2GeV exclusive cross section measurement: Selection update

Shyam Bhuller

University of Bristol

April 29, 2024

Introduction

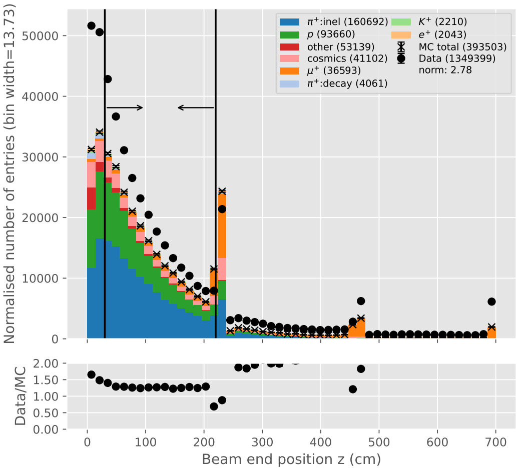
- first talk
- Beam particle selection updated to include fiducial cuts.
- Region selection categorises events by the exclusive process using reco information.
- Adopted Kang's γ, $\pi^{\pm}$ and π^{0} selection, found that in 2GeV, most regions are largely background dominated $\rightarrow$ Affects the fit and background subtraction performance.
- Updated region selection to better suit the 2GeV analysis.
- Studying 2GeV PDSP Data and MC.

	Short name	**Description**
Absorption	abs	$\pi^{+} + \text{Ar} \rightarrow 0\pi^{\pm} + 0\pi^{0} + X$
Charge exchange	cex	$\pi^{+} + \text{Ar} \rightarrow 0\pi^{\pm} + 1\pi^{0} + X$
Single pion production	spip	$\pi^{+} + \text{Ar} \rightarrow 1\pi^{\pm} + 0\pi^{0} + X$
Pion production	pip	$\pi^{+} + \text{Ar} \rightarrow \geq 1\pi^{\pm} + \geq 1\pi^{0} + X$
		$\pi^{+} + \text{Ar} \rightarrow 0\pi^{\pm} + > 1\pi^{0} + X$
		$\pi^{+} + \text{Ar} \rightarrow > 1\pi^{\pm} + 0\pi^{0} + X$

Beam particle selection

fiducial region

- Fiducial region for the beam interaction defined as 30 cm $< z <$ 220 cm, both in reco and truth
- In reco, this excludes events past APA3, conveniently vetos muons and the first 30cm of the TPC contains many backgrounds which can be excluded.



Beam trigger + preselection

- Information from the beam instrumentation is used to select pion and muon like beam particles → reduces proton background.
- After beam selection and defining the fiducial region, sample purity is close to 90%.
- Preselection ensures that selected events:
    - Have calorimetric information.
    - Have final state PFOs reconstructed (otherwise they can't be categorised).

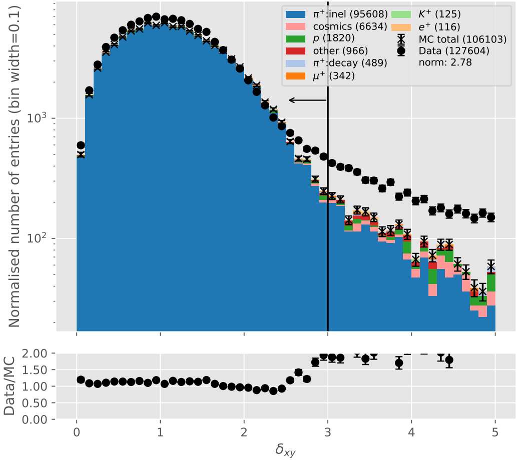
	Remaining events	π^+:inel	π^+:decay	μ^+	e^+	p	K^+	other	cosmics
No selection	141548	57803	1461	13163	735	33691	795	19115	14785
Fiducial region	56621	35177	217	166	58	16804	393	442	3364
Beam trigger	38931	35020	182	137	43	706	47	375	2421
Preselection	38112	34342	176	123	42	654	45	347	2383

Beam quality δ_{xy}

- Reject beam particles with large deviations in the lateral position.

$$\delta_{xy} = \sqrt{\left(\frac{x-\mu_x}{\sigma_x}\right)^2 + \left(\frac{y-\mu_y}{\sigma_y}\right)^2} \qquad (1)$$

- μ_i and σ_i are obtained from gaussian fits to the truncated start x and y positions (see backups).



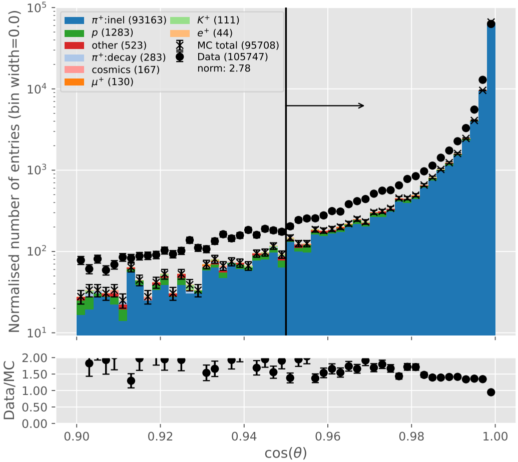
	μ_x	μ_y	σ_x	σ_y
MC	-36.91 ± 0.03	416.25 ± 0.03	3.89 ± 0.04	3.64 ± 0.04
Data	-33.01 ± 0.02	417.59 ± 0.01	4.13 ± 0.02	3.83 ± 0.02

Beam quality $\cos(\theta)$

- Make selection on beam trajectory $\hat{n}$

$$\cos(\theta) = \hat{n}_x \mu_{\hat{n}_x} + \hat{n}_y \mu_{\hat{n}_y} + \hat{n}_z \mu_{\hat{n}_z}. \quad (2)$$

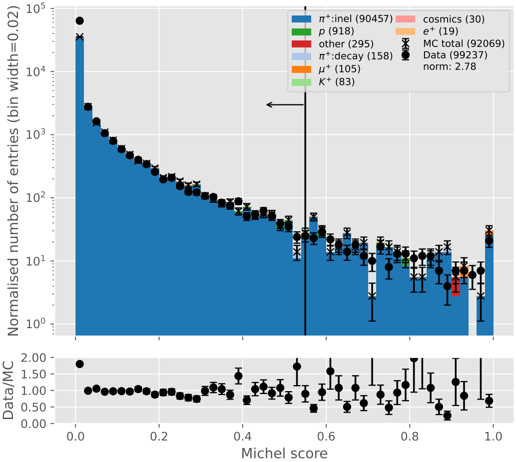
- $\mu_{\hat{n}^i}$ is the average value of the direction component.



	$\boldsymbol{\mu_{\hat{n}_x}}$	$\boldsymbol{\mu_{\hat{n}_y}}$	$\boldsymbol{\mu_{\hat{n}_z}}$
MC	-0.1828 ± 0.0007	-0.183 ± 0.001	0.9150 ± 0.0008
Data	-0.1531 ± 0.0005	-0.2036 ± 0.0008	0.8912 ± 0.0005

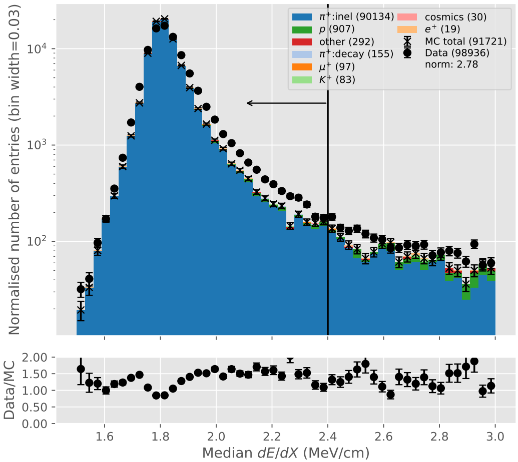
Michel score

- Michel score output from CNN can be used to distinguish pions from muons.
- At 2GeV, muons travel further into the TPC before stopping, so fiducial region actually does a really good job of excluding muons.



Median dE/dX

- Use mean energy loss to distinguish protons from pions.
- Again, fiducial region results in many less misidentified beam tracks, so selection has minimal impact on selection purity.

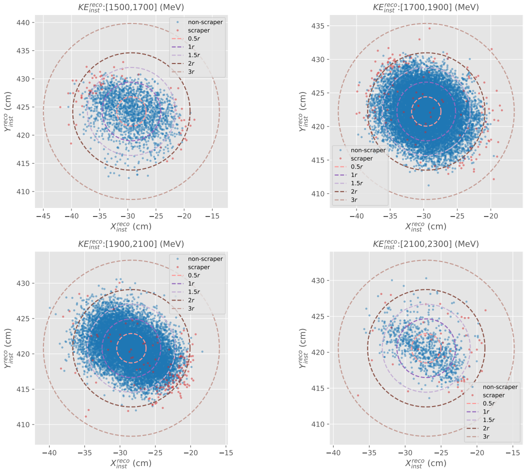


Beam Scraper

- Exclude events with inconsistent beam momenta measured by the beam instrumentation.
- Use same method in this talk to identify beam scrapers in MC.
- Beam scraper effect noticeably smaller than 1GeV (check slide 7)
- Cut on normalised radial position.

$$r_{inst} = \sqrt{\left(\frac{X^{reco}_{inst} - \mu_X}{\sigma_X}\right)^2 + \left(\frac{Y^{reco}_{inst} - \mu_Y}{\sigma_Y}\right)^2}, \quad (3)$$

- μ_i and σ_i are arithmetic mean and standard deviation of positions measured by beam instrumentation.
- Set cut value at $r_{inst} < 1.5$.



Updated selection

	Counts	Purity (%)	Efficiency (%)
No selection	57803	40.8	100.0
30 cm $<$ Reco beam end z position $<$ 220 cm	40073	53.9	69.3
30 cm $<$ True beam end z position $<$ 220 cm	35177	62.1	60.9

	Counts	Purity (%)	Efficiency (%)
Fiducial region	35177	62.1	100.0
Beam trigger	35020	90.0	99.6
Preselection	34342	90.1	97.6
$\delta_{xy} < 3$	33464	97.3	95.1
$\cos(\theta) > 0.95$	32492	98.2	92.4
Michel score $<$ 0.55	32376	98.3	92.0
Median $dE/dX < 2.4$ MeV/cm	30794	98.9	87.5
$r_{inst} < 1.5$	19550	98.9	55.6

$$\text{purity} = \frac{\text{number of beam pions selected}}{\text{total number of particles selected}}, \tag{4}$$

$$\text{efficiency} = \frac{\text{number of beam pions selected}}{\text{total number of beam pions in the fiducial region}}. \tag{5}$$

- Tables are of 2GeV MC.
- Selection is very pure, efficiency in the fiducial region is 55.6%, total efficiency (counting true beam pions outside fiducial region) is 35%.
- michel score, and median dE/dX have minimal impact on selection, beam scraper results in large reduction in efficiency.

Counts in Data and MC

	data	mc
No selection	1349399	141548
30 cm $<$ Reco beam end z position $<$ 220 cm	263960	74313
30 cm $<$ True beam end z position $<$ 220 cm	263960	56621

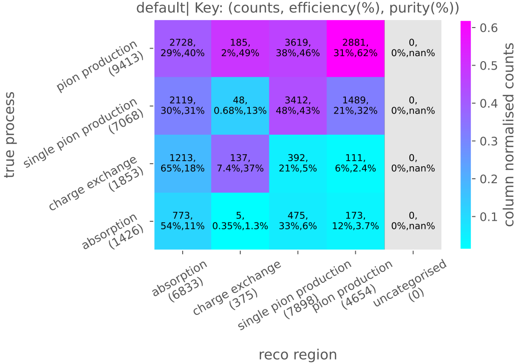
	data	mc
No selection	263960	56621
Beam trigger	132008	38931
Preselection	127604	38112
$\delta_{xy} < 3$	105747	34378
$\cos(\theta) > 0.95$	99237	33071
Michel score $<$ 0.55	98936	32946
Median $dE/dX < 2.4$ MeV/cm	92588	31135
$r_{inst} < 1.5$	56451	19762

PFO Selection

Old approach

- Kang's approach was to select $\pi^{\pm}$ and π^{0}s, then assign events to regions by counting the number of selected pions.

	$\pi^{\pm}$	π^{0}
abs	0	0
cex	0	1
spip	1	0
pip	>1	≥ 0
	>0	1



- No events are uncategorised.
- Abs, cex and spip selections are not great.
- Performance of selection impact performance of fit + background subtraction.

New selections

- Strict particle selection to select PFOs with high purity. Useful when we need to positively ID objects.
- Loose particle selection to select PFOs with high efficiency. Useful to veto objects

loose γ
$(\chi^2/ndf)_p > 61.2$
track score < 0.45
nHits > 31
$d < 114$
$b < 80$

strict γ
$(\chi^2/ndf)_p > 61.2$
track score < 0.45
nHits > 80
$3 < d < 90$
$b < 20$

loose $\pi^{\pm}$
$(\chi^2/ndf)_p > 61.2$
track score > 0.39
median $dE/dX < 6.3$

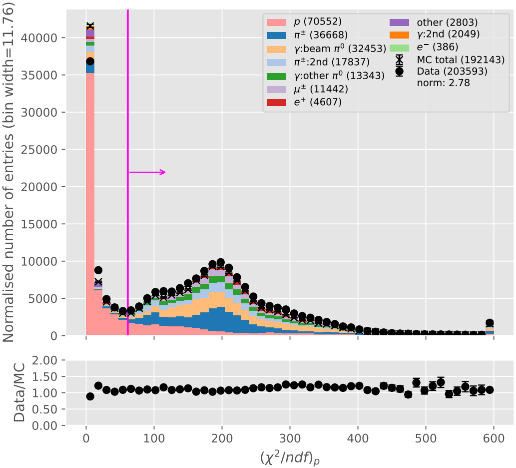
strict $\pi^{\pm}$
$(\chi^2/ndf)_p > 61.2$
track score > 0.5
nHits > 20
0.5 < median $dE/dX < 2.8$

π^0
strict γ = 2
$m_{\gamma\gamma} < 225$
$\theta < 60$

- Note that strict particles are a subset of the loose particles.

Proton veto

- Exclude large majority of protons using $(\chi^2/ndf)_p =$"**`reco_daughter_allTrack_Chi2_proton`** / **`reco_daughter_allTrack_Chi2_ndof`**"
- Same cut is used for **all** particle selections.

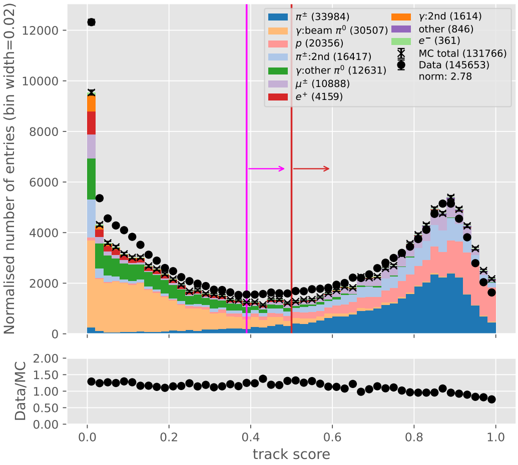


$\pi^{\pm}$ selection: track score

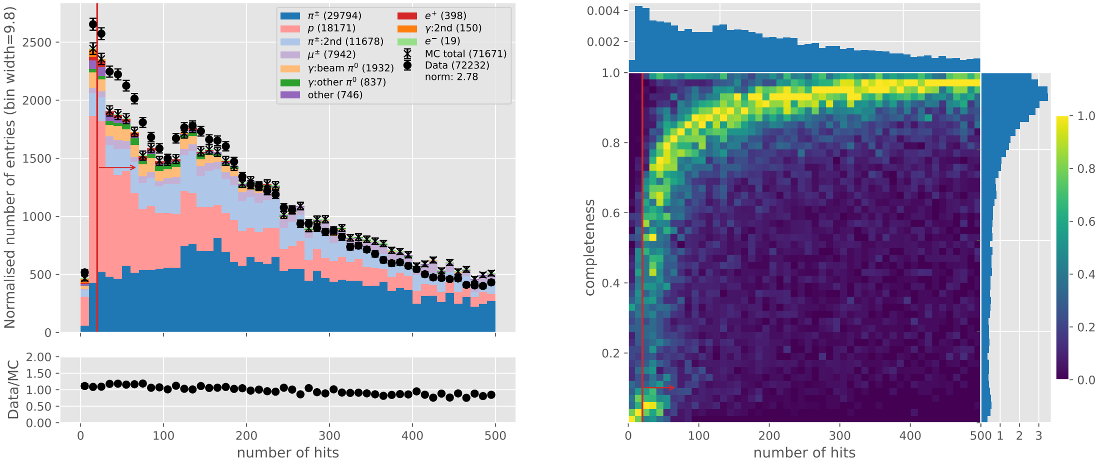
loose $\pi^{\pm}$
$(\chi^2/ndf)_p > 61.2$
track score > 0.39
median $dE/dX < 6.3$

strict $\pi^{\pm}$
$(\chi^2/ndf)_p > 61.2$
track score > 0.5
nHits > 20
$0.5 <$ median $dE/dX < 2.8$

- Use CNN track score output to select track-like PFOs.
- For the loose $\pi^{\pm}$, cut indicated by magenta arrow, strict $\pi^{\pm}$ indicated by red.



$\pi^{\pm}$ selection: nHits



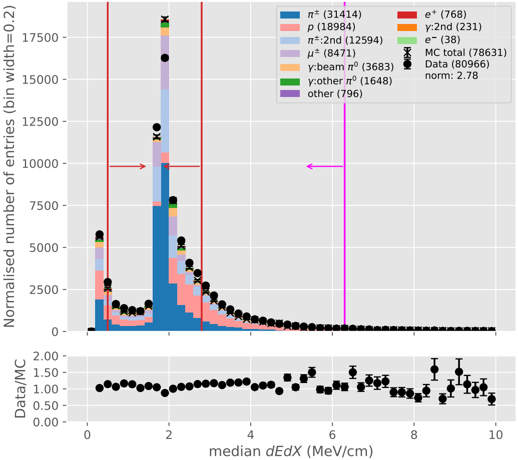
- Number of hits excludes some low completeness PFOs.

$$\text{completeness} = \frac{\text{number of hits in the PFO}}{\text{number of hits produced by the true particle associated to the PFO}}, \tag{6}$$

- **Cut not used for loose $\pi^{\pm}$ selection.**

$\pi^{\pm}$ selection: median dE/dX

- Proton still some significant proton background, so cut on median dE/dX.



for the loose $\pi^{\pm}$, cut indicated by magenta arrow, strict $\pi^{\pm}$ indicated by red

Performance

loose $\pi^{\pm}$

	Counts	Purity (%)	Efficiency (%)
Beam particle selection	13171	19.1	100.0
$(\chi^2/ndf)_p > 61.2$	12207	25.8	92.7
track score > 0.39	11284	40.0	85.7
Median $dE/dX < 6.3$ **MeV/cm**	11229	40.6	85.3

strict $\pi^{\pm}$

	Counts	Purity (%)	Efficiency (%)
Beam particle selection	13171	19.1	100.0
$(\chi^2/ndf)_p > 61.2$	12207	25.8	92.7
track score > 0.5	10702	41.6	81.3
Number of hits > 20	10500	42.7	79.7
0.5 **MeV/cm** $<$ **Median** $dE/dX < 2.8$ **MeV/cm**	8823	48.0	67.0

$$\text{Purity} = \frac{\text{selected PFOs which are } \pi^{\pm} \text{ from the beam interaction}}{\text{total number of PFOs selected}}, \quad (7)$$

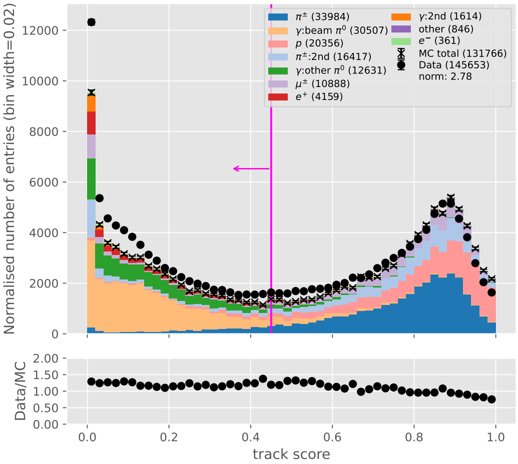
$$\text{Efficiency} = \frac{\text{selected PFOs which are } \pi^{\pm} \text{ from the beam interaction}}{\text{number of } \pi^{\pm} \text{ from the beam interaction}}. \quad (8)$$

- Loose $\pi^{\pm}$ selection efficiency is high, but the strict $\pi^{\pm}$ purity is below 50%.
- At 2GeV, more pion production and single pion production occurs, so in general, higher selection efficiency is favoured.

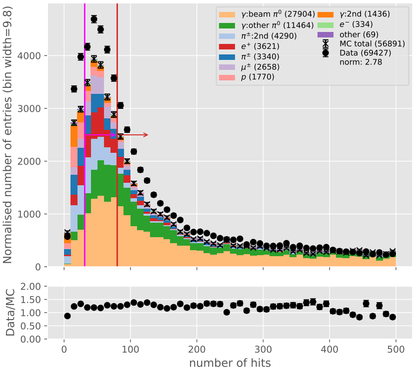
γ selection: track score

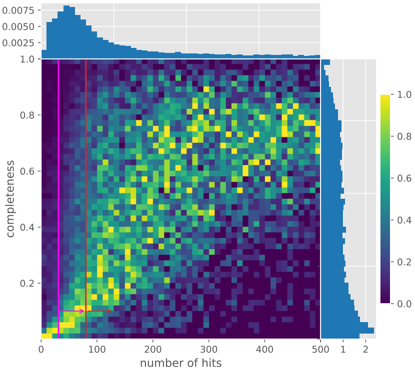
loose γ	strict γ
$(\chi^2/ndf)_p > 61.2$	$(\chi^2/ndf)_p > 61.2$
track score < 0.45	track score < 0.45
nHits > 31	nHits > 80
$d < 114$	$3 < d < 90$
$b < 80$	$b < 20$

- Start with track score to select shower like objects.
- Same cut is used both for strict and loose selections.



γ selection: nHits

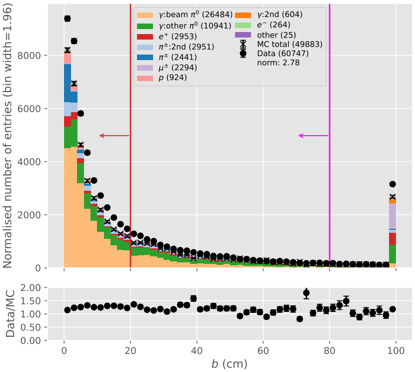


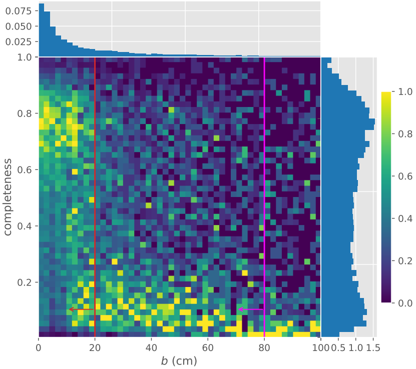


- Number of hits removes low purity objects, which tend to be tracks with low number of hits.

for the loose γ, cut indicated by magenta arrow, strict γ indicated by red

γ selection: impact parameter





- Impact parameter of shower vs beam interaction vertex is:

$$b = |(\vec{r}_{beam} - \vec{s}_{shower}) \times \hat{n}_{shower}|. \quad (9)$$

- Removes low completeness objects as well as cosmics and other photons which don't come from a beam π^0.

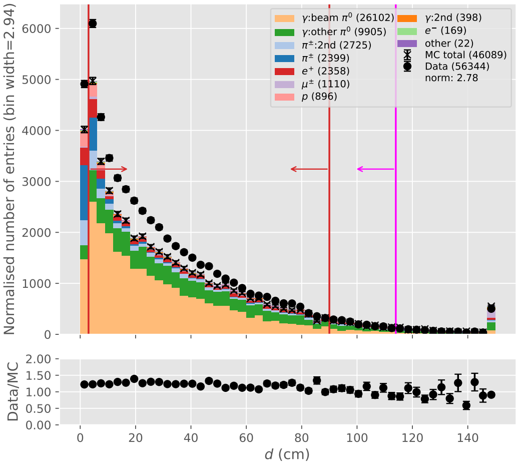
for the loose γ, cut indicated by magenta arrow, strict γ indicated by red

γ selection: distance to beam vertex

- Distance from PFO start to beam vertex is

$$d = |\vec{r}_{beam} - \vec{s}_{shower}|. \qquad (10)$$

- Photons will travel some distance before producing a shower.
- PFOs far away from the interaction vertex are less likely to have been produced from the beam interaction.



for the loose γ, cut indicated by magenta arrow, strict γ indicated by red

Performance

loose γ	Counts	Purity (%)	Efficiency (%)
Beam particle selection	11657	16.9	100.0
$(\chi^2/ndf)_p > 61.2$	10958	23.2	94.0
track score < 0.45	10023	49.0	86.0
Number of hits > 31	9513	53.1	81.6
$b < 80$ **cm**	9376	56.6	80.4
$d < 114$ **cm**	9280	57.7	79.6

strict γ	Counts	Purity (%)	Efficiency (%)
Beam particle selection	11657	16.9	100.0
$(\chi^2/ndf)_p > 61.2$	10958	23.2	94.0
track score < 0.45	10023	49.0	86.0
Number of hits > 80	7281	61.7	62.5
$b < 20$ **cm**	6064	69.7	52.0
3 cm $< d < 90$ **cm**	5601	71.7	48.0

$$\text{Purity} = \frac{\text{selected PFOs which are } \gamma \text{ from a beam } \pi^0}{\text{total number of PFOs selected}}, \tag{11}$$

$$\text{Efficiency} = \frac{\text{selected PFOs which are } \gamma \text{ from a beam } \pi^0}{\text{number of } \gamma \text{ from a beam } \pi^0}. \tag{12}$$

- Strict selection has good purity and loose selection has decent purity and good efficiency.

π^0 selection

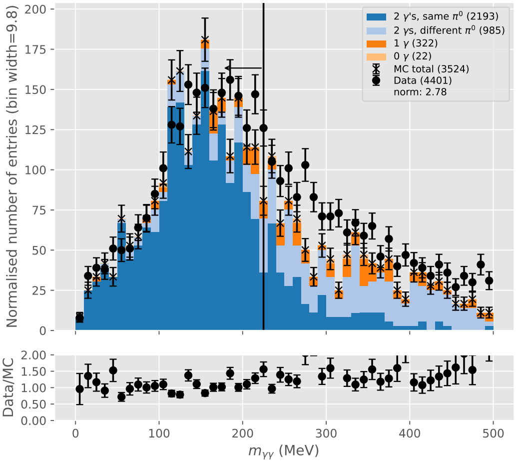
Invariant mass

π^0
strict γ = 2
$m_{\gamma\gamma} < 225$
$\theta < 60$

- Events with exactly 2 strict γ selection are candidates for π^0s.
- Invariant mass is

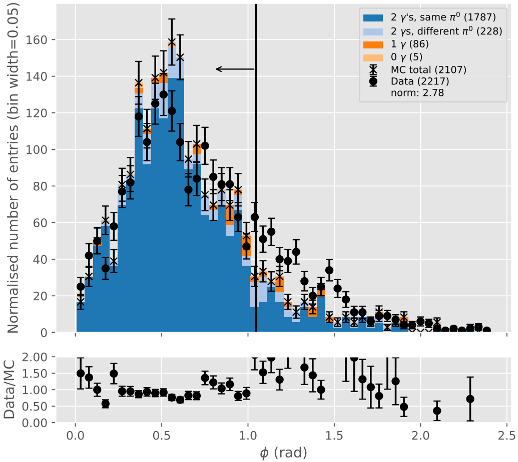
$$m_{\gamma\gamma} = \sqrt{2E_1E_2(1 - \cos(\phi))} \qquad (13)$$

- opening angle between showers is ϕ.
- E_1 and E_2 are **corrected** shower energies.
- Shower energy correction is similar technique to Kang's (see backups)
- cut on $m_{\gamma\gamma}$ to exclude misidentified shower pairs.



Opening angle

- Also cut on opening angle, but this has less impact than the mass cut.



Performance

	Counts	Purity (%)	Efficiency (%)
Beam particle selection	804	4.1	9.5
Number of photons = 2	788	62.2	9.3
$\boldsymbol{m_{\gamma\gamma} < 225}$ **MeV**	642	84.8	7.5
$\boldsymbol{\phi < 60}$ **deg**	581	87.4	6.8

$$\text{Purity} = \frac{\text{number of events with 2 photon showers which originate from a beam } \pi^0}{\text{total number of events selected}}, \tag{14}$$

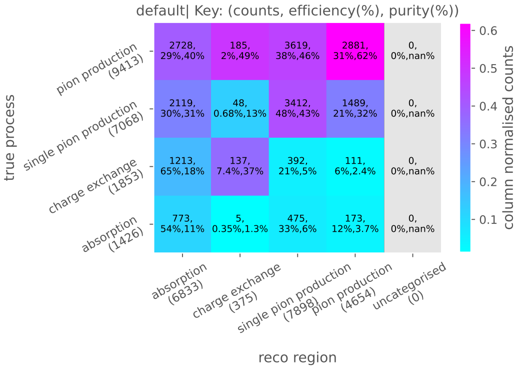
$$\text{Efficiency} = \frac{\text{number of events with 2 photon showers which originate from a beam } \pi^0}{\text{number of } \pi^0 \text{ after beam particle selection}}. \tag{15}$$

- A lot of events with a π^0 do not have two reconstructed shower pairs, so total efficiency is fairly low.
- Relative efficiency and purity is fairly good.
- Various misreconstruction effects (shower fragmenting, shower merging etc.) results in the low number of events with 2 π^0 showers.
- reconstruction method can only reconstruct 1 π^0 per event

Region selection

Old region selection

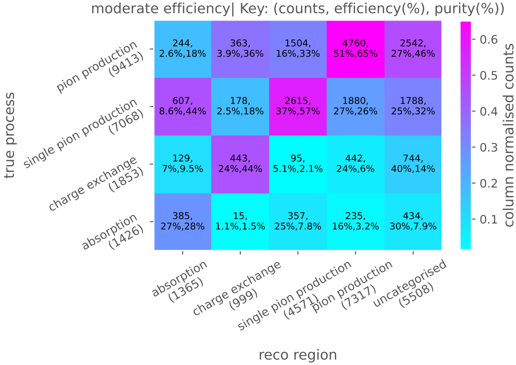
	$\pi^{\pm}$	π^{0}
abs	0	0
cex	0	1
spip	1	0
pip	>1	≥ 0
	>0	1



- For reference, the old selection.

New region selection

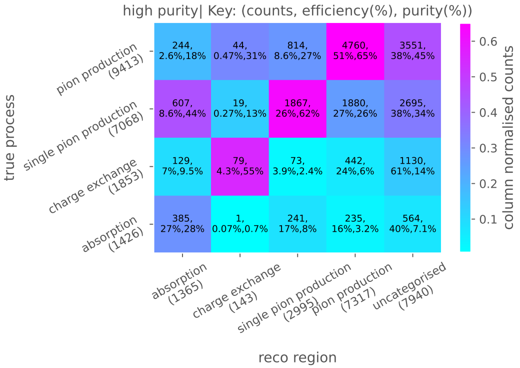
	loose $\pi^{\pm}$	loose γ	$\pi^{\pm}$	γ	π^{0}
abs	0	0	0	0	-
cex	0	0	0	2	1
	0	0	0	1	-
spip	-	0	1	0	-
pip	-	-	>1	-	-
	-	-	>0	2	1
	-	-	-	>2	-
	-	>2	-	-	-



- New selection defines regions based on all five particle types
- Loose particles used to veto, others used for positive ID
- In events where positive ID is difficult e.g. an event with just 1 loose γ, leave this event as uncategorised
- Overall region purity and efficiency improves in comparison to the old selection
- Abs region has the most strict selection possible, but still is background dominated, so not much more can be done without better particle ID
- Cex and spip regions could have more strict definitions

Higher purity?

	loose $\pi^{\pm}$	loose γ	$\pi^{\pm}$	γ	π^{0}
abs	0	0	0	0	-
cex	0	0	0	2	1
spip	0	0	1	0	-
pip	-	-	>1	-	-
	-	-	>0	2	1
	-	-	-	>2	-
	-	>2	-	-	-



	high purity	moderate efficiency	old
abs	0.08	0.08	0.06
cex	0.02	0.11	0.03
spip	0.16	0.21	0.21
pip	0.33	0.33	0.19

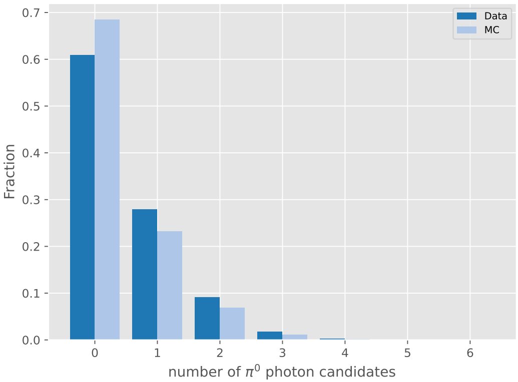
- More strict regions results in a much better cex purity, but at a massive cost in efficiency
- Comparing the purity $\times$ efficiency of selections, moderate efficiency is the best.
- High purity selection has too large of an efficiency cost for the purity increase to justify this selection

Summary

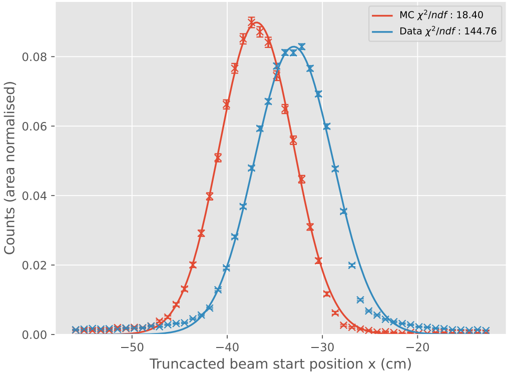
- New beam particle selection adopts fiducial region in reco and truth.
- New PFO selection and region selection, uses more particle types to define an exclusive process.
- It is possible to use a strict region definition, but at a large cost in efficiency $\rightarrow$ not great as MC needs to be used as a template for fit and unfolding.
- High purity selection may be justifiable if MC stat was much higher, but for now, I stick with the higher efficiency region definitions.
- Outstanding tasks:
  1. Second look at pull study (done)
  2. Measurement with Data (in progress)
  3. Systematics (in progress)

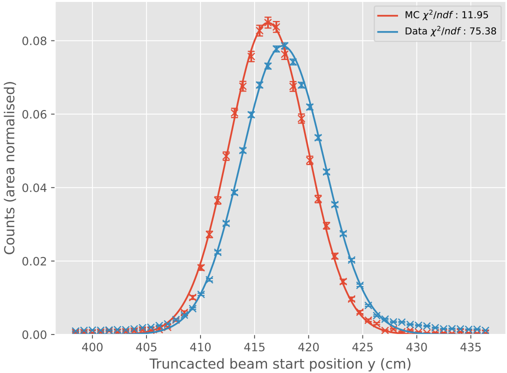
Backup

number of selected photons in an event

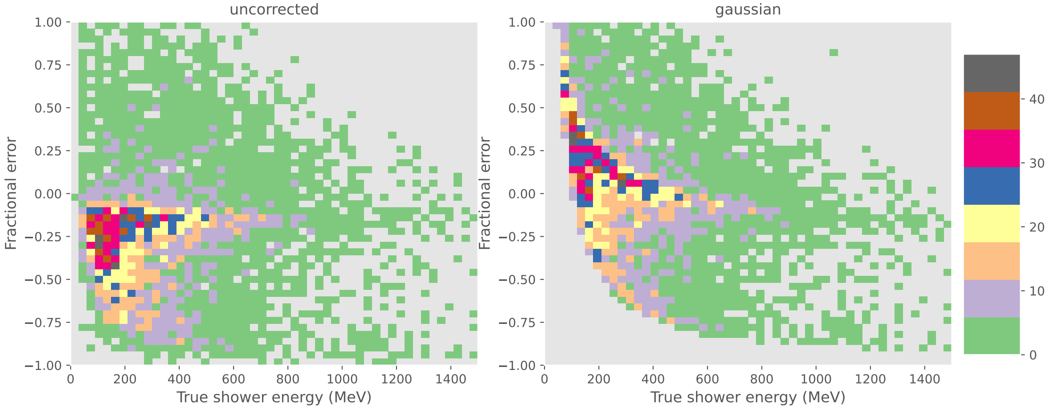


beam quality fits





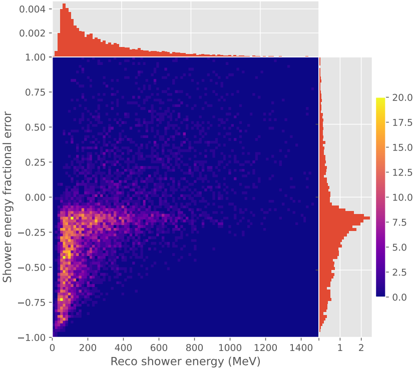
Shower energy correction

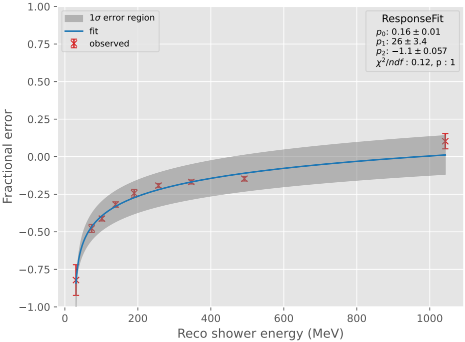


	μ	σ
uncorrected	-0.232	0.372
corrected	0.026	0.388

- overall improvement in the bias
- spread of shower energies increases slightly

Shower energy correction





- reconstructed shower energies are not great, many showers have underestimated enegies, and a $\sim$ 20% bias in the shower energies are observed.
- in bins of E_{reco} estimate the most probable shower energy fractional error, then fit a response function to these (plot on the right)

$$C(E_{reco}) = p_0 \ln(E_{reco} - p_1) + p_2. \tag{16}$$

- correction defined as

$$E^{c}_{reco} = \frac{E_{reco}}{1 + C(E_{reco})}. \tag{17}$$