Use of calorimetric profiles to separate merged showers in Pandora

Maria Brigida Brunetti

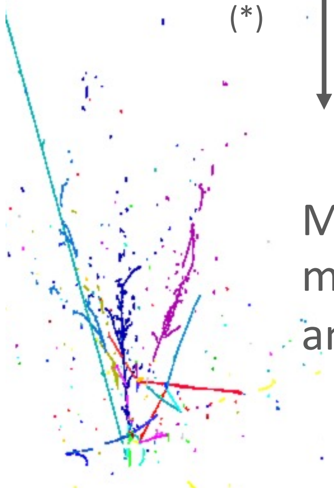
FD Sim/Reco meeting - 9 May 2022

WARWICK
THE UNIVERSITY OF WARWICK

Calorimetric profiles

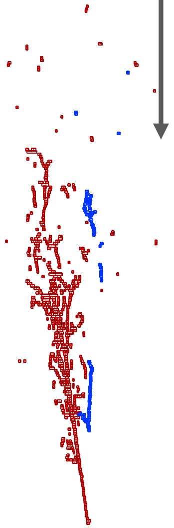
- Previous sim/reco presentation on π^0 mass reconstruction: https://indico.fnal.gov/event/53402/
  Merging of showers identified as a leading cause of inefficiency
- Exploring use of calorimetric shower profiles to split merged showers
- Start from **transverse profiles**, look at **longitudinal profiles** at a later stage

(*)



Most useful when merged showers have angular separation

(*)

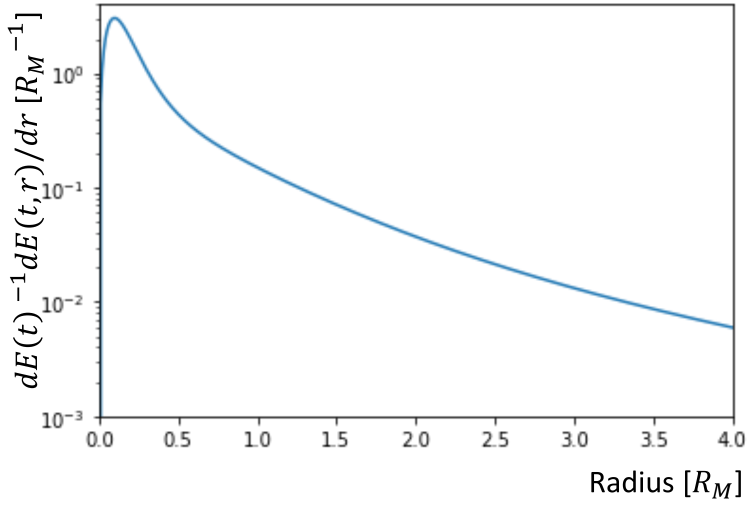


Most useful when two merged showers start at different depths along the principal axis

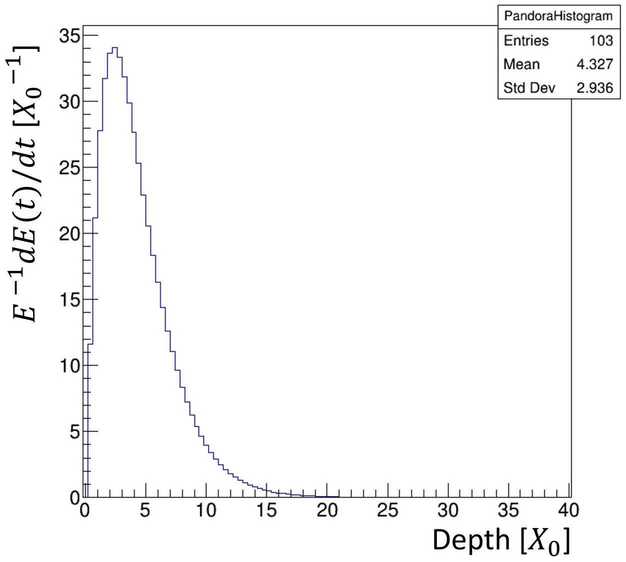
*cheated patrec

Calorimetric profiles (2)

- Previous sim/reco presentation on π^0 mass reconstruction: https://indico.fnal.gov/event/53402/
  Merging of showers identified as a leading cause of inefficiency
- Exploring use of calorimetric shower profiles to split merged clusters
- Start from **transverse profiles**, look at **longitudinal profiles** at a later stage

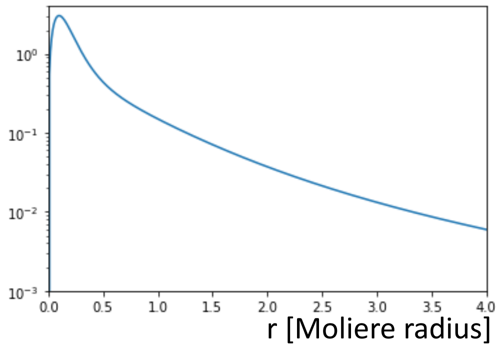


Example expected radial profile for a single 500 MeV shower

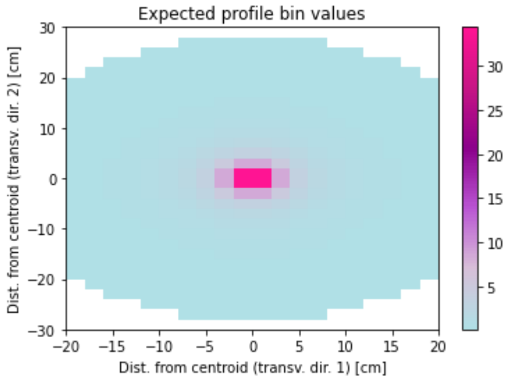


Example expected longitudinal profile for a single 500 MeV shower

Transverse profiles

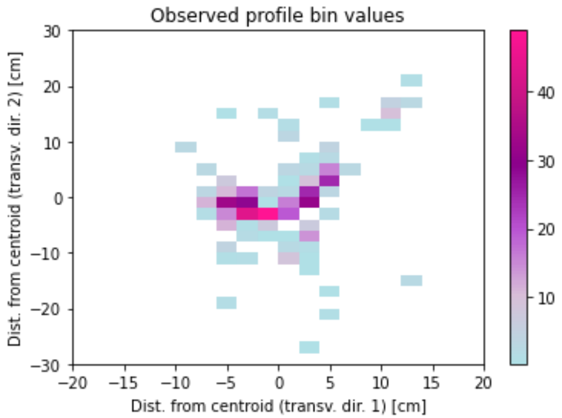


From radial to 2D transverse

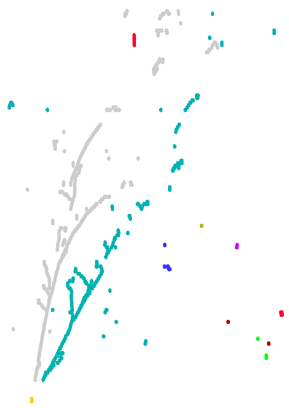


Expected

- Transverse profiles in a plane orthogonal to the shower principal axis (PA)
- For expected profiles, used parametrization from arXiv:hep-ex/0001020v1
- Expected profiles follow a radially symmetric distribution with shape determined by the depth along the PA and the deposited energy
- In observed profiles, different structures can arise due to merging



Observed (ADCs summed and converted to MeV)



NC RES in FD
500 MeV
merged cluster

Transverse profiles - parametrization

- from arXiv:hep-ex/0001020v1

$$f(r) = \frac{1}{dE(t)}\frac{dE(t,r)}{dr},$$

$$\begin{aligned} f(r) &= pf_C(r) + (1-p)f_T(r) \\ &= p\frac{2rR_C^2}{(r^2+R_C^2)^2} + (1-p)\frac{2rR_T^2}{(r^2+R_T^2)^2} \end{aligned}$$

$$\begin{aligned} R_{C,hom}(\tau) &= z_1 + z_2\tau \\ R_{T,hom}(\tau) &= k_1\{\exp(k_3(\tau - k_2)) + \exp(k_4(\tau - k_2))\} \\ p_{hom}(\tau) &= p_1 \exp\left\{\frac{p_2-\tau}{p_3} - \exp\left(\frac{p_2-\tau}{p_3}\right)\right\} \end{aligned}$$

with

$$\begin{aligned} z_1 &= 0.0251 + 0.00319 \ln E \\ z_2 &= 0.1162 + -0.000381 Z \\ k_1 &= 0.659 + -0.00309 Z \\ k_2 &= 0.645 \\ k_3 &= -2.59 \\ k_4 &= 0.3585 + 0.0421 \ln E \\ p_1 &= 2.632 + -0.00094 Z \\ p_2 &= 0.401 + 0.00187 Z \\ p_3 &= 1.313 + -0.0686 \ln E \end{aligned}$$

t = longitudinal shower depth in units of radiation length
τ = t/T = shower depth in units of the depth of the shower maximum
r = radial distance from the shower axis in Moliere radius units
E = shower energy in units of critical energy

ADC->MeV conversion factor=0.0075 MeV/ADC
Argon properties
https://pdg.lbl.gov/2014/AtomicNuclearProperties/HTML/liquid_argon.html

Transverse profiles in Pandora

- Use transverse profiles to drive reclustering in Pandora:
  - Compare expected and observed transverse profiles bin-by-bin
  - Define a figure of merit (FOM) quantifying agreement between observed and predicted profiles

$$FOM = \sqrt{\sum_{Bins} \frac{(N_{OBS} - N_{EXP})^2}{N_{EXP}}}$$

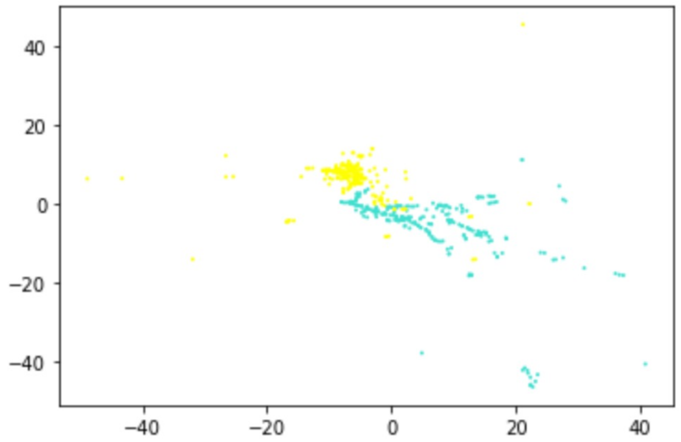
  - Use clustering algorithm (e.g. k-Means) to predict new cluster centers and distributions in the transverse plane under the hypothesis of N=2 (or more) clusters
  - Recalculate FOM for new clusters
  - Compare new and old FOMs to decide whether the particle should be dissolved and N particles should be created in Pandora
  - The plan is to base the reclustering off the 3D clusters

Transverse profiles - approximation

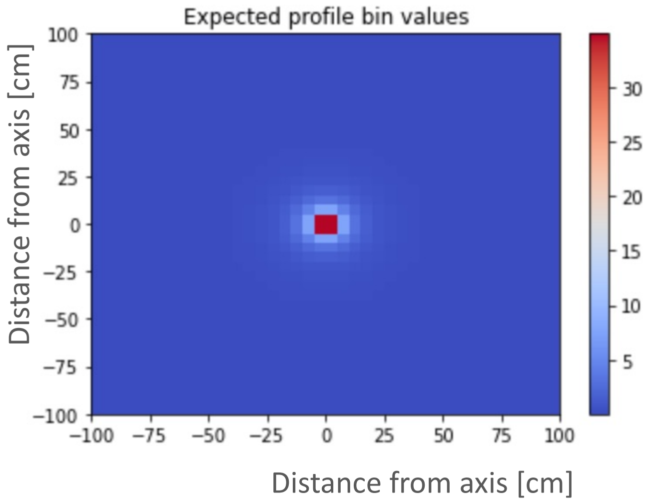
- The used parametrization gives expected transverse profiles at a specific shower depth
- Inputs are the shower depth and the energy deposited at that depth
- In Pandora, we want to make use of all hits in the shower
- First approximation:
  - Fix the shower depth to be the maximum
  - Assume profile shape is the that of the maximum across whole shower, so that we can consider energy deposits from all hits
  - Binned transverse profile is normalised to entire energy in the shower
- Second approximation:
  - Sample a few shower depths and consider hits in slices around the sampling depth

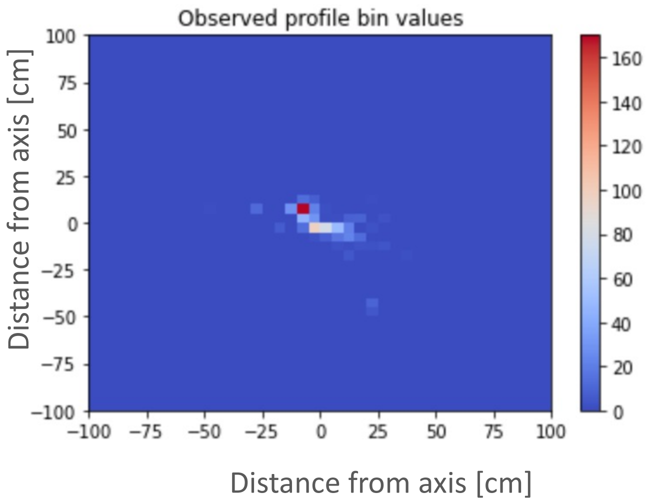
Using Pandora 3D hits as inputs (MCC11 nue events, reconstructed with streaming) and the first approximation, I performed a study on a small sample – examples in next slides

Example – two merged showers



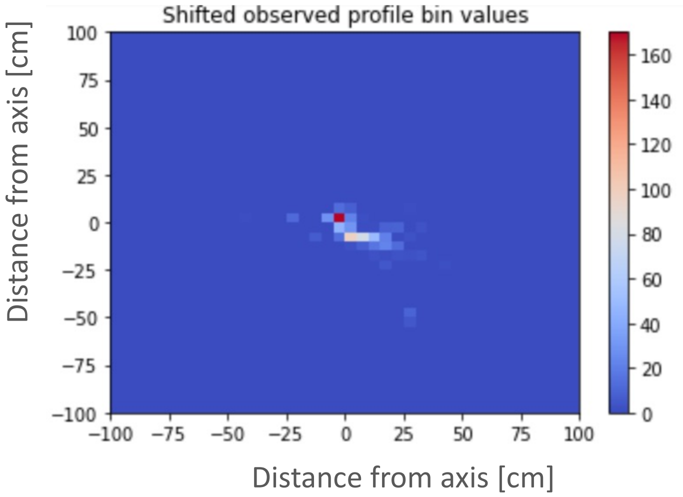
Same hit color → same main contributing MC particle





FOM = 70

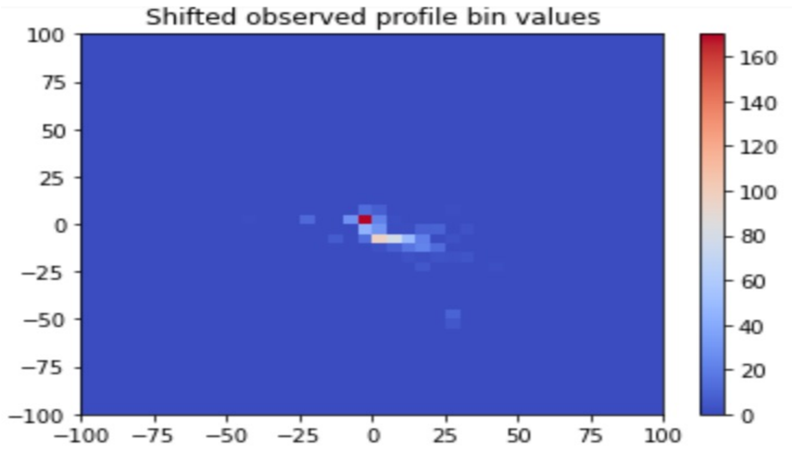
Can try to shift observed profile to find minimum FOM:



FOM = 61

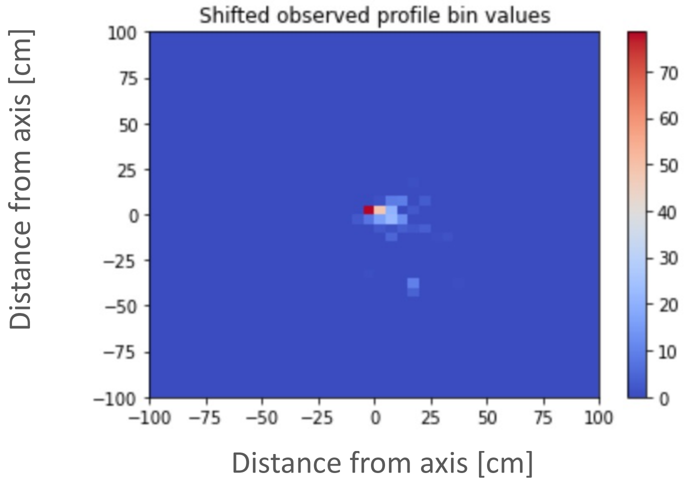
Example – two merged showers (2)

What happens if I split the merged shower in two, using truth info?



CLUSTER 1

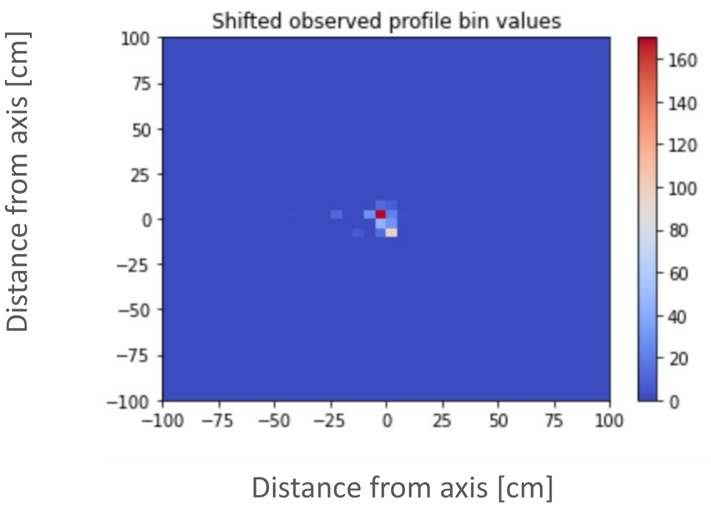
(*)



FOM = 40 < 61

CLUSTER 2

(*)

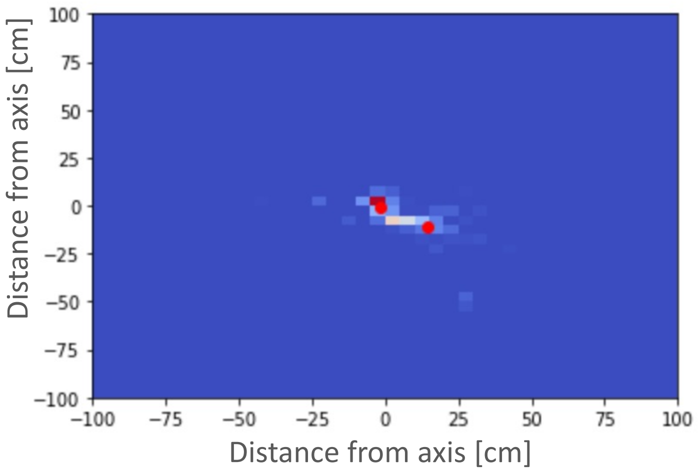


FOM = 43 < 61

* Shifted to recenter on expected profile:
shower axis is **not** recalculated in this study!

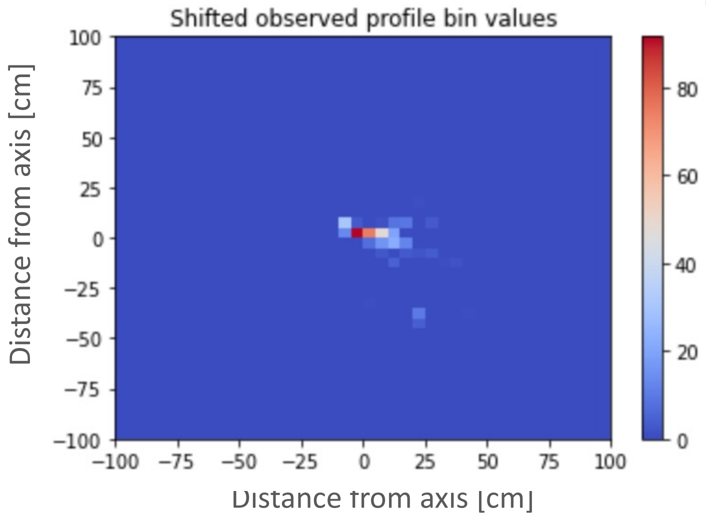
Example – two merged showers (3)

What happens if I split the merged shower in two, using kMeans?



Red dots represented new cluster centers predicted by kMeans

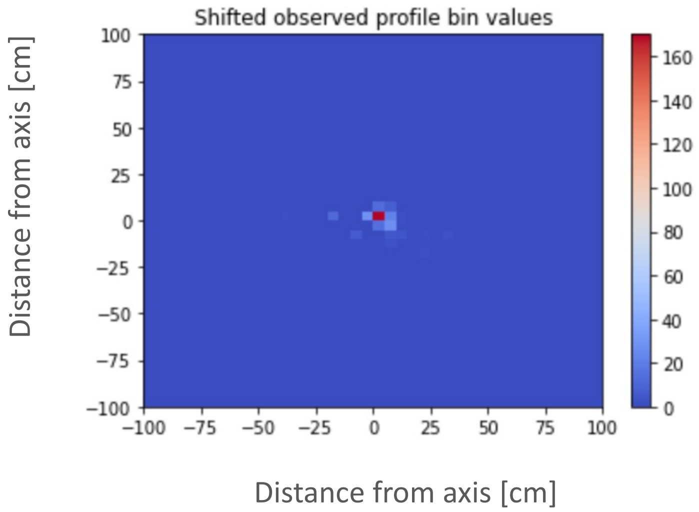
CLUSTER 1



(*)

FOM = 45 < 61

CLUSTER 2



(*)

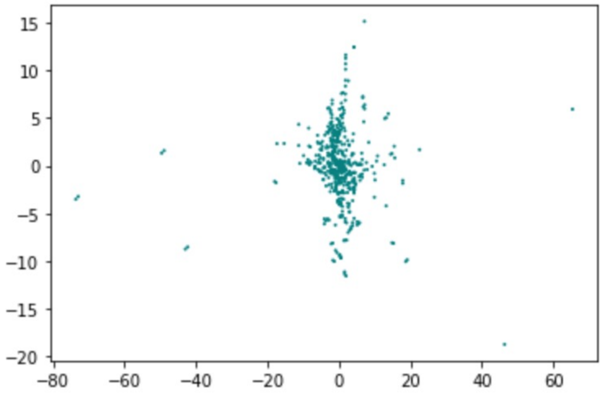
FOM = 35 < 61

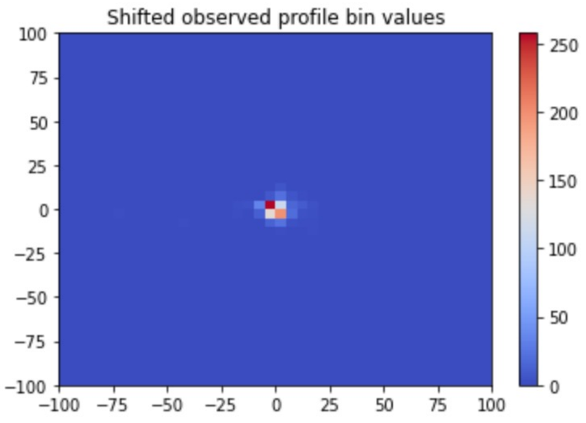
* Shifted to recenter on expected profile: principal axis is **not** recalculated in this study!

An example - single shower

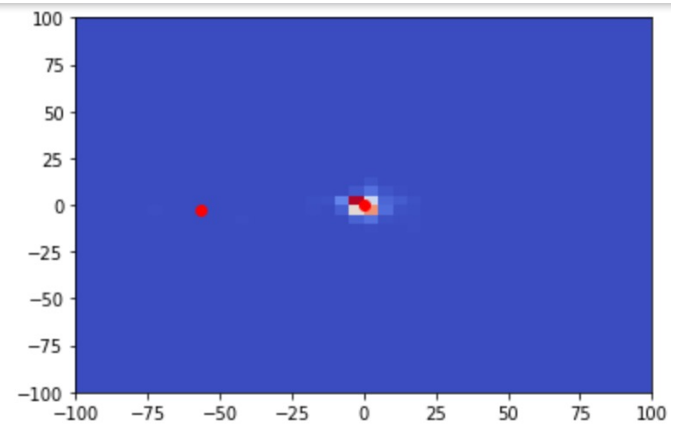
905 MeV

(All profiles shifted to find minimum Chi2)





FOM = 34.9



Read dots = kMeans predicted cluster centers

This cluster should not be split in two!

kMeans creates a cluster with very little energy

→ Add a threshold on new cluster energy

CLUSTER 1

Energy = 902 MeV
FOM = 34.5

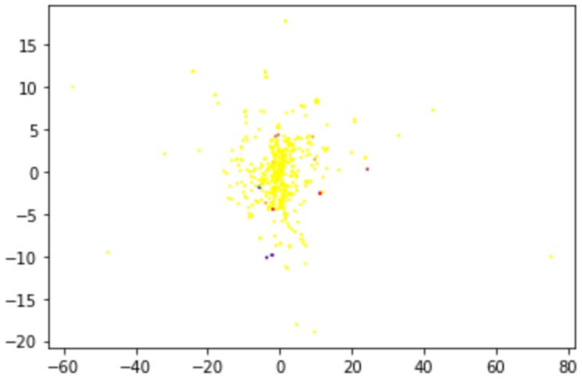
CLUSTER 2

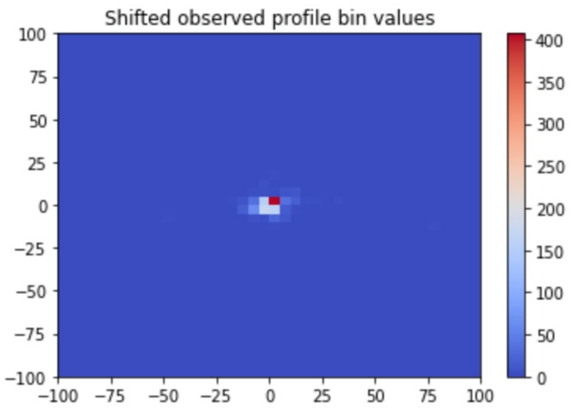
Energy = 3 MeV
FOM = 14

An example where N=2 should make things worse (2)

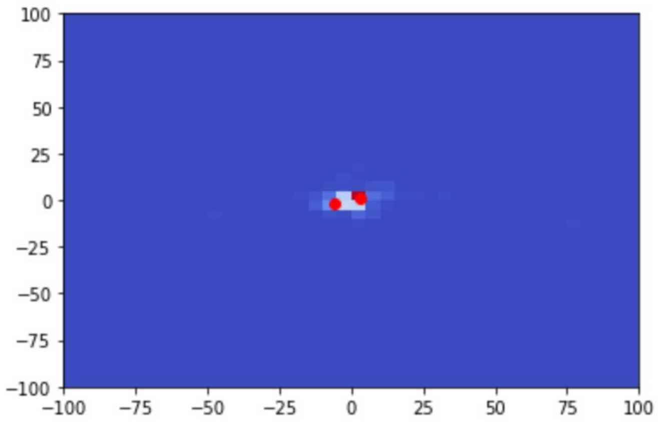
(All profiles shifted to find minimum Chi2)

1190 MeV





FOM = 43



Read dots = kMeans predicted cluster centers

This cluster should not be split in two!

Only one of the two clusters gets a better FOM than the initial one → reject such cases

CLUSTER 1

Energy = 467 MeV
FOM = 33.5

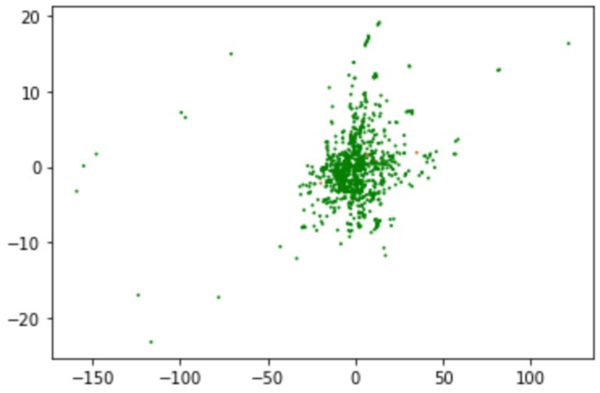
CLUSTER 2

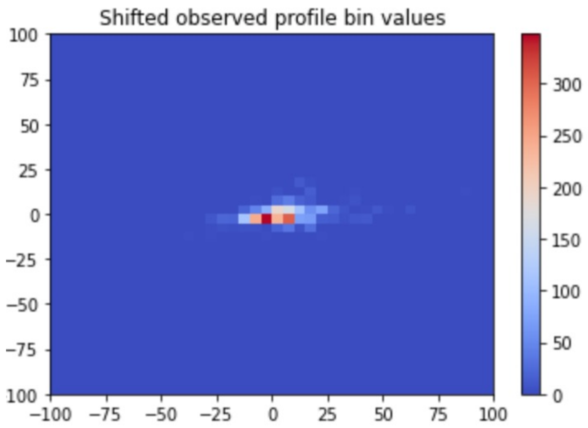
Energy = 722 MeV
FOM = 53

An example where N=2 should make things worse (2)

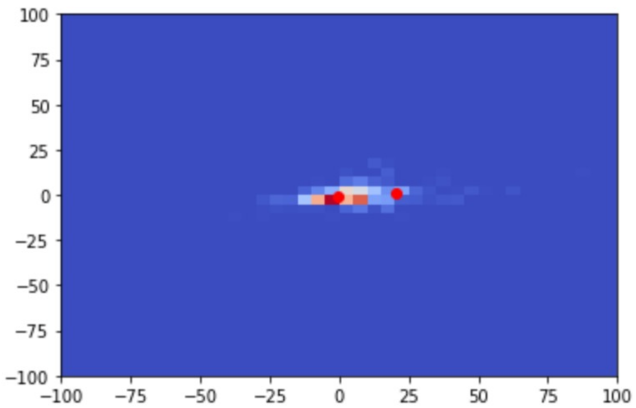
(All profiles shifted to find minimum Chi2)

2545 MeV





FOM = 82

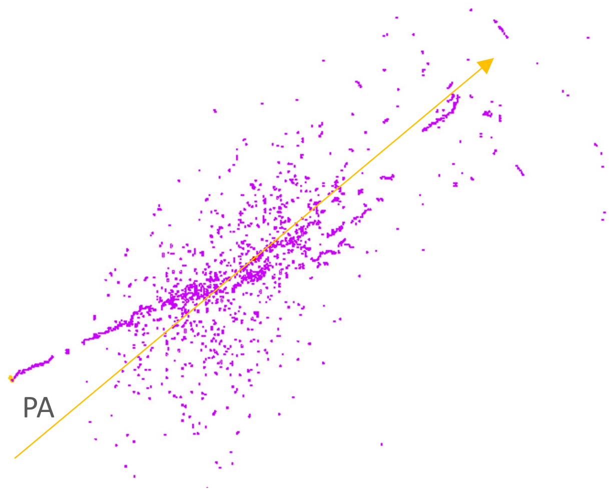


Read dots = kMeans predicted cluster centers

This cluster would be split even if it should not be…

I think the resulting splits clusters just get more symmetric

Looking at event display, 3D hits for this event are not perfect



CLUSTER 1

Energy = 1950 MeV
Chi2: = 74

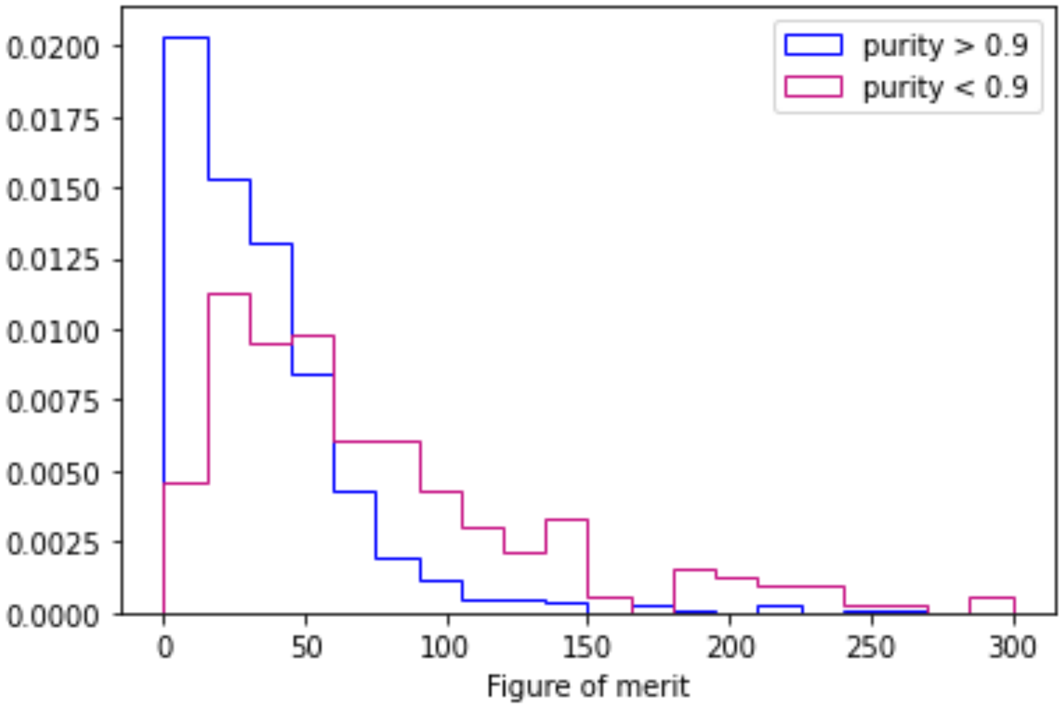
CLUSTER 2

Energy = 592 MeV
Chi2: = 42

A look at a small sample

- Start with 1000 pfos
- Using truth information, defined two samples:
  - Main MC particle contributes at least 90% hits **(single shower)**
  - Main MC particle contributes < 90% hits **(merged shower)**

(Note the two samples have unequal number of events)



- Additional selection cuts, common to both samples
  - Minimum energy: 20 MeV
  - Main MC particle is a photon or electron

$$FOM(68\%) = \frac{41\ (single\ showers)}{99\ (merged\ showers)}$$

First indication that splitting merged showers would yield a lower FOM

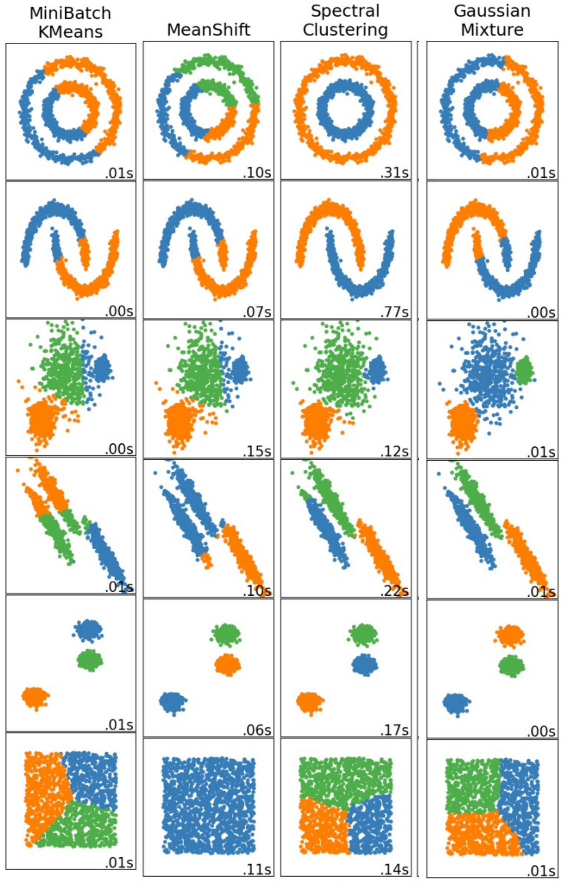
Conclusions and to-do

- First positive indications that transverse shower profiles can be used to split merged showers in Pandora
- Look at higher stats
- Test second approximation using multiple shower depths
- Test different clustering algorithms (kMeans is not very good with elongated clusters)
- Test different figures of merit
- Implement reclustering in Pandora, starting from 3D clusters, following ILC approach

Clustering algorithms

- Look at other algorithms besides k-Means clustering

(kMeans not good for elongated profile shapes)

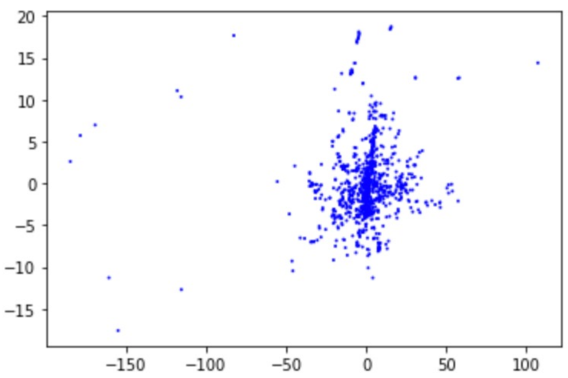


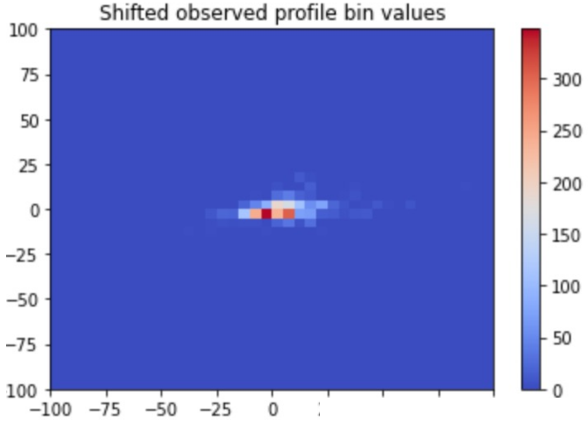
For us **samples** = bins in the transverse profile

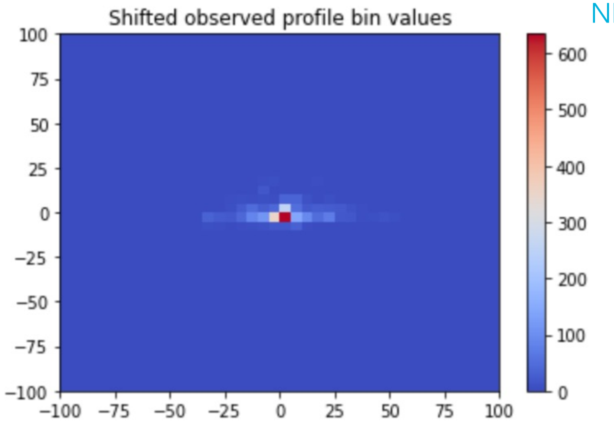
K-Means: one parameter (number of clusters), scales with large number of samples, fast, supports weights
Mean Shift: one parameter (bandwidth, size of region to search) not scalable (requires multiple nearest neighbour searches during the execution of the algorithm), slow
Spectral Clustering: one parameter (number of clusters), scales with medium n samples and small number of clusters, quite slow
Gaussian Mixture: many parameters, not scalable, fast. Mixture models generalize k-means clustering to incorporate information about the covariance structure of the data as well as the centers of the latent Gaussians.

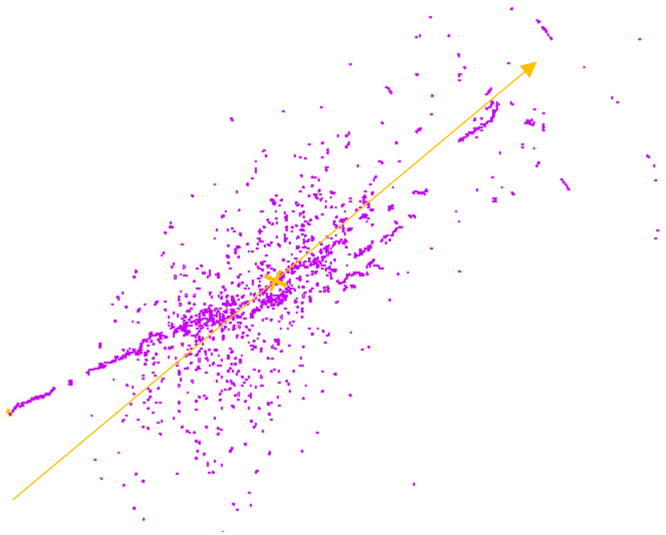
(need to figure out how to include weights in the different approaches)

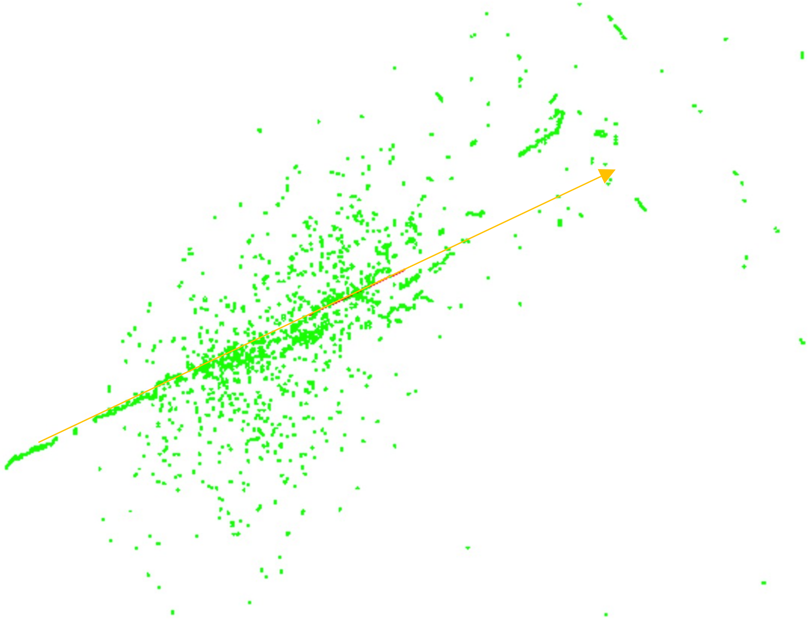
Particle n. 32, energy – 2545 MeV (higher E!)   (All profiles shifted to find minimum Chi2)











But the shape of the profile doesn't change much when I use the MC true direction of the particle instead.
I think lots of 3D hits have been reconstructed in the same plane