First look at the π^0 event reconstruction in ProtoDUNE-I 2 GeV data

Vikas Gupta

Hadron Analysis meeting

10-May-2023

Content

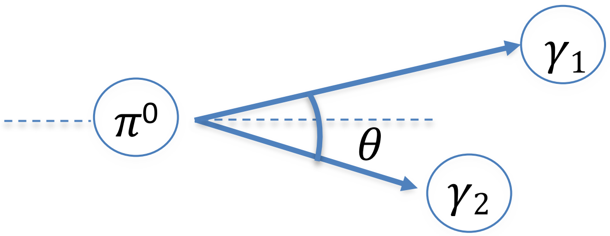
- π^0-decay analysis overview
- π^0 shower selection
- Shower energy correction
- π^0 event selection and $m_{\gamma\gamma}$
- π^0-decay reconstruction at event level
- Extending analysis to lower nHits data
- Summary and Future plans

π^0-decay analysis overview

π^0 decay is an useful candle for validating shower reconstruction in (Proto)DUNE

- π^0 are produced commonly from hadronic interactions and decay promptly to two photons (EM showers).
- The showers are boosted significantly in the lab frame:

  $\theta << 180$, and $E_1 \neq E_2$.



- Motivation for analysing π^0 events in ProtoDUNE:
  1. Identifying π^0 showers and reconstructing the invariant mass is a good test for validating shower energy and direction reconstruction.
  2. Discriminating electron showers from a photon shower (of π^0 decay) is also an important background challenge for DUNE.

$$m_{\gamma\gamma}^2 = 2E_1E_2(1 - \cos\theta)$$

Dataset used in this analysis

- I have used analysis ntuples created by **Jake Calcutt** for the **2 GeV** dataset:
  - Data: xroot://fndca1.fnal.gov:1094/pnfs/fnal.gov/usr/dune/tape_backed/dunepro/protodune-sp/root-tuple/2022/detector/physics/PDSPProd4/00/00/54/29/PDSPProd4_data_2GeV_reco2_ntuple_v09_42_03_01.root
  - MC: xroot://fndca1.fnal.gov:1094/pnfs/fnal.gov/usr/dune/tape_backed/dunepro/protodune-sp/root-tuple/2022/mc/physics/PDSPProd4a/20/91/32/85/PDSPProd4a_MC_2GeV_reco1_sce_datadriven_v1_ntuple_v09_41_00_03.root

- Ntuple variables used in this analysis:
  - Beam tag: `reco_beam_type`
  - Beam particle (for MC): `true_beam_PDG`
  - Pandora tag: `reco_daughter_pandora_type`
  - CNN score : `reco_daughter_PFP_emScore_collection`
  - nHits : `reco_daughter_PFP_nHits_collection`
  - E_{shower} , Dir_{shower} : `reco_daughter_allShower_energy`, `reco_daughter_allShower_dirX(/Y/Z)`

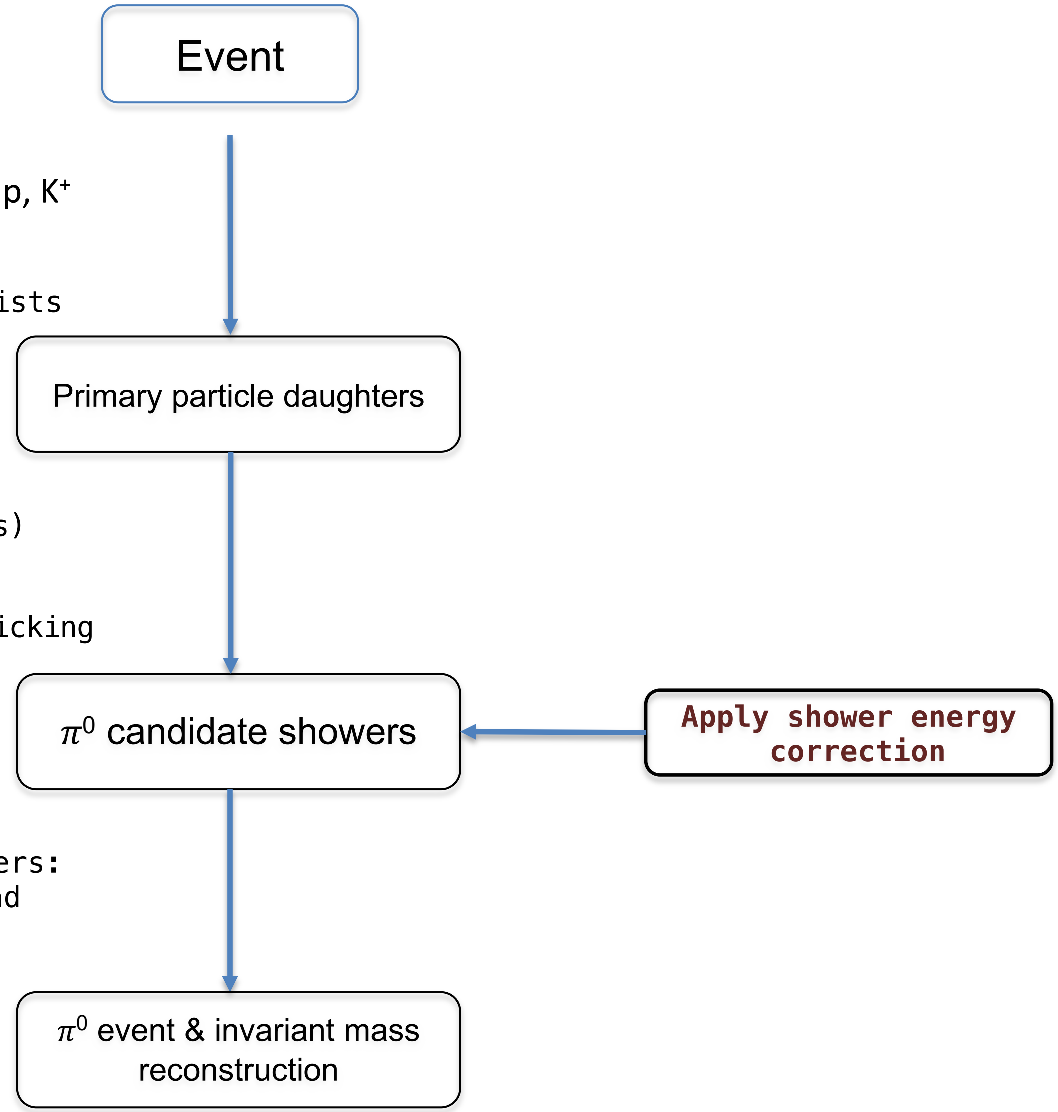
Complete analysis flowchart for selecting π^0 events

1.1 Beam selection
- Beam tag : track
- Beam particle (**for MC**): π^+, p, K^+

1.2 Sanity check
- E_{shower} >0
- *CNN score* exists, *nHits* exists

2. π^0 Shower selection
- Pandora tag : 11 (EM showers)
- 0.6 > *CNN score* > 0.95
- nHits > 80
- Two showers in the event (picking only two if more)

3. π^0 event selection
- opening angle between showers:
  0.1 rad < ($\theta_{\pi 0}$) < 0.7 rad
- π^0 energy:
  ($E_{\pi 0}$) < 1000 MeV/c²

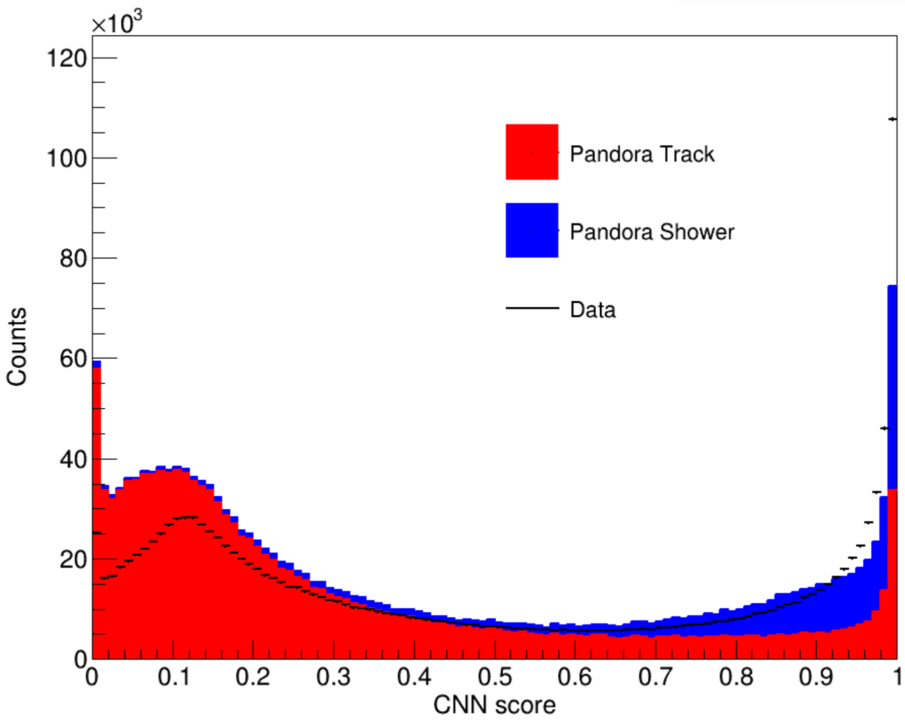


π^0 shower selection

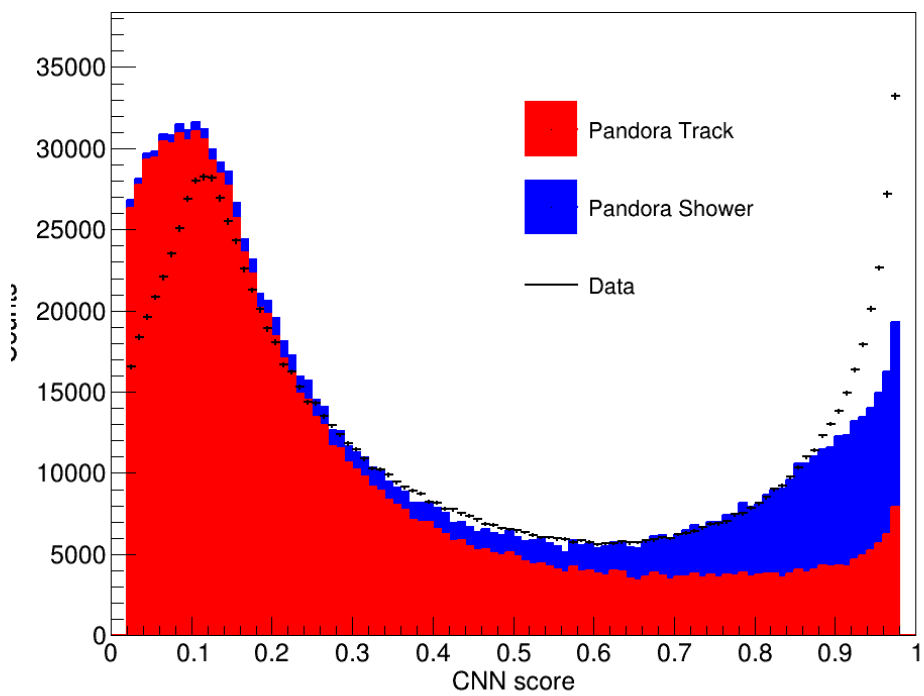
CNN score for all primary particle daughters

- Some disagreement for daughter particle (**shower**) CNN score between MC and data.
  - weighted CNN score only affects distribution at < ~0.25 (doesn't affect analysis using **showers**)

By Pandora tag



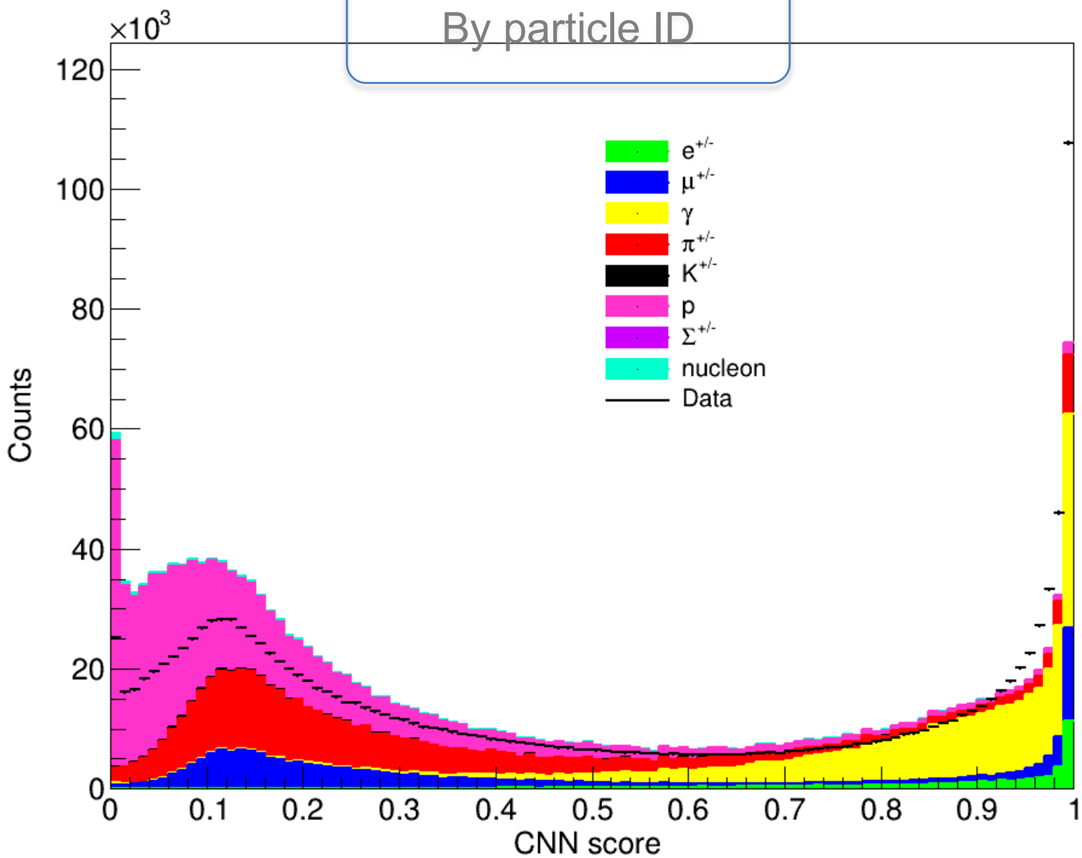
0.02< CNN score <0.98



CNN score for all primary particle daughters

- Some disagreement for daughter particle (**shower**) CNN score between MC and data.
  - weighted CNN score only affects distribution at < ~0.25 (doesn't affect analysis using **showers**)

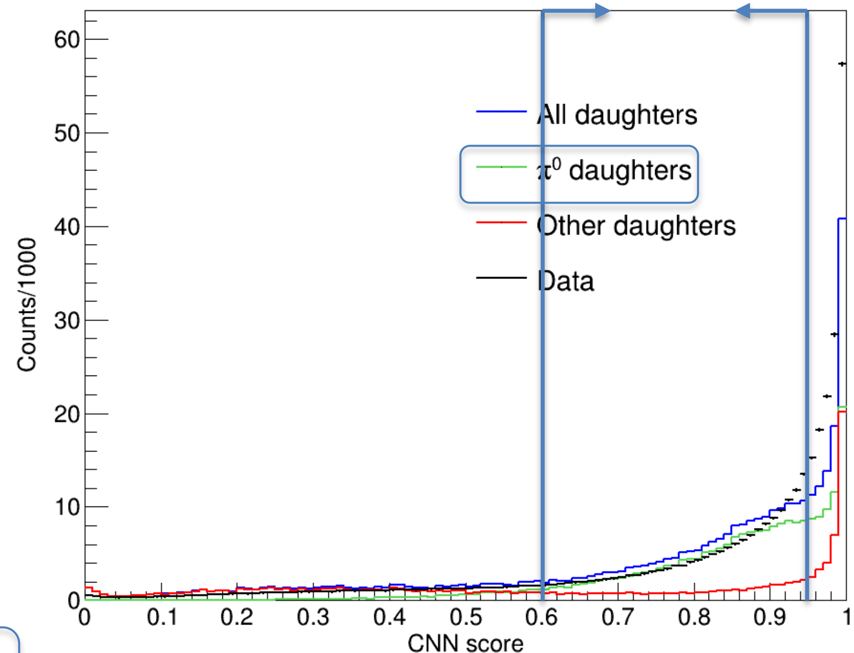
By particle ID



CNN score for all primary particle daughter tagged by Pandora as showers (primary daughter showers)

- The primary daughter showers are mostly photons coming from π^0 decay:
  - Taking data above CNN score of 0.6 will give mostly π^0 daughters.

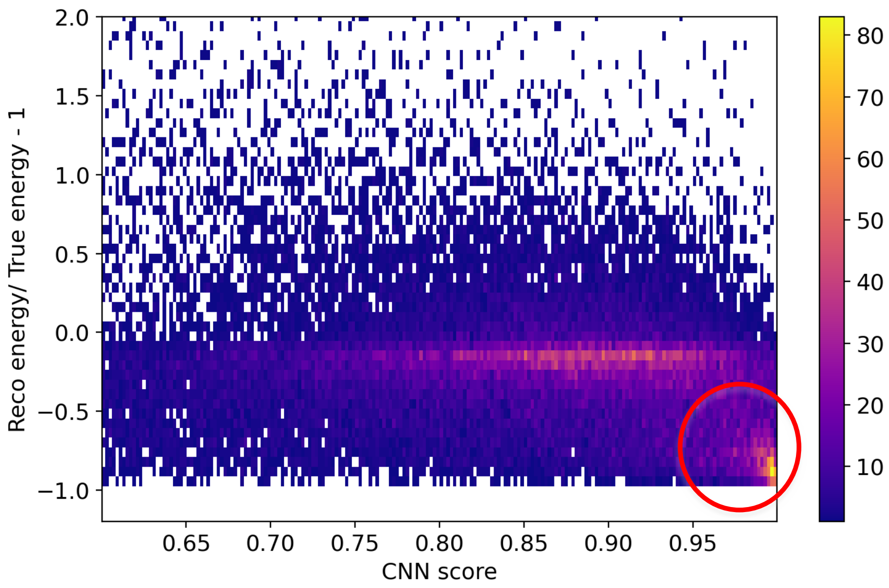
Pandora showers only



CNN score: 0.6–0.95

CNN score for all primary daughter showers

- The primary daughter showers are mostly photons coming from π^0 decay:
  - Taking data above CNN score of 0.6 will give mostly π^0 daughters.
  - above CNN score of 0.95, we have a lot of broken showers (low completeness) that should be removed.
    - Alternatively, they are also removed with nHits>80 cut.



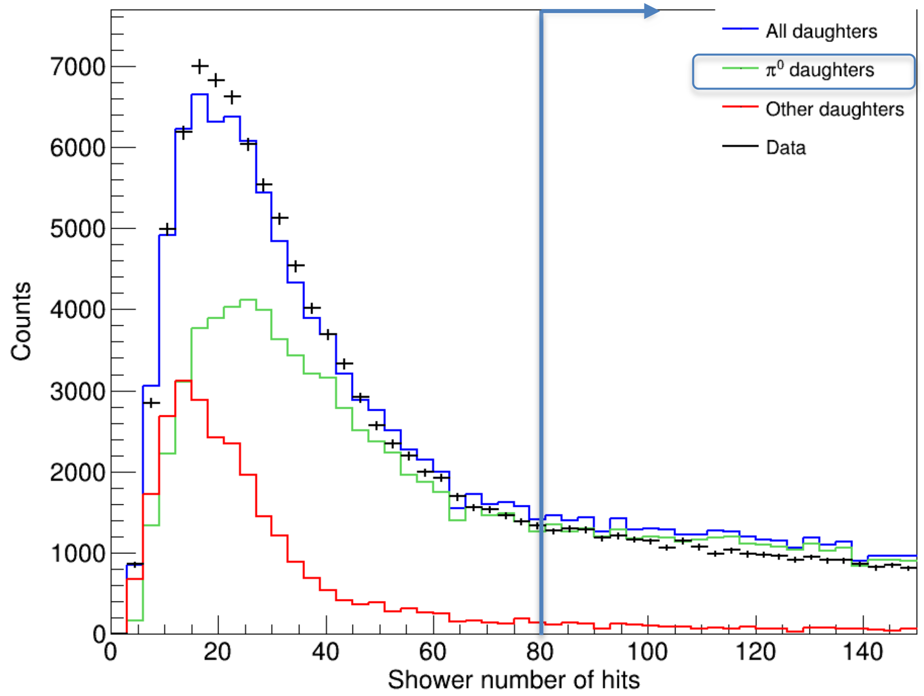
nHits >20 applied for better visibility

A higher CNN score for low completeness showers should point towards some improvement needed in the reconstruction.

CNN score: 0.6–0.95

Number of hits for all primary daughter showers

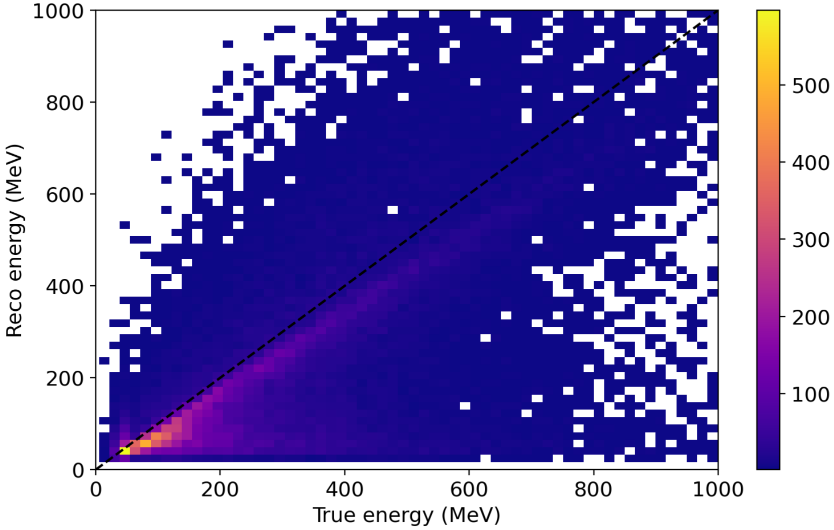
- The primary daughter showers are mostly photons coming from π^0 decay:
  - Ideally, we would like to consider data above 20 nHits to maximize the efficiency.



nHits > 80

Number of hits for all primary daughter showers

- The primary daughter showers are mostly photons coming from π^0 decay:
  - Ideally, we would like to consider data above 20 nHits to maximize the efficiency.
  - Stricter cuts on are needed due to low completeness of showers at low nHits:
    - we get artificially high statistics at low energy due to these broken showers.

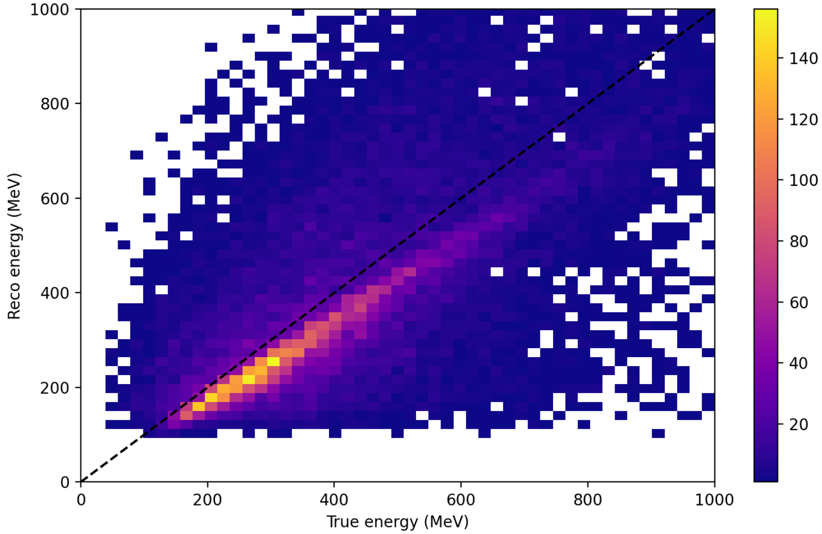


nHits >20

nHits > 80

Number of hits for all primary daughter showers

- The primary daughter showers are mostly photons coming from π^0 decay:
  - Ideally, we would like to consider data above 20 nHits to maximize the efficiency.
  - Stricter cuts on are needed due to low completeness of showers at low nHits:
    - we get artificially high statistics at low energy due to these broken showers.
    - A lot of outliers are removed for nHits > 80. The showers energy looks underestimated by 10%-20%



nHits >80

nHits > 80

Shower selection in numbers

$$\text{Purity} = \frac{N_{\pi0\_daughters}}{N_{sample}}$$

$$\text{Efficiency} = \frac{N_{\pi0\_daughters}}{N_{\pi0\_daughters}\ available}$$

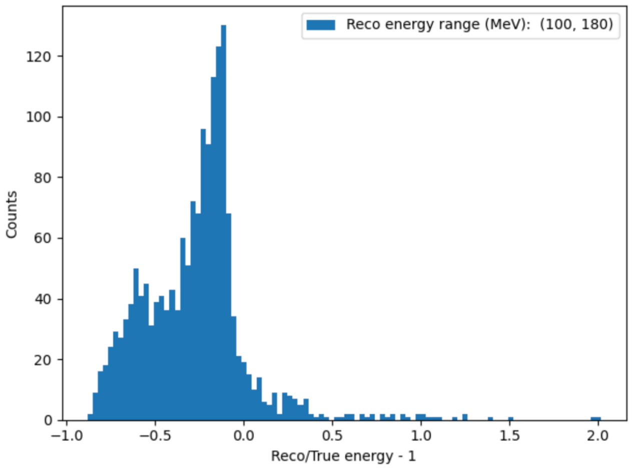
1. Selecting all π^0 showers in events

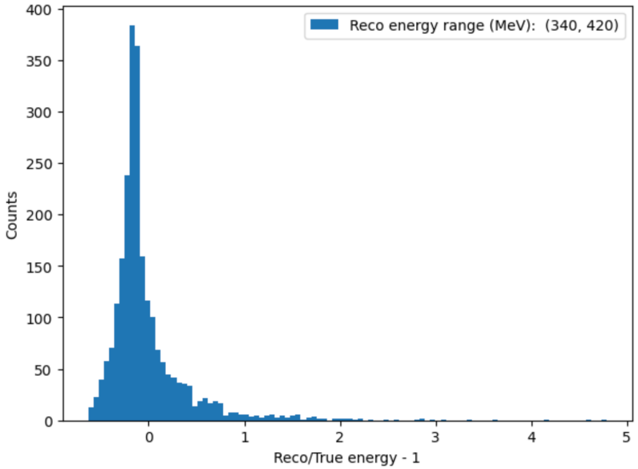
Number of PFParticle daughters in dataset	Data	MC	Ratio (Data/MC)	Purity (Cumulative)	Efficiency (Cumulative)
All daughters	1580360	372250	4.24		
Sanity check	1384840	332605	4.16	0.201	1
Pandora showers	361952	71555	5.06	0.638	0.682
0.6 < CNN score < 0.95	166477	37554	4.43	0.815	0.457
nHits >80	76369	17372	4.39	0.934	0.242

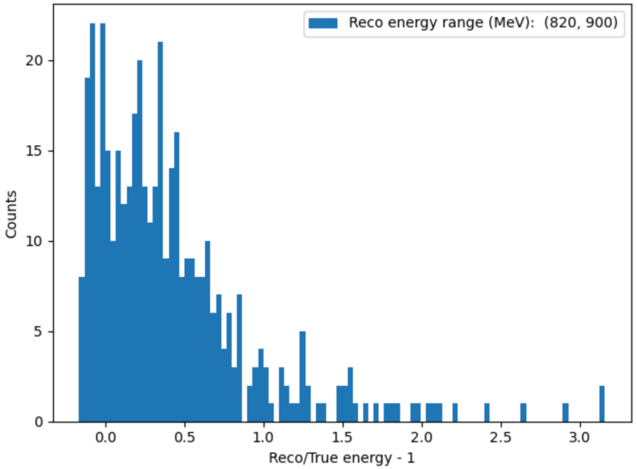
Shower energy correction

Estimating shower energy correction function

- The correction factor can be calculated as a function of reconstructed energy:
  - Find most likely fraction bias (E_{reco}/E_{true} –1) in bins of E_{reco} .

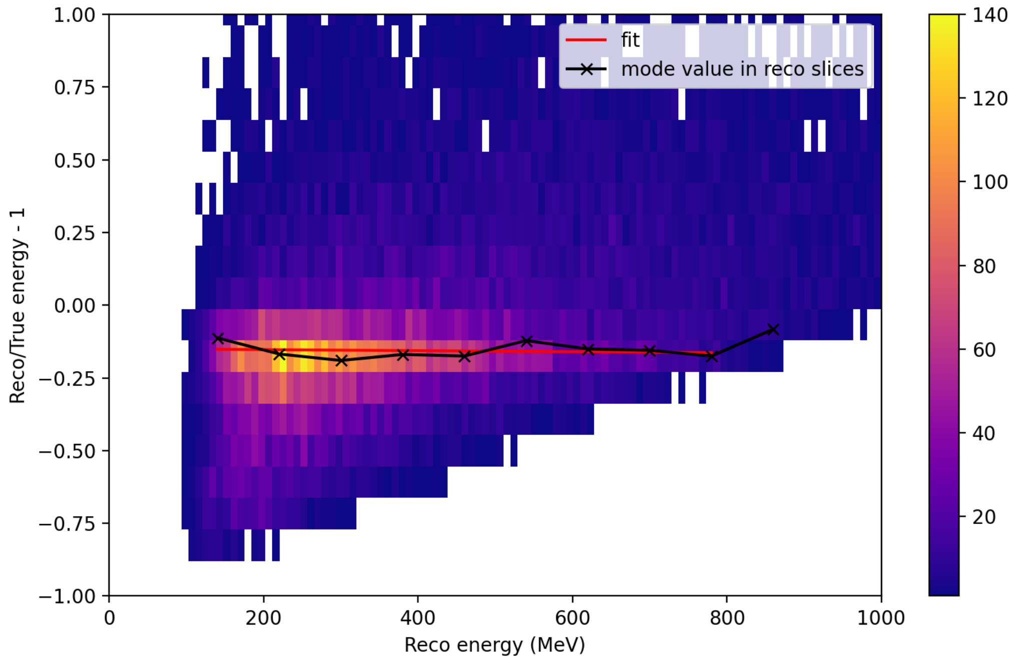






Estimating shower energy correction function

- The correction factor can be calculated as a function of reconstructed energy:
  - Find most likely fraction bias (E_{reco}/E_{true} –1) in bins of E_{reco} .
  - Fit the most likely value per bin to get correction factor as function of E_{reco} .

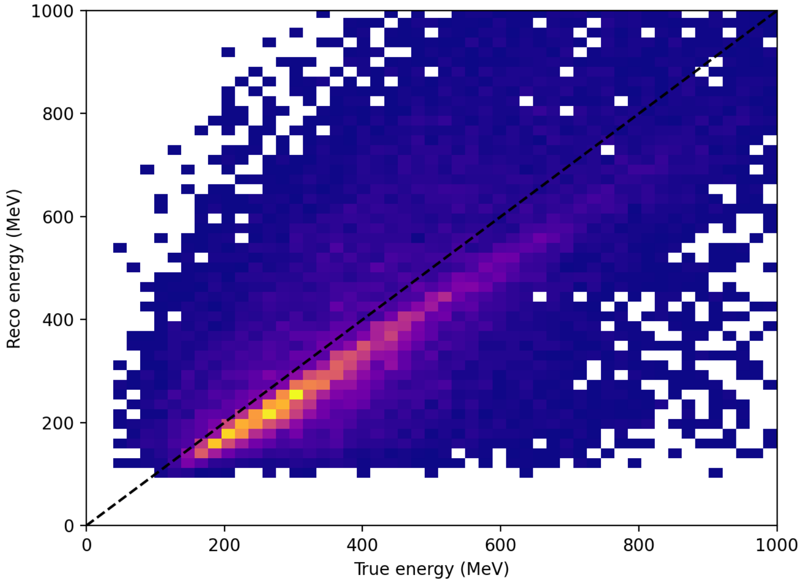


Linear fit with
m = -1.76 e-05 ± 4.38 e-05
c = - 0.149 ± 0.022

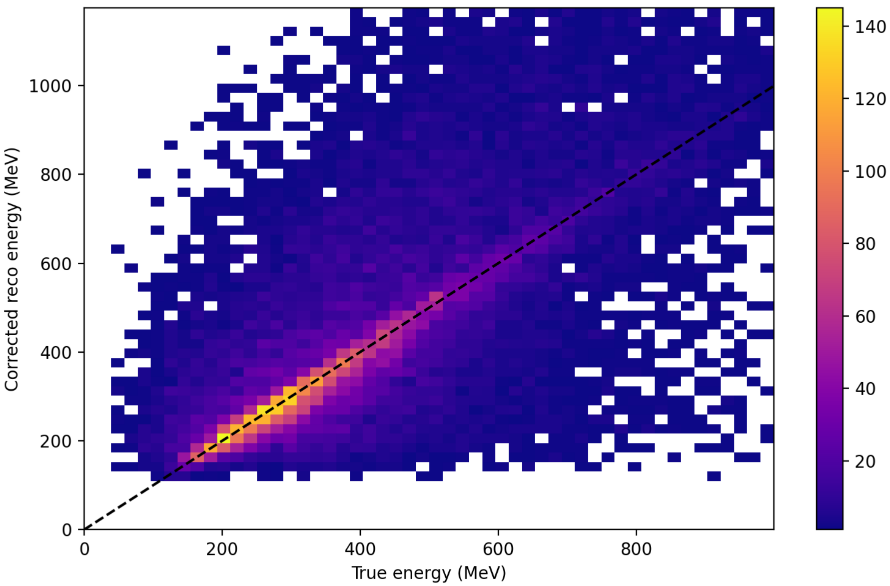
$E_{corr} \approx E_{reco} / 0.851$

Energy reconstruction bias is greatly reduced after applying energy correction

- The correction method does a good job and after correction, the bias is reduced greatly.
  - For nHits >80 data, an energy independent correction factor is adequate.



Before correction



After correction

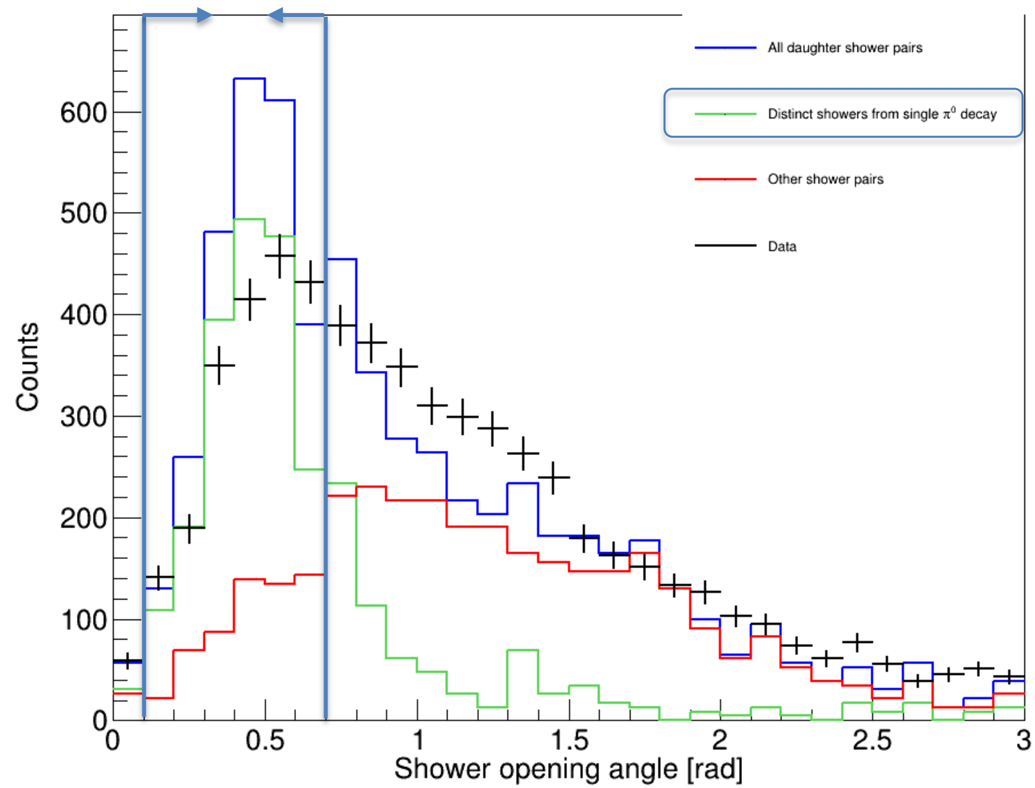
π^0 event selection and $m_{\gamma\gamma}$

Shower pair selection for π^0 reconstruction

- Most events have either one or two showers that pass initial cuts:
  - Choose the two most energetic showers in the event, if more than two showers.
- Assume the two showers come from a π^0-decay and reconstruct:
  - Opening angle from shower directions
  - Energy of π^0 (= $E_{shower1} + E_{shower2}$)
- Calculate $m_{\gamma\gamma}$ for shower pair that pass cuts on angle and π^0 energy.

Opening angle of shower pairs

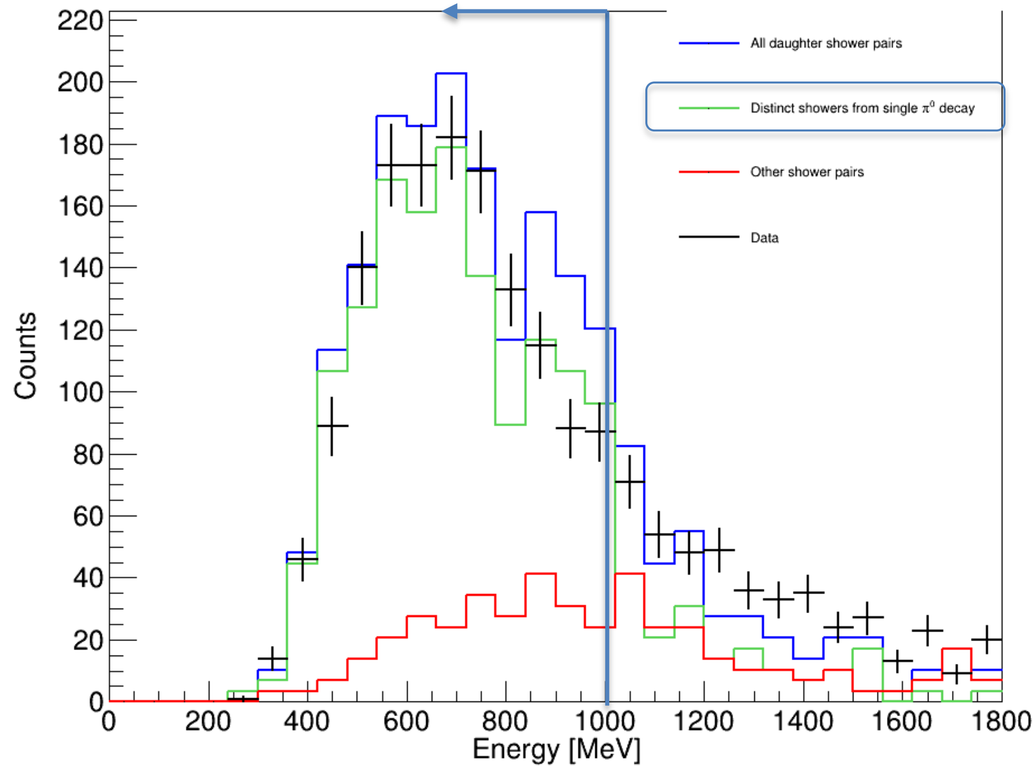
- π^0 showers peak at low opening angle and above 0.7 radian, the total background starts to dominate
- Below 0.1 rad, also low signal/bkg.



$0.1 \text{ rad} < \theta_{\pi 0} < 0.7 \text{ rad}$

π^0 energy

- π^0 energy can be estimated by summing corrected energy for both showers.
- Removing data above 1000 MeV as signal/background is quite low at higher energies.

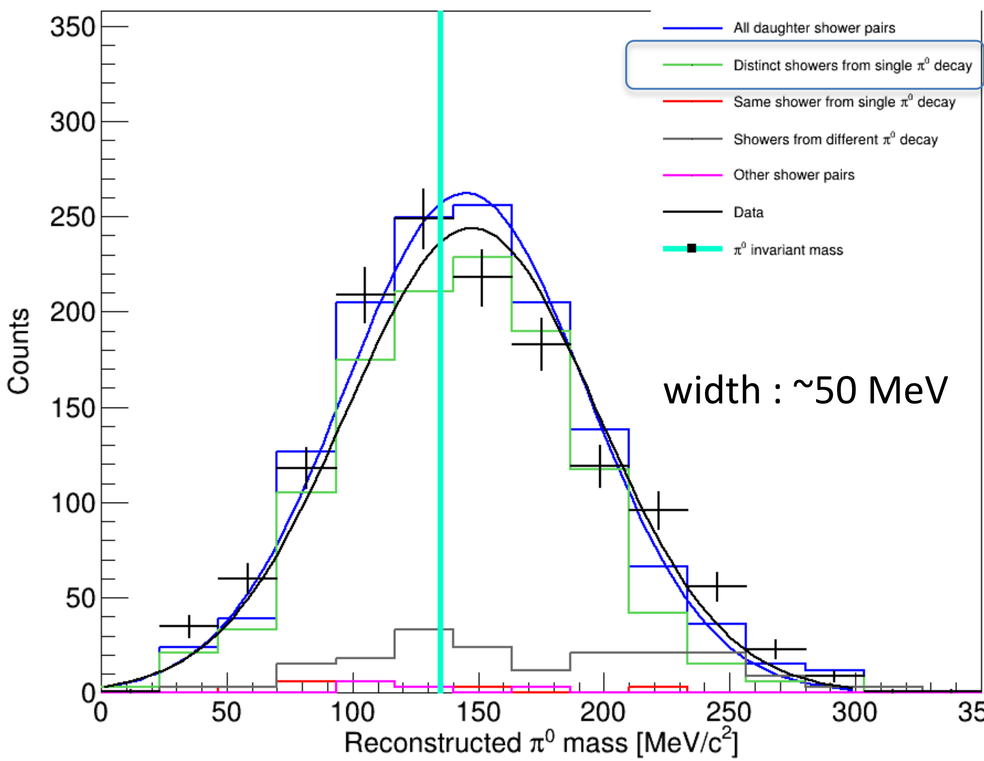


$E_{\pi 0}$: < 1000 MeV

Reconstructed π^0 mass

- Fit results for π^0 reconstructed mass disagrees slightly with the true invariant mass of 134.97 MeV/c^2 for both data and MC.

- Statistics are quite low due to
  - Extremely low efficiency of finding both π^0 showers in an event.
  - nHits > 80 cut required for good shower energy reconstruction



Gaussian fit results ($m_{\gamma\gamma}$):
MC: 145.024 ± 2.35 MeV/c^2 (signal only: 142.1 ± 2.32 MeV/c^2)
Data: 147.622 ± 1.68 MeV/c^2

π^0 event selection in numbers

$$\text{Purity} = \frac{N_{\pi0\_\mathbf{unique}\_daughter\_pairs}}{N_{sample}}$$

$$\text{Efficiency} = \frac{N_{\pi0\_\mathbf{unique}\_daughter\_pairs}}{N_{\pi0\_\mathbf{unique}\_daughter\_pairs}\ available}$$

2. Selecting **two** distinct π^0 showers in each event

Number of PFParticle daughters in dataset	Data	MC	Ratio (Data/MC)	Purity (Cumulative)	Efficiency (Cumulative)
Shower pair	11970	2764	4.33	0.449	1 (0.0185 of all showers)
0.1 rad < Opening angle < 0.7 rad	3970	1156	3.43	0.763	0.710
π^0 energy < 1000 MeV/c^2	2756	916 (768)	3	0.838	0.618

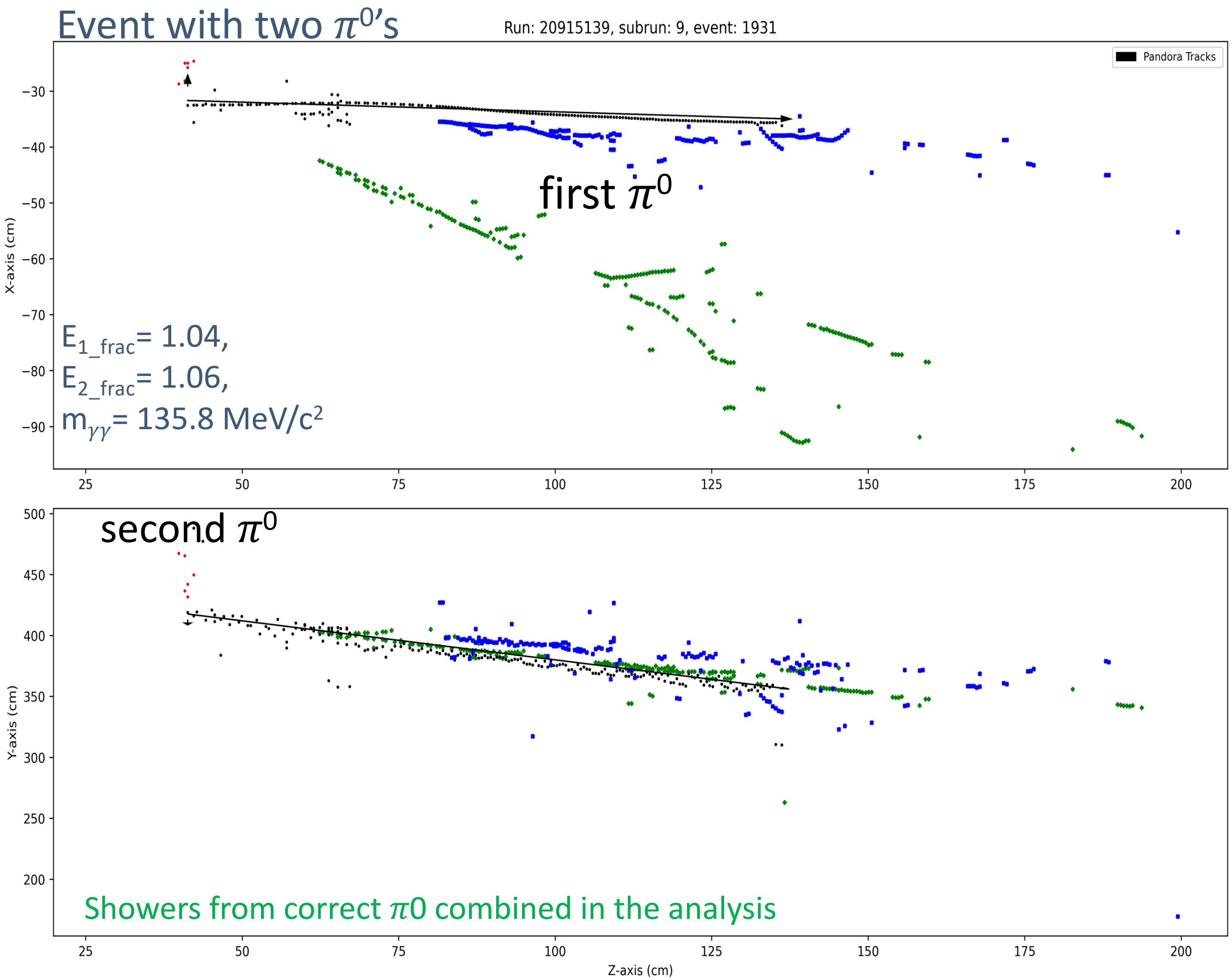
768 daughters : 384 π^0 events in MC

π^0-decay reconstruction at event level

How does the reconstructed event look? (only showing primary daughters of PFParticle)

Color scheme:

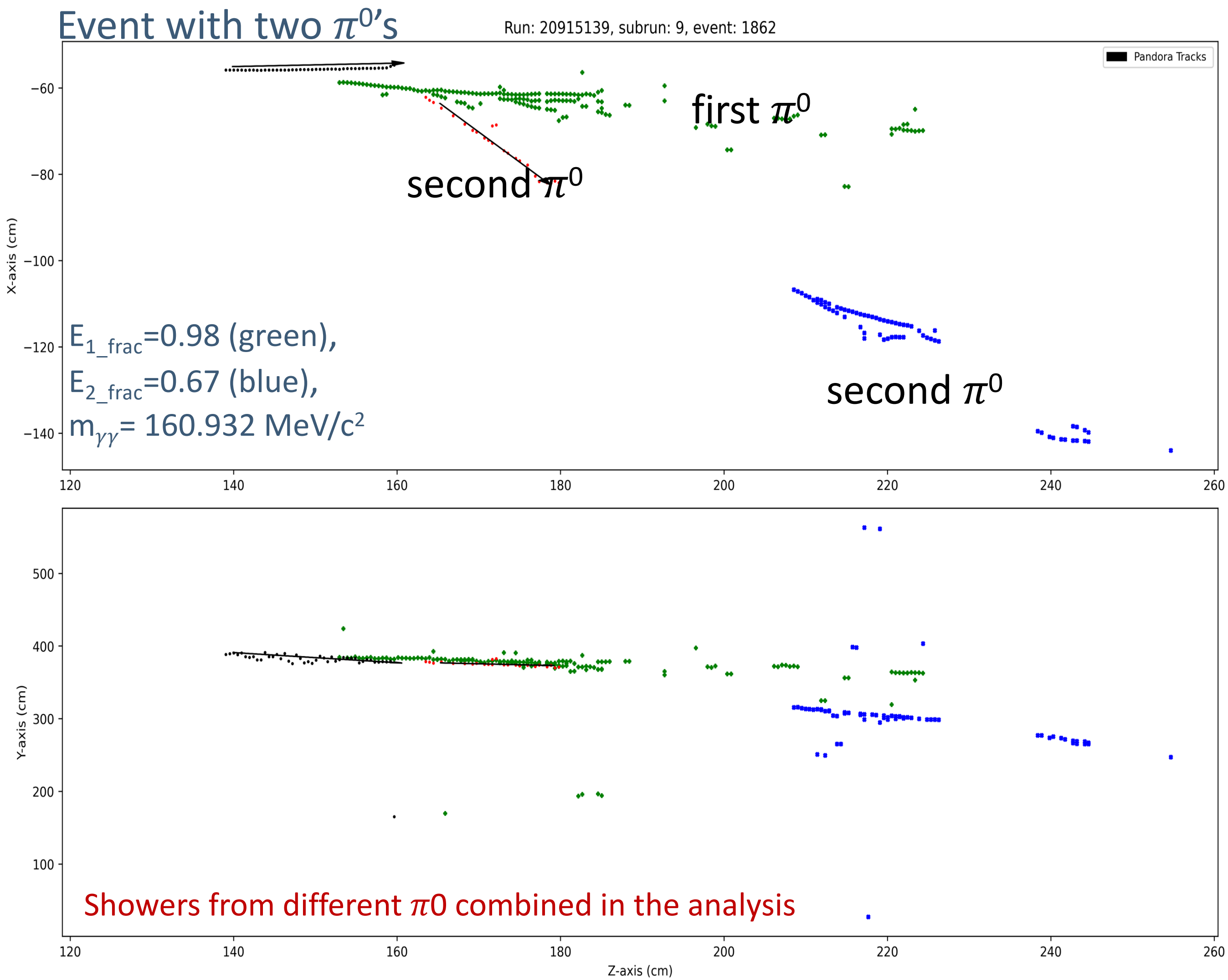
1. Hit cluster doesn't belong to reco object assigned to a π^0 daughter ->black.
2. reco object is tagged as track -> black arrow.
3. If a π^0 was produced in the event:
   - Shower hits are red
   - Shower hits are green and blue **if I did** reconstruct a π^0 in my analysis.



How does the reconstructed event look? (only showing primary daughters of PFParticle)

Color scheme:

1. Hit cluster doesn't belong to reco object assigned to a π^0 daughter ->black.
2. reco object is tagged as track -> black arrow.
3. If a π^0 was produced in the event:
   - Shower hits are red
   - Shower hits are green and blue **if I did** reconstruct a π^0 in my analysis.

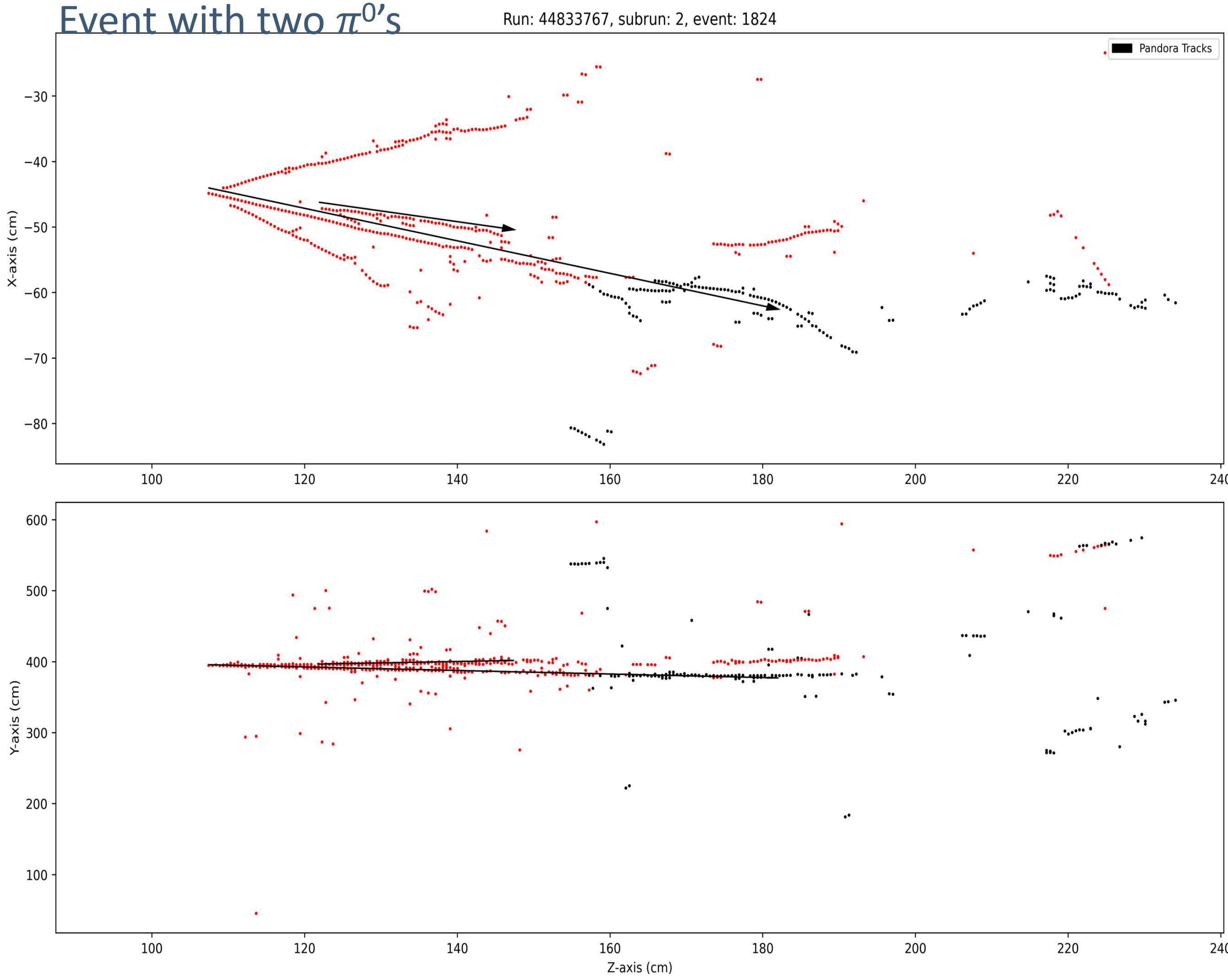


How does the reconstructed event look? (only showing primary daughters of PFParticle)

Color scheme:

1. Hit cluster doesn't belong to reco object assigned to a π^0 daughter ->black.
2. reco object is tagged as track -> black arrow.
3. If a π^0 was produced in the event:
   - Shower hits are red

Event with two π^0's

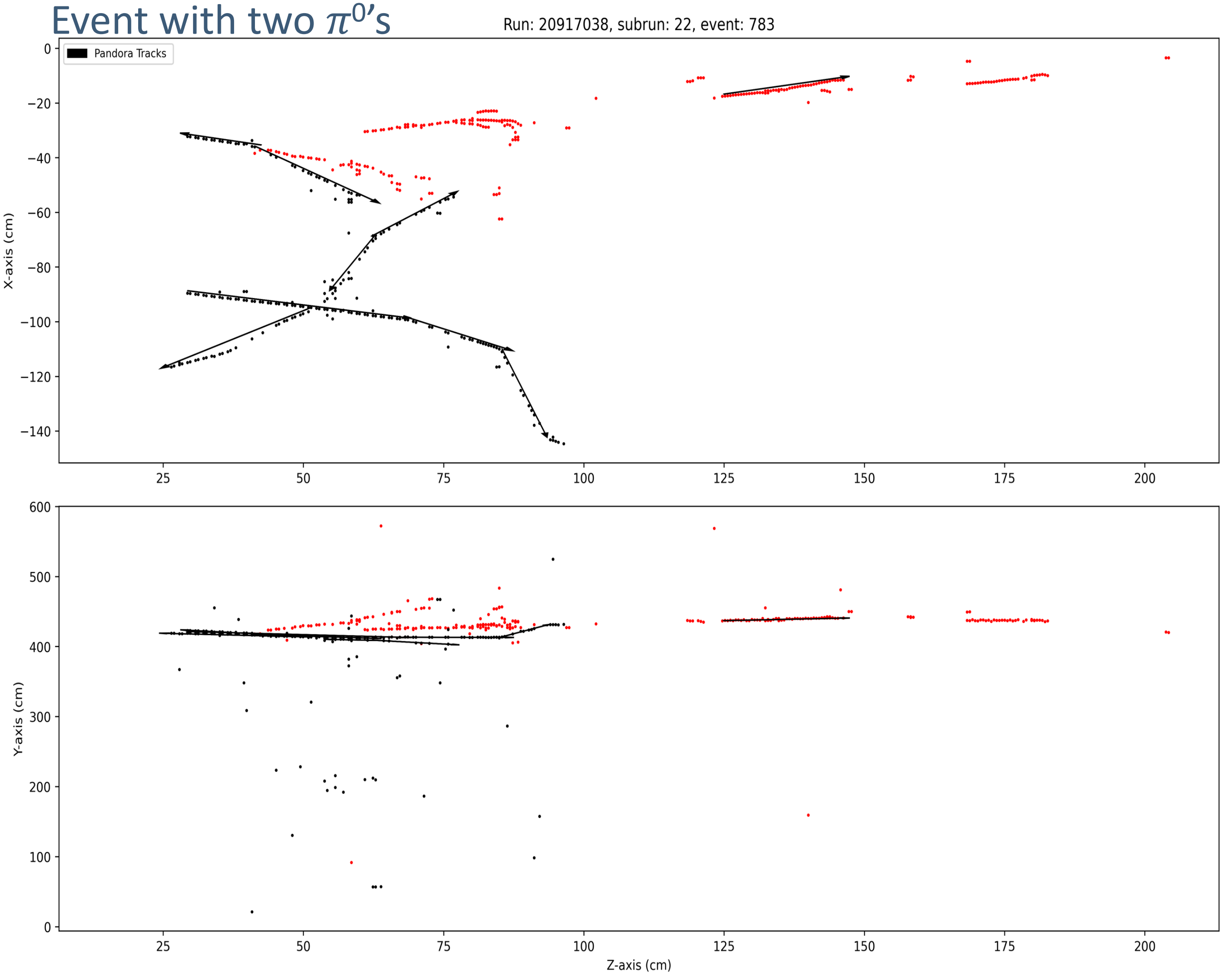


How does the reconstructed event look? (only showing primary daughters of PFParticle)

Color scheme:

1. Hit cluster doesn't belong to reco object assigned to a π^0 daughter ->black.
2. reco object is tagged as track -> black arrow.
3. If a π^0 was produced in the event:
   - Shower hits are red

Event with two π^0's

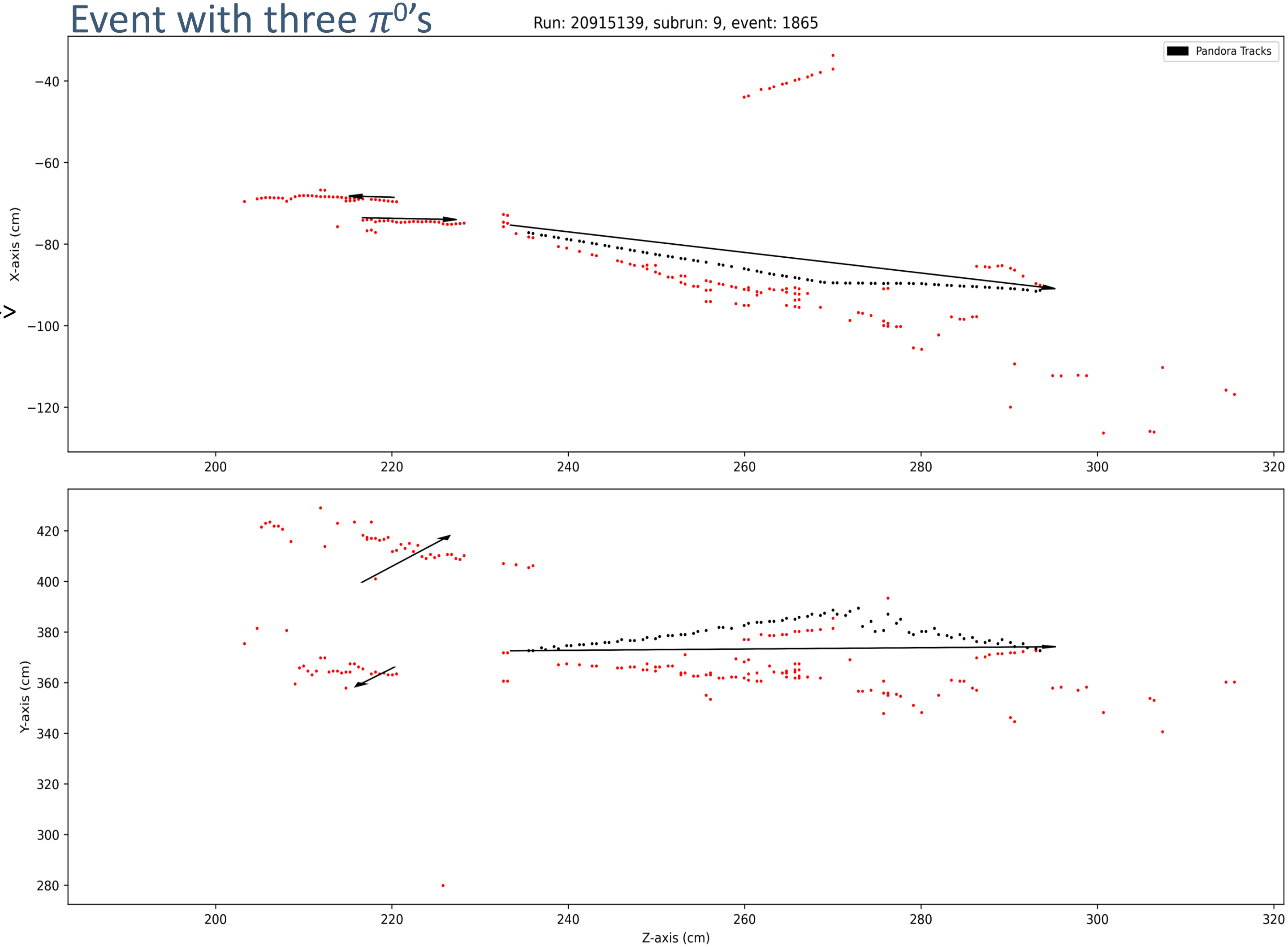


How does the reconstructed event look? (only showing primary daughters of PFParticle)

Color scheme:

1. Hit cluster doesn't belong to reco object assigned to a π^0 daughter ->black.
2. reco object is tagged as track -> black arrow.
3. If a π^0 was produced in the event:
   - Shower hits are red

Event with three π^0's



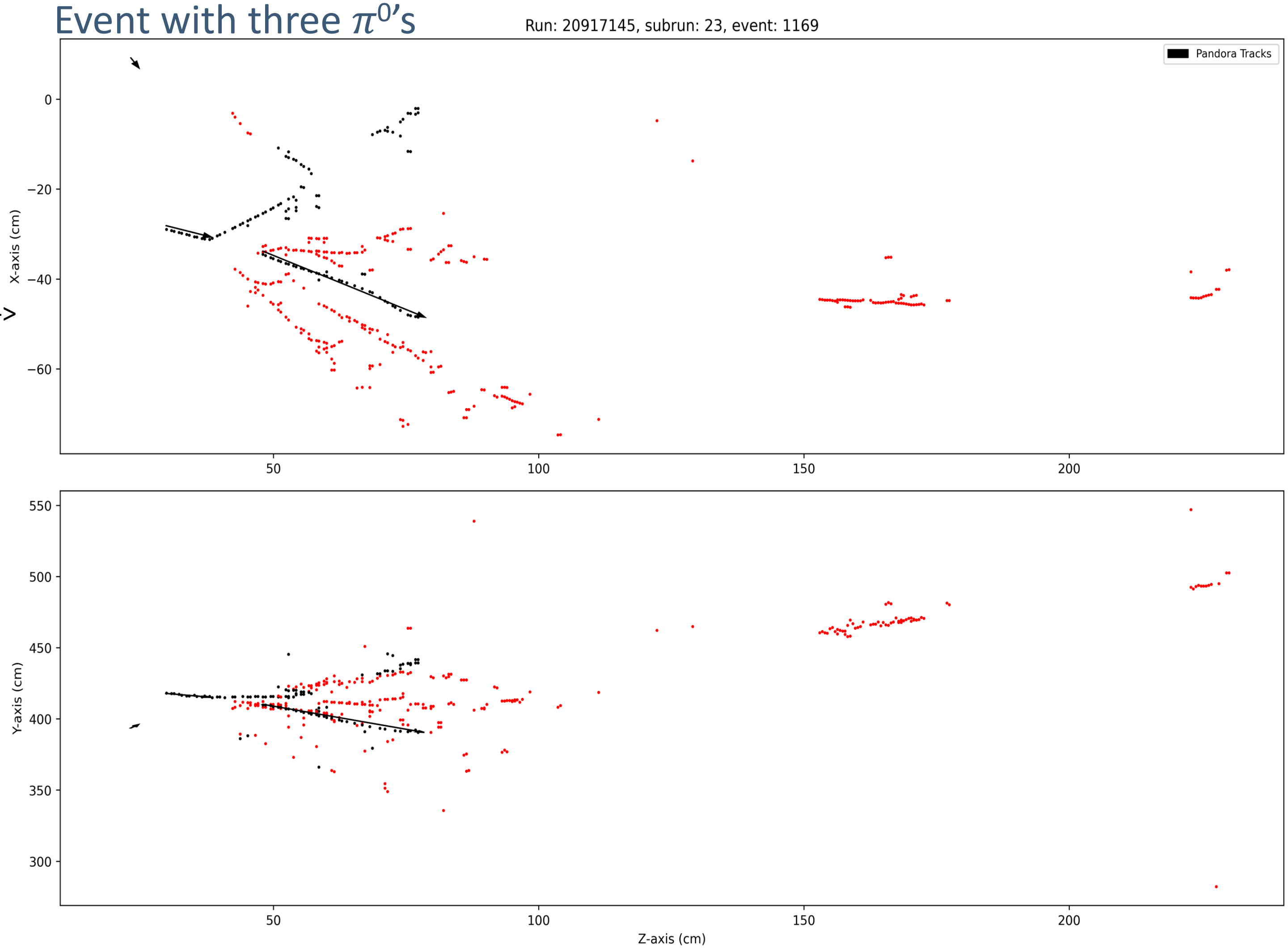
Nikhef DUNE

How does the reconstructed event look? (only showing primary daughters of PFParticle)

Color scheme:

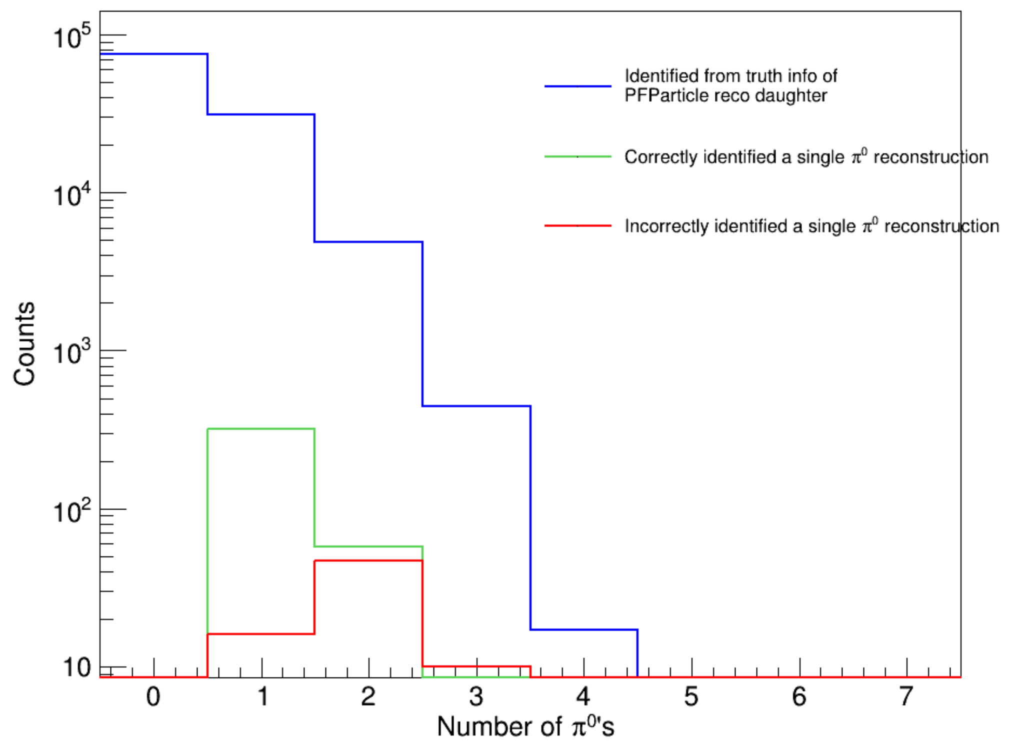
1. Hit cluster doesn't belong to reco object assigned to a π^0 daughter ->black.
2. reco object is tagged as track -> black arrow.
3. If a π^0 was produced in the event:
   - Shower hits are red

Event with three π^0's



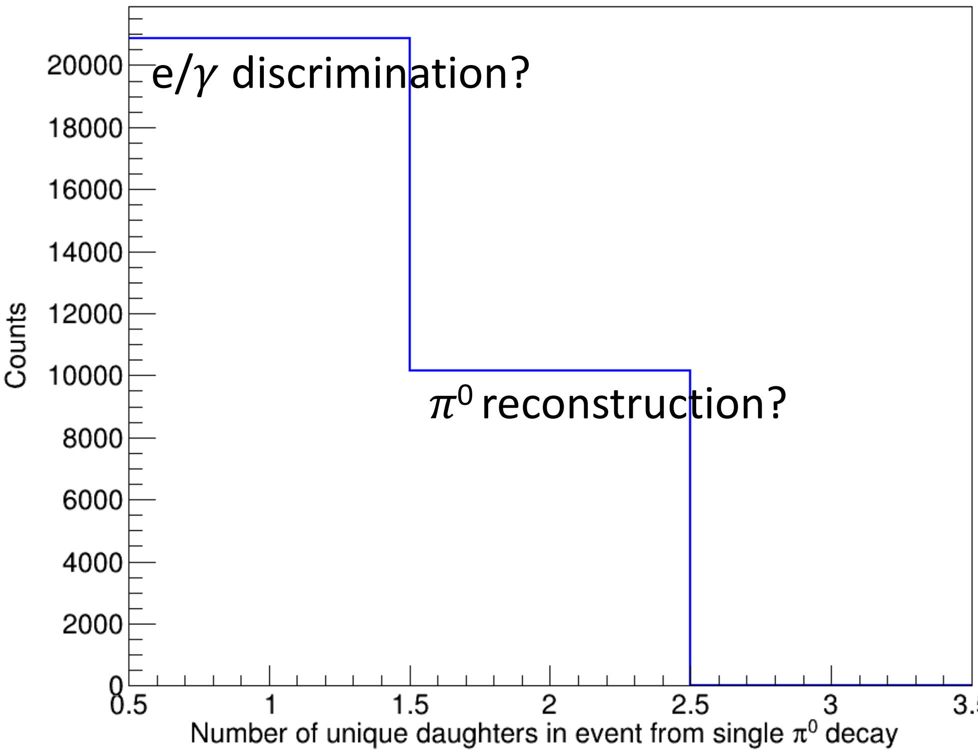
How many π^0's are contributing to PFParticle daughters?

- I am looking at data with only sanity check applied.
- We are only able to identify a very small number of π^0 decays that are produced in ProtoDUNE events.
- Note: Zero π^0's means that we see no reco object from its decay in PFParticle daughters (not that none were produced).



How often do we see both unique showers in events with only a single π^0 decay?

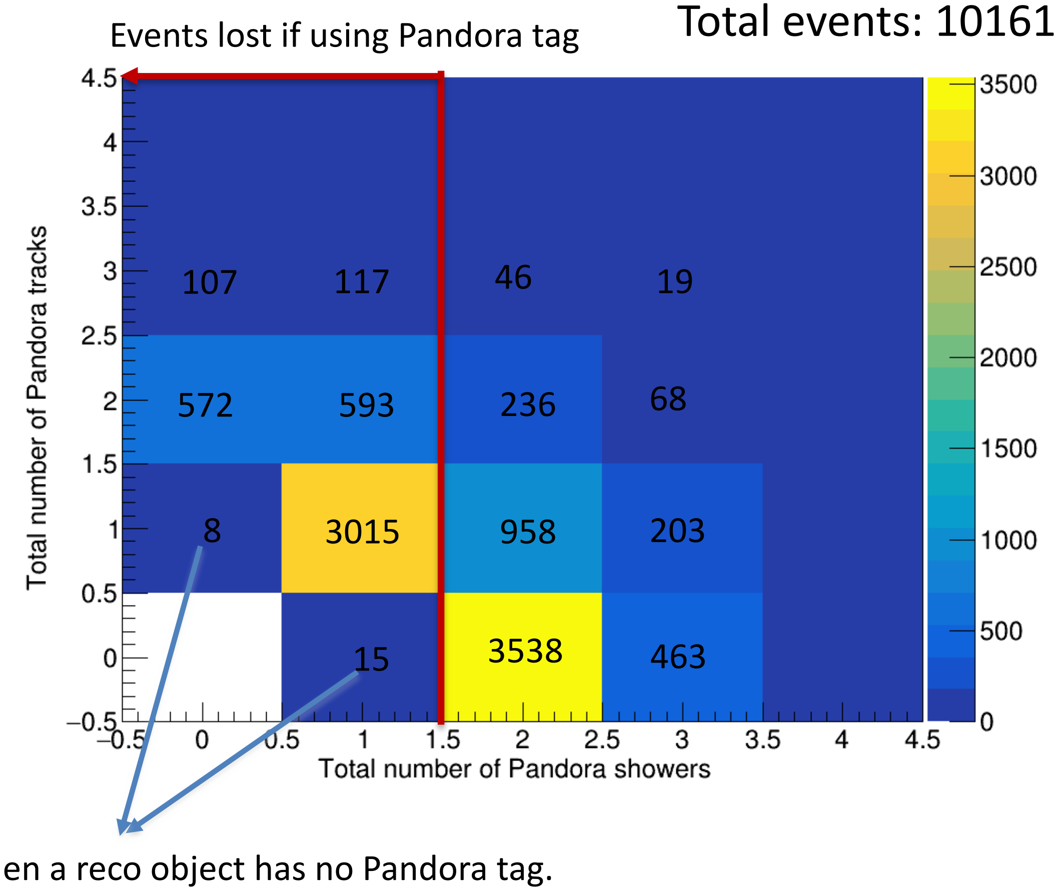
- For events with only a single π^0 decay contributing:
  - About 2/3rd decay do not have both photons in the PFParticle daughter objects.



How many reco object are created from a single π^0 daughter (with both daughter seen in the event)?

- For events with only a single π^0 decay contributing:
  - About 2/3rd decay do not have both photons in the PFParticle daughter objects.
  - Quite often the true π^0 daughters do not leave only a single Pandora shower each (Ideal scenario).
  - **How often is the energy well reconstructed in the Ideal scenario?**

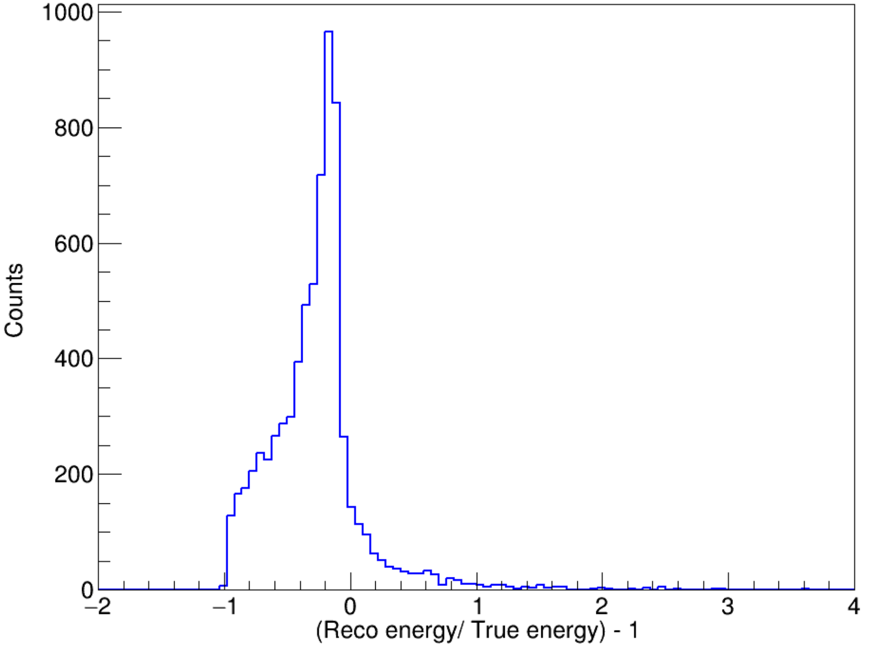
Total events: 10161



How often is the energy well reconstructed in the Ideal scenario?

- Using events that have one Pandora shower each for π^0 daughter. (3538 events)
- We still have a lot of low completeness showers:
  - Shower hits are getting absorbed by nearby reco objects?
  - Does the first photon absorb hits from the second one often as well?
- Will study these events further.

(no energy correction applied)



Extending analysis to lower nHits data

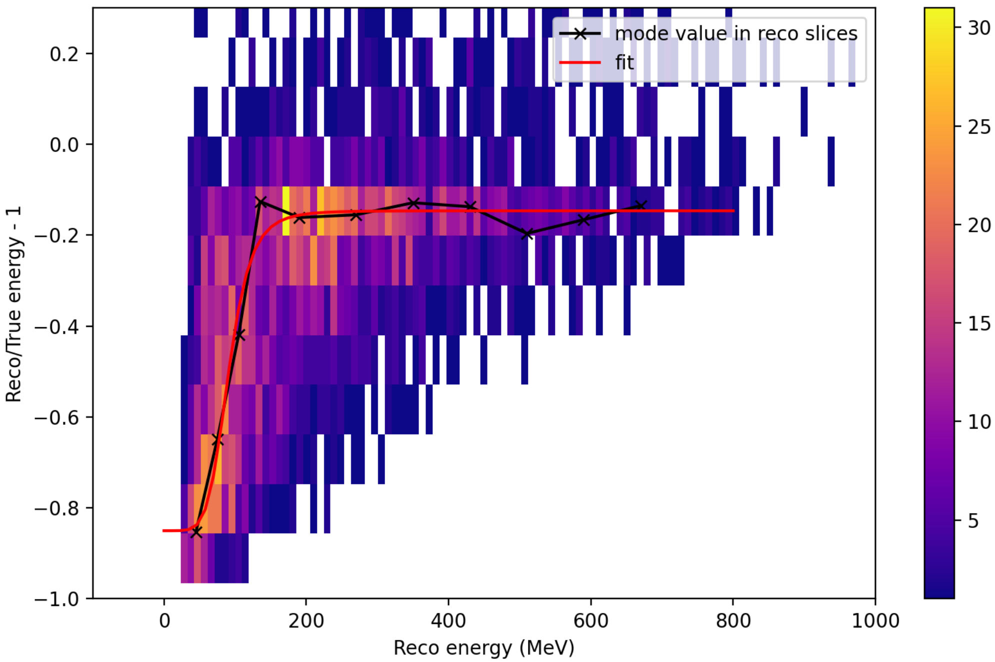
Shower energy correction with nHits > 20

+ using only showers that pass π^0 cuts as well

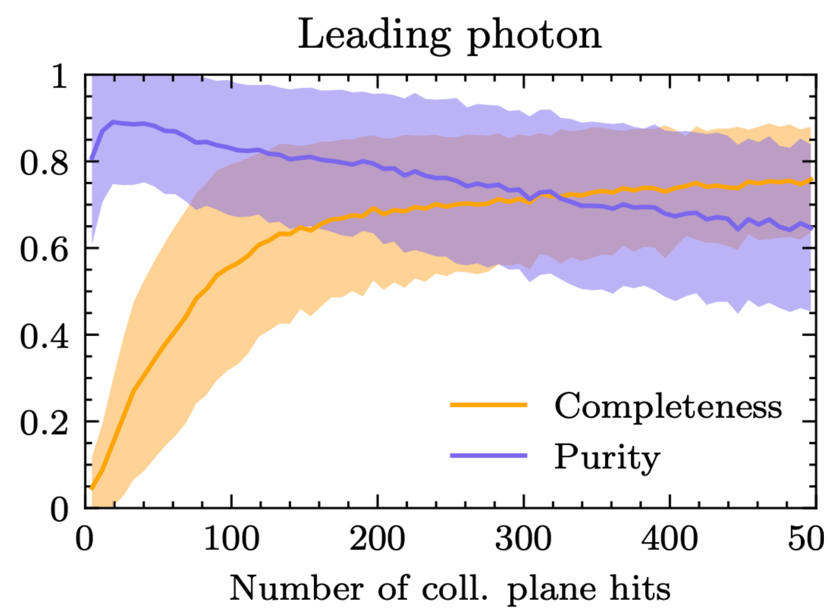
Function taken from
Kang Yang's π^0 analysis

$$p_3 + \frac{(p_0 - p_3)}{1 + \left(\frac{x}{p_2}\right)^{p_1}}$$

fit with non-linear function:
p0 =-0.85 +/- 0.05
p1 =5.85 +/- 1.41
p2 =91.17 +/- 5.07 MeV
p3 =-0.146 +/- 0.018

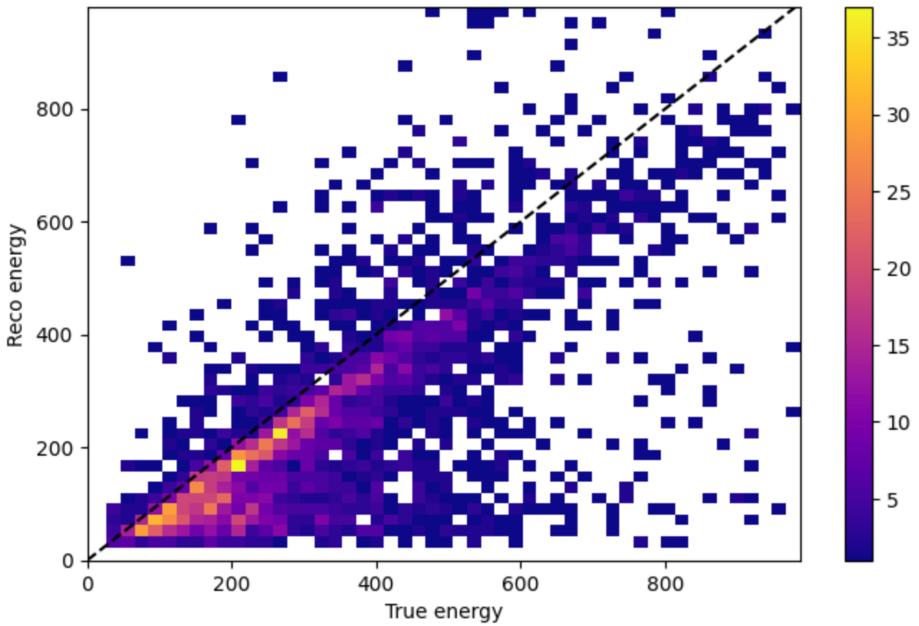


(Plot taken from **Milo Vermulin's** PhD thesis)

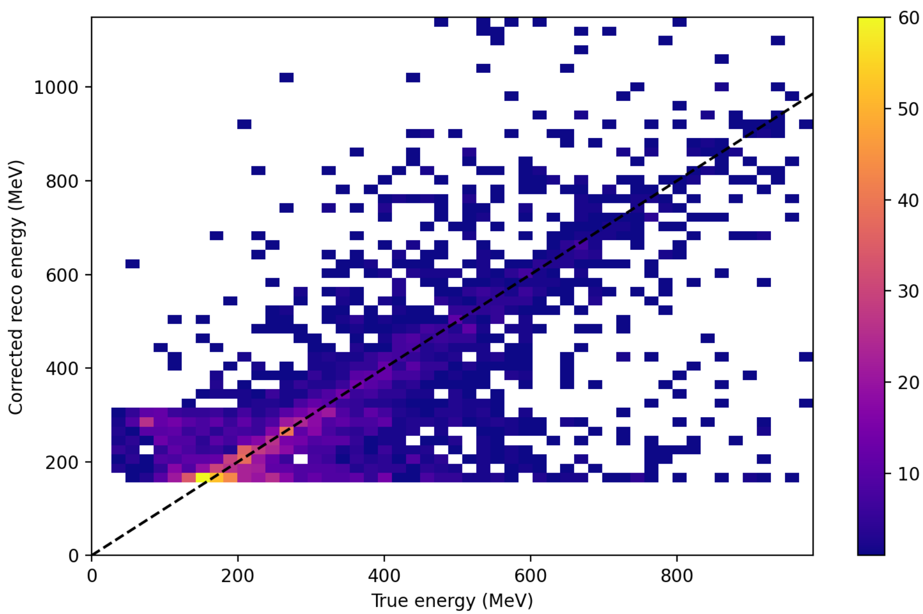


Effect of energy correction on energy reconstruction bias

- The correction method does a reasonable job but:
  - showers with a wide true shower energy profile contribute to low nHits reco showers.
  - A single correction factor can only redistribute the offset from mostly negative to both negative and positive sides at best.



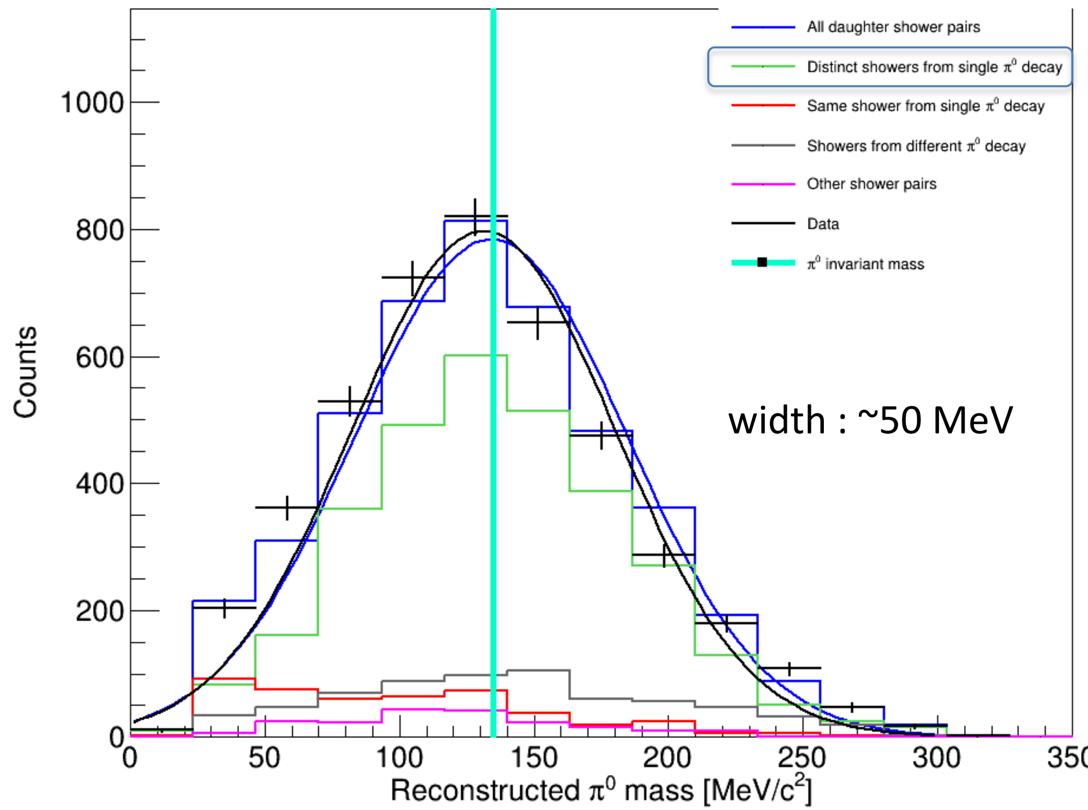
Before correction



After correction

Reconstructed π^0 mass for Nhits>20

- Fit results for π^0 reconstructed mass agree well with the true invariant mass for both data and MC.
  - But as we see from the corrected reco energy, it is not possible to find a meaningful energy correction at low energy/nHits.
  - We need improvement at reco level for this analysis.



Gaussian fit result:
MC: 134.24 ± 1.5 MeV/c^2
Data: 131.41 ± 0.88 MeV/c^2

π^0 event selection in numbers for Nhits>20

	Data	MC	Ratio (Data/MC)	Purity (Cumulative)	Efficiency (Cumulative)
nHits >20	140012	31708	4.41	0.871	0.412
Shower pair	43552	10400	4.18	0.442	1 (0.068 of all showers)
0.1 rad < Opening angle < 0.7 rad	11078	3260	3.39	0.674	0.482
π^0 energy < 1000 MeV/c^2	8828	2804 (1966)	3.14	0.701	0.431

~3 times more events compared to nHits>80.

Summary and Future plans

Summary / Future plans

- Summary:
  - The π^0 analysis works reasonably for well reconstructed data (i.e. nHits >80):
    1. We have a very low efficiency in finding both π^0 showers due to missing/incomplete reconstruction.
    2. Including data from low nHits is currently not useful as shower energy correction doesn't work.

- Future plans:
  - Any suggestions on what should be done next are welcome.
  - I will work on Pandora reconstruction to:
    - understand why we don't see most π^0 events completely during the Pandora reconstruction stage.
    - Improve completeness of showers in general (if possible).

Thanks for your attention

- I would also like to thank a few particular people for their contribution in this work:
  - Milo Vermeulen and Kang Yang for their excellent work on π^0 event reconstruction.
  - Jake Calcutt for creating the analysis ntuples.
  - Miguel Ángel García-Peris for many helpful discussions on ProtoDUNE.

Backup slides

Event reconstruction examples

```
==================================
==================================
Print full event
particle PDG code: 211 , particle id: 1 Is a track
Daughter index , PDG code , particle id , parent end process , Pandora classification , parent PDG , parent id
0 , -211 , 600591 , pi-Inelastic, Track , 211 ,  1
1 , 1000010020 , 601363 , hIoni, Track , 211 ,  600590
2 , 1000010020 , 601363 , hIoni, Track , 211 ,  600590
3 , 22 , 601947 , conv, Track , 111 ,  601351
4 , 22 , 601947 , conv, Shower , 111 ,  601351
5 , 22 , 601947 , conv, Shower , 111 ,  601351
6 , 2212 , 600711 , hIoni, Track , 2112 ,  600594
==================================
level 2
==================================
Daughter index , PDG code , particle id , parent end process , Pandora classification , parent PDG , parent id
==================================
==================================
```

a photon is reconstructed as three daughters, second shower is missing

Event reconstruction examples

```
Print full event

particle PDG code: 211 , particle id: 1 Is a track

Daughter index , PDG code , particle id , parent end process , Pandora classification , parent PDG , parent id

0 , -13 , 640056 , Decay, Track , 211 ,  11

1 , -13 , 640056 , Decay, Track , 211 ,  11

2 , 22 , 647282 , conv, Shower , 111 ,  647138

3 , -13 , 640056 , Decay, Track , 211 ,  11

4 , 22 , 647282 , conv, Shower , 111 ,  647138

5 , -13 , 640056 , Decay, Shower , 211 ,  11

==================================

level 2

==================================

Daughter index , PDG code , particle id , parent end process , Pandora classification , parent PDG , parent id

primary daughter 1 has 1 secondary daughters

0 , -13 , 640056 , Decay, Track , 211 ,  11

==================================

==================================
```

a photon is reconstructed as two daughters, second shower is missing

Event reconstruction examples

```
Print full event

particle PDG code: 211 , particle id: 1 Is a track

Daughter index , PDG code , particle id , parent end process , Pandora classification , parent PDG , parent id

0 , 211 , 1 , pi+Inelastic, Track , -999 ,  -999

1 , 2212 , 737979 , hIoni, Track , 211 ,  1

==================================

level 2

==================================

Daughter index , PDG code , particle id , parent end process , Pandora classification , parent PDG , parent id

primary daughter 1 has 2 secondary daughters

0 , 211 , 737976 , Decay, Track , 211 ,  1

1 , 22 , 738742 , conv, Shower , 111 ,  737975

==================================

==================================

primary is beam particle 1

Got primary CNN score 0.101487, true PDG code: 211

Pandora track: 0x1c04a9b8, shower: 0
```

a photon is reconstructed as secondary daughter,
second shower is missing

Event reconstruction examples

```
Print full event
particle PDG code: 211 , particle id: 1 Is a track
Daughter index , PDG code , particle id , parent end process , Pandora classification , parent PDG , parent id
0 , 211 , 823487 , pi+Inelastic, Track , 211 ,  1
1 , 22 , 824115 , conv, Track , 111 ,  823485
2 , 2212 , 823488 , hIoni, Track , 211 ,  1
==================================
level 2
==================================
Daughter index , PDG code , particle id , parent end process , Pandora classification , parent PDG , parent id
primary daughter 1 has 1 secondary daughters
0 , 22 , 824115 , conv, Shower , 111 ,  823485
==================================
==================================
primary is beam particle 1
Got primary CNN score 0.168034, true PDG code: 211
Pandora track: 0x127ec1f0, shower: 0
```

first part is considered track by pandora

a photon is reconstructed as its own daughter, second shower is missing

Event reconstruction examples

```
==================================
==================================
Print full event
particle PDG code: 211 , particle id: 1 Is a track
Daughter index , PDG code , particle id , parent end process , Pandora classification , parent PDG , parent id
0 , 22 , 680782 , conv, Shower , 111 ,  680351
1 , 22 , 680783 , conv, Shower , 111 ,  680351
==================================
level 2
==================================
Daughter index , PDG code , particle id , parent end process , Pandora classification , parent PDG , parent id
==================================
==================================
primary is beam particle 1
Got primary CNN score 0.38757, true PDG code: 211
Pandora track: 0x33c38b68, shower: 0
```

both shower are reconstructed as primary daughters

Event reconstruction examples

```
==================================

==================================

Print full event

particle PDG code: 211 , particle id: 1 Is a track

Daughter index , PDG code , particle id , parent end process , Pandora classification , parent PDG , parent id

0 , 211 , 664181 , pi+Inelastic, Track , 211 ,  1

1 , 22 , 666247 , conv, Shower , 111 ,  664179

2 , 22 , 664581 , conv, Track , 111 ,  664182

==================================

level 2

==================================

Daughter index , PDG code , particle id , parent end process , Pandora classification , parent PDG , parent id

==================================

==================================

primary is beam particle 1

Got primary CNN score 0.378347, true PDG code: 211

Pandora track: 0x5efcd768, shower: 0
```

Two showers from two different Pi0's,
only one is identified as a shower