex reconstructions and results for efficiency
 1690 and purity were compared. Table 18 shows selection efficiencies, table 19 shows
 1691 purity comparisons. Based on this, the decisions to use the algorithms were
 1692 made. Please note that the tables don't compare directly to the selections de-
 1693 scribed in the note since smaller changes were applied during the development
 1694 and definitions of purity and efficiency were slightly modified.

Table 18: Selection efficiency for different algorithms.

Track reco	Vertex reco			
	Pandora ν vtx	pandoraCosmic (all)	pandoraNu (all)	pmtrack
pandoraNuKHit	9.3%	9.6%	9.9%	9.2%
pandoraCosmic	10.3%	8.2%	11.0%	10.5%
pandoraNu	10.9%	11.4%	11.4%	10.6%
pmtrack	10.0%	10.7%	10.4%	11.3%
pandoraNuPMA	10.7%	11.2%	11.0%	10.5%
trackkalmanhit	10.4%	7.7%	11.4%	10.4%

Table 19: Selection purity for different algorithms.

Track reco	Vertex reco			
	Pandora ν vtx	pandoraCosmic (all)	pandoraNu (all)	pmtrack
pandoraNuKHit	42.5%	39.1%	42.8%	42.4%
pandoraCosmic	51.7%	52.3%	50.5%	50.3%
pandoraNu	46.7%	47.9%	42.3%	50.9%
pmtrack	56.1%	50.0%	53.7%	45.7%
pandoraNuPMA	49.8%	52.4%	48.8%	51.5%
trackkalmanhit	49.7%	33.5%	48.1%	51.1%

1695 **References**

- 1696 [1] T. Kobilarcik, T. Miceli, Z. Pavlovic. BNB Protons on Target for Micro-
1697 BooNE. *MicroBooNE internal note (docdb 5719)*, 2016.
- 1698 [2] R.C.Fernandez, T. Miceli, Z. Pavlovic, A. Wickremasinghe. Booster Neu-
1699 trino Beam Flux at MicroBooNE. *MicroBooNE internal note (docdb 5787)*,
1700 2016.
- 1701 [3] C. Andreopoulos et al. The GENIE Neutrino Monte Carlo Generator. *Nucl.*
1702 *Instrum. Meth.*, A614:87–104, 2010. doi: 10.1016/j.nima.2009.12.009.
- 1703 [4] Costas Andreopoulos, Christopher Barry, Steve Dytman, Hugh Gallagher,
1704 Tomasz Golan, Robert Hatcher, Gabriel Perdue, and Julia Yarba. The
1705 GENIE Neutrino Monte Carlo Generator: Physics and User Manual. 2015.
- 1706 [5] A. A. Aguilar-Arevalo, et al. First measurement of the muon neutrino
1707 charged current quasielastic double differential cross section. *Phys. Rev.*
1708 *D*, 81:092005, May 2010. doi: 10.1103/PhysRevD.81.092005. URL <http://link.aps.org/doi/10.1103/PhysRevD.81.092005>.
- 1710 [6] The MicroBooNE Collaboration. Measurement of the Electronegative Con-
1711 taminants and Drift Electron Lifetime in the MicroBooNE Experiment. *Mi-*
1712 *croBooNE public note (docdb 4928)*, 2015.