dune-tms make_plots.py histograms

Jeffrey Kleykamp

2023-11-30

THE UNIVERSITY of MISSISSIPPI

Location of files and documentation

- dune-tms code:
  - https://github.com/DUNE/dune-tms
- ND production (outdated compared to the newer version that Alex Booth is working on, and should only be used for small tests, not full production)
  - https://github.com/DUNE/ND_Production
- dune-tms tutorial
  - https://github.com/DUNE/dune-tms/wiki/Tutorial
- dune-tms datasets:
  - Using geometry v3
    - /dune/data/users/kleykamp/tms_testing_files/2023-11-22/2023-11-20_areal_density_fix_processing_geom_v_1.0.3.root
  - https://github.com/DUNE/dune-tms/wiki/Datasets

Exactly how these plots were made

- First run det sim and reco sim
  - Using geometry v3
  - And dune-tms version tagged kleykamp_areal_density_fix_processing_2023-11-20
    - and kleykamp_make_hists_2023-11-30 for make_hists.py
  - **Do not manually run ProcessND.py commands without explicit permission from ND production**

```
python ProcessND.py --manual_geometry_override nd_hall_with_lar_tms_sand_TDR_Production_geometry_v_1.0.3.gdml  --
geometry_location /pnfs/dune/persistent/physicsgroups/dunendsim/geometries/TDR_Production_geometry_v_1.0.3/
nd_hall_with_lar_tms_sand_TDR_Production_geometry_v_1.0.3.gdml --geometry nd_hall_with_lar_tms_nosand --pot 1e15 --
outdir /pnfs/dune/scratch/users/kleykamp/nd_production_output/2023-11-20_areal_density_fix_processing_geom_v_1.0.3
--topvol volDetEnclosure --stages tmsreco --indir /pnfs/dune/persistent/users/kleykamp/nd_production_output/2023-10-
24_test_geom_v_1.0.3

jobsub_submit --group dune --role=Analysis -N 100 --OS=SL7 --expected-lifetime=12h --memory=4000MB --tar_file_name
dropbox:///pnfs/dune/persistent/users/kleykamp/dune-tms_tarfiles/2023-11-20_areal_density_fix_processing.tar.gz
file://processnd.sh

# Then hadd (combine) files
hadd /dune/data/users/kleykamp/tms_testing_files/2023-11-22/2023-11-
20_areal_density_fix_processing_geom_v_1.0.3.root /pnfs/dune/scratch/users/kleykamp/nd_production_output/2023-11-
20_areal_density_fix_processing_geom_v_1.0.3/tmsreco/FHC/00m/00/neutrino.*.tmsreco.root

# Then make plots using dune-tms tagged with kleykamp_make_hists_2023-11-30
python make_hists.py --f /dune/data/users/kleykamp/tms_testing_files/2023-11-22/2023-11-
20_areal_density_fix_processing_geom_v_1.0.3.root --preview --outdir /dune/data/users/kleykamp/tms_density_fix/plots
--name validation_2023-11-20_tagged_version.root
```

More information about creation of genie/edep files

- **Do not run ProcessND.py** commands without getting some explicit permissions from ND production
  - Hours need to be accounted for, and don't want to replicate work
  - And test jobs should be completed before running
  - And files should be saved in common locations for everyone

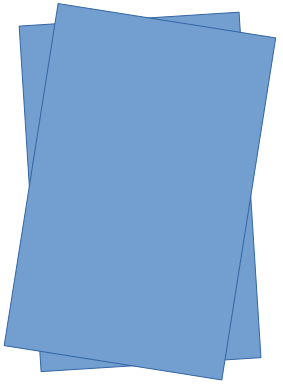
```
python ProcessND.py --manual_geometry_override
nd_hall_with_lar_tms_sand_TDR_Production_geometry_v_1.0.3.gdml  --geometry_location
/pnfs/dune/persistent/physicsgroups/dunendsim/geometries/TDR_Production_geometry_v_1.0.
3/nd_hall_with_lar_tms_sand_TDR_Production_geometry_v_1.0.3.gdml --geometry
nd_hall_with_lar_tms_nosand --pot 1e15 --outdir
/pnfs/dune/scratch/users/kleykamp/nd_production_output/2023-10-24_test_geom_v_1.0.3 --
topvol volDetEnclosure

jobsub_submit --group dune --role=Analysis -N 100 --OS=SL7 --expected-lifetime=12h --
memory=4000MB --tar_file_name dropbox:///pnfs/dune/persistent/users/kleykamp/dune-
tms_tarfiles/2023-10-24_dune-tms_no_det_sim.tar.gz file://processnd.sh

These files are then used as input "indir" in the previous slide. This is possible
because only the det sim and reco steps changed so we didn't need to rerun genie and
edep sim. Both of those require more computing power so it's nice to avoid
rerunning if possible.
```

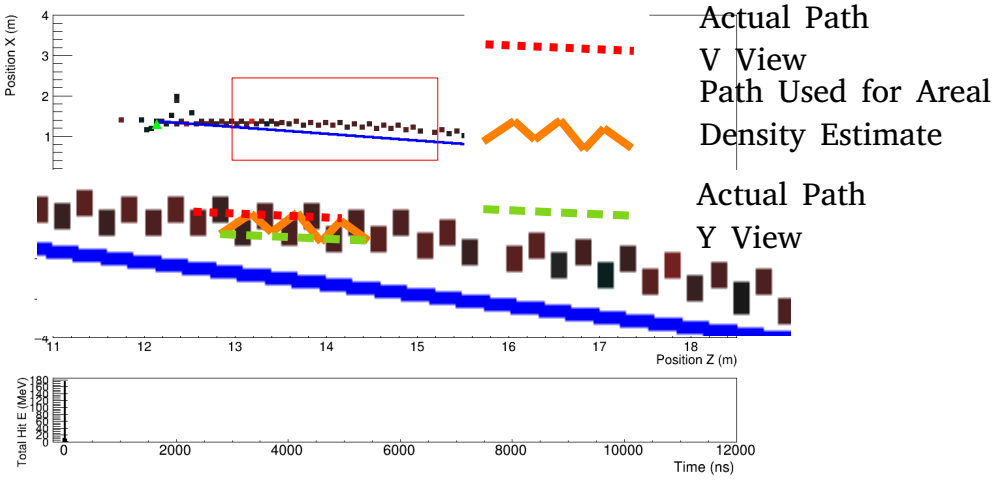
General information about make_hists.py

- The script was originally created to better understand what's going on in the Reco scripts
- It was also used to try to figure out the areal density issue
- Because of this, it has various plots that don't make too much sense now – **labeled "can ignore"**
  - Since areal density was broken, and reco couldn't be trusted, tried to define "max dz" which is the distance from the highest to lowest z hit. It's a very simplified proxy for track length
    - Now that areal density issue is fixed, these plots are basically useless
- Some hist titles are poor
  - Could be nice task to fix it and better understand the underlying histograms at the same time

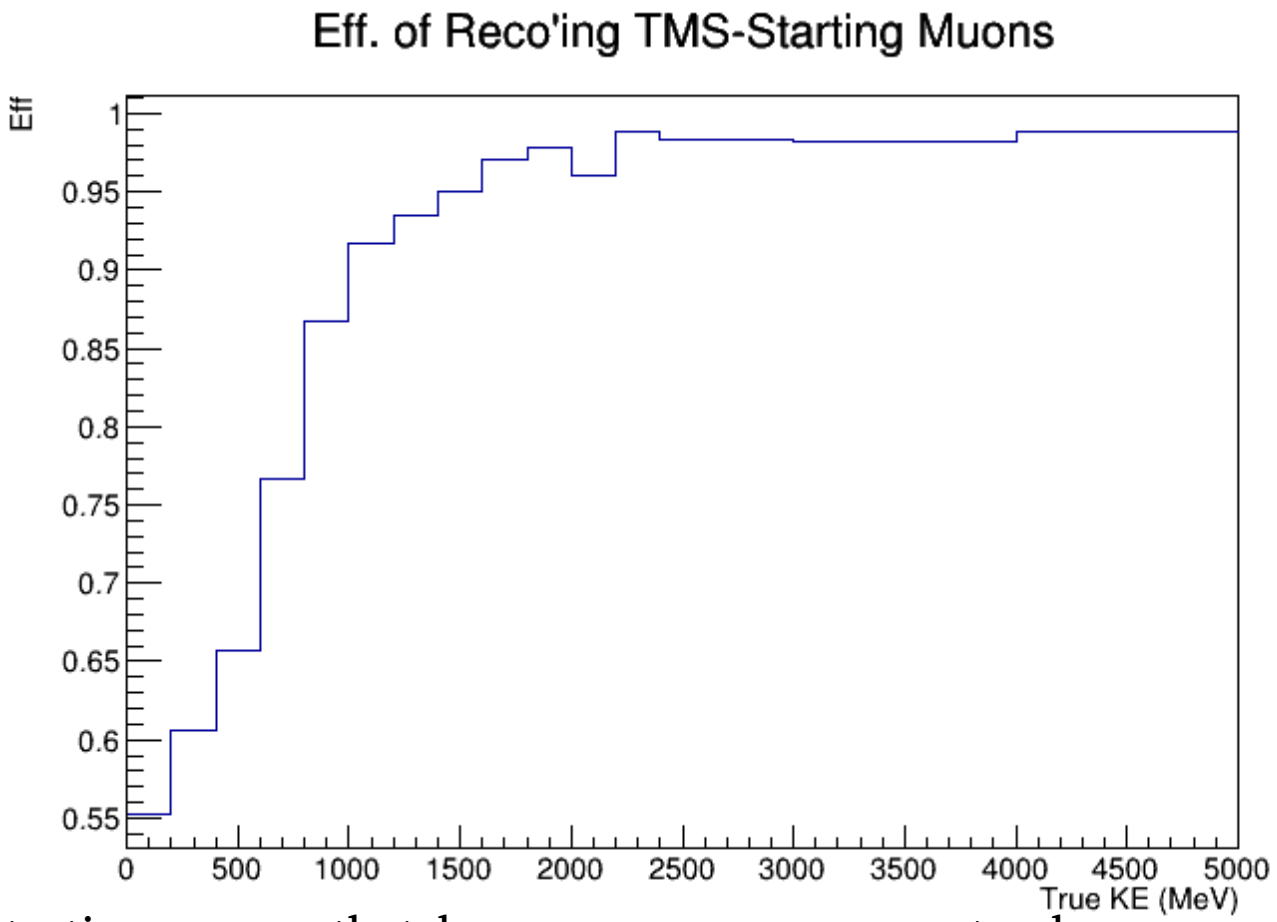
Summary of areal density fix

- tms planes are rotated
- code has areal density function that returns areal density of track
- This function would sometimes skip sections of reco track
  - Specifically because dist between nodes > 10 cm
- Would result in severe underestimate of areal density and reco muon KE
- Only a problem for geom v2+ because geom v1 had overlapping planes
- Fixed in files used for plots and
  - https://github.com/DUNE/dune-tms/pull/44





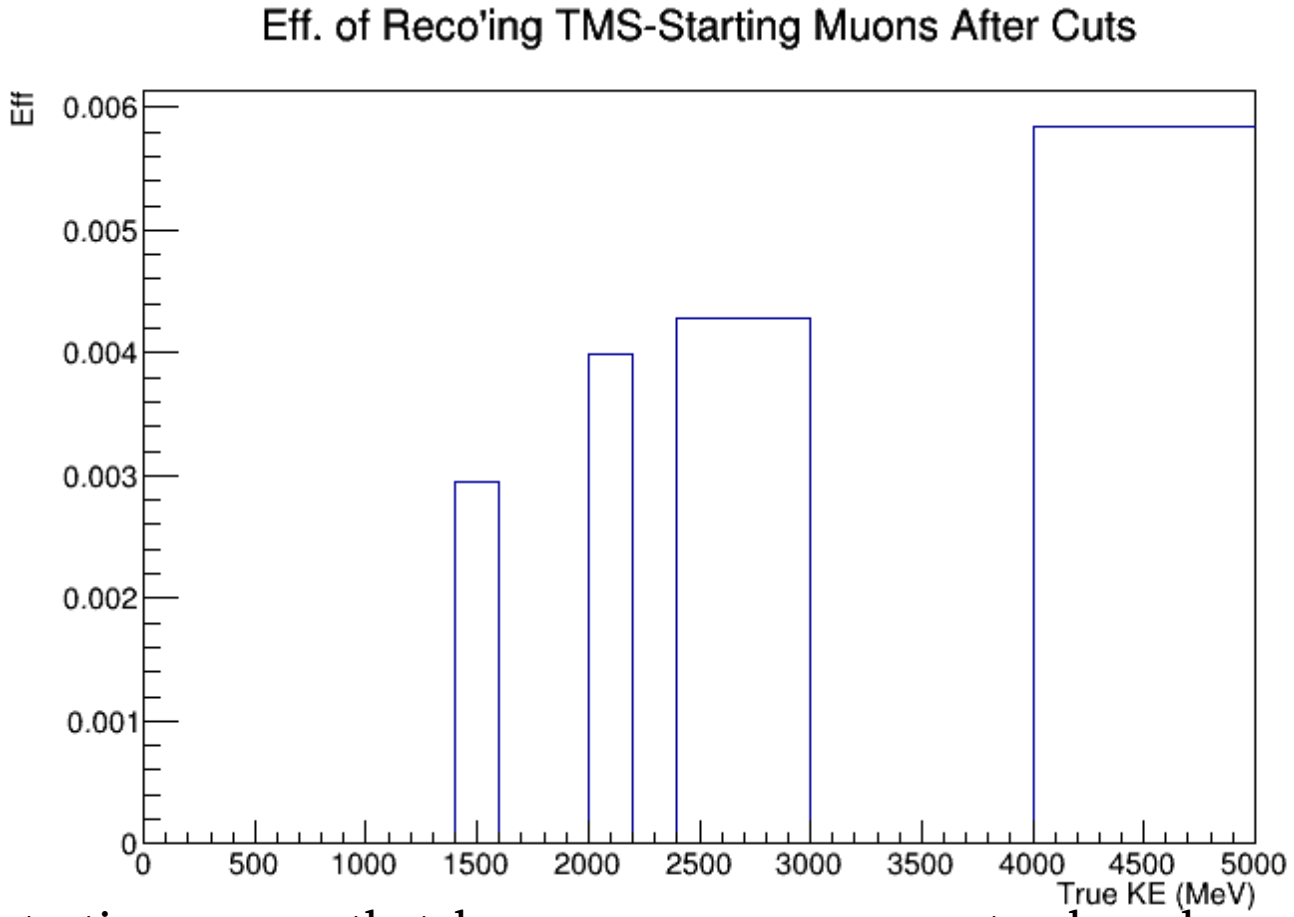
Plots in alphabetical order by filename
Plots also located in
/dune/data/users/kleykamp/tms_density_fix/plots/previews/validation_2023-11-20_tagged_version



One important note is that we're not checking that the event actually reconstructed a muon. We're only checking that there was one reco track in the event. In most high-E cases, that track is the muon, but we need to add additional information to properly study these things

$$\text{Eff} = \frac{\text{TMS-starting muons that have one or more reco track}}{\text{all TMS-starting muons}}$$

Eff including muonke.cpp-based cuts

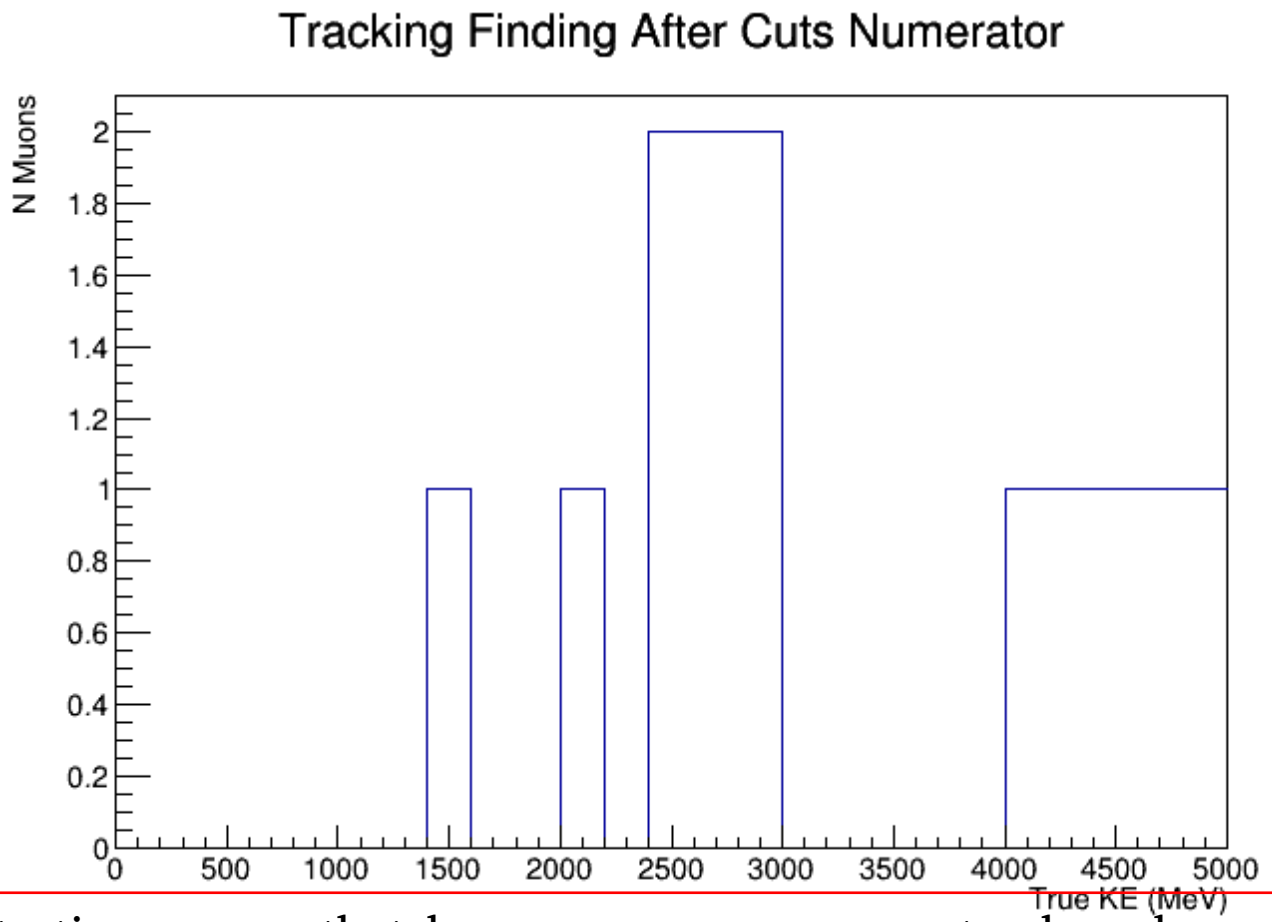


Passes these cuts. These cuts are based on cuts used in muonke.cpp:
occupancy > 0.5
track_start_z < 11472
track_end_z > 18134
abs(track_start_x) < 3320
abs(track_end_x) < 3320
-3864 < true_muon_y < 1159

See code here.

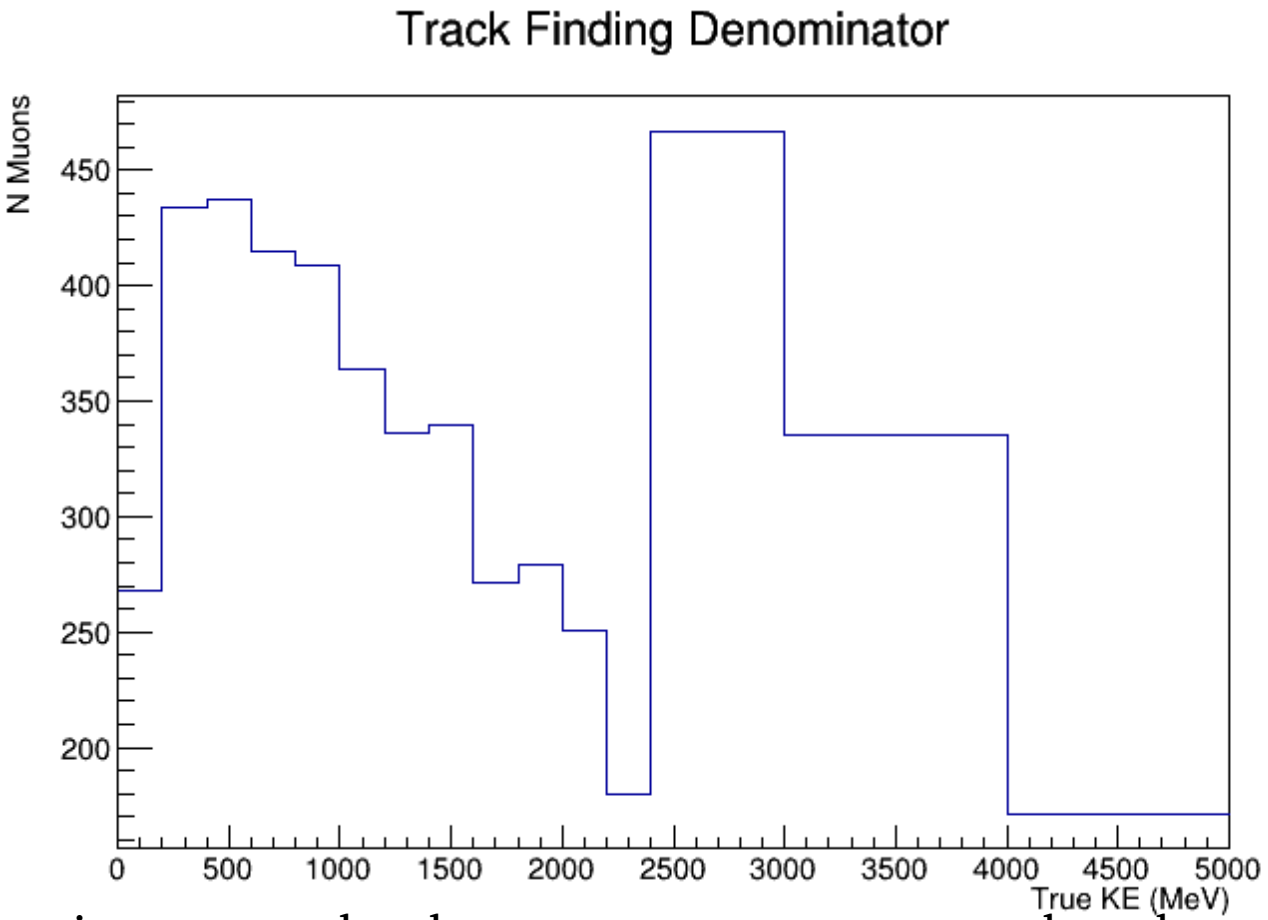
$$\text{Eff} = \frac{\text{TMS-starting muons that have one or more reco track and passes cuts}}{\text{all TMS-starting muons}}$$

Numerator of eff histogram (red box)



$$\text{Eff} = \frac{\text{TMS-starting muons that have one or more reco track and passes cuts}}{\text{all TMS-starting muons}}$$

Denominator of hist histogram



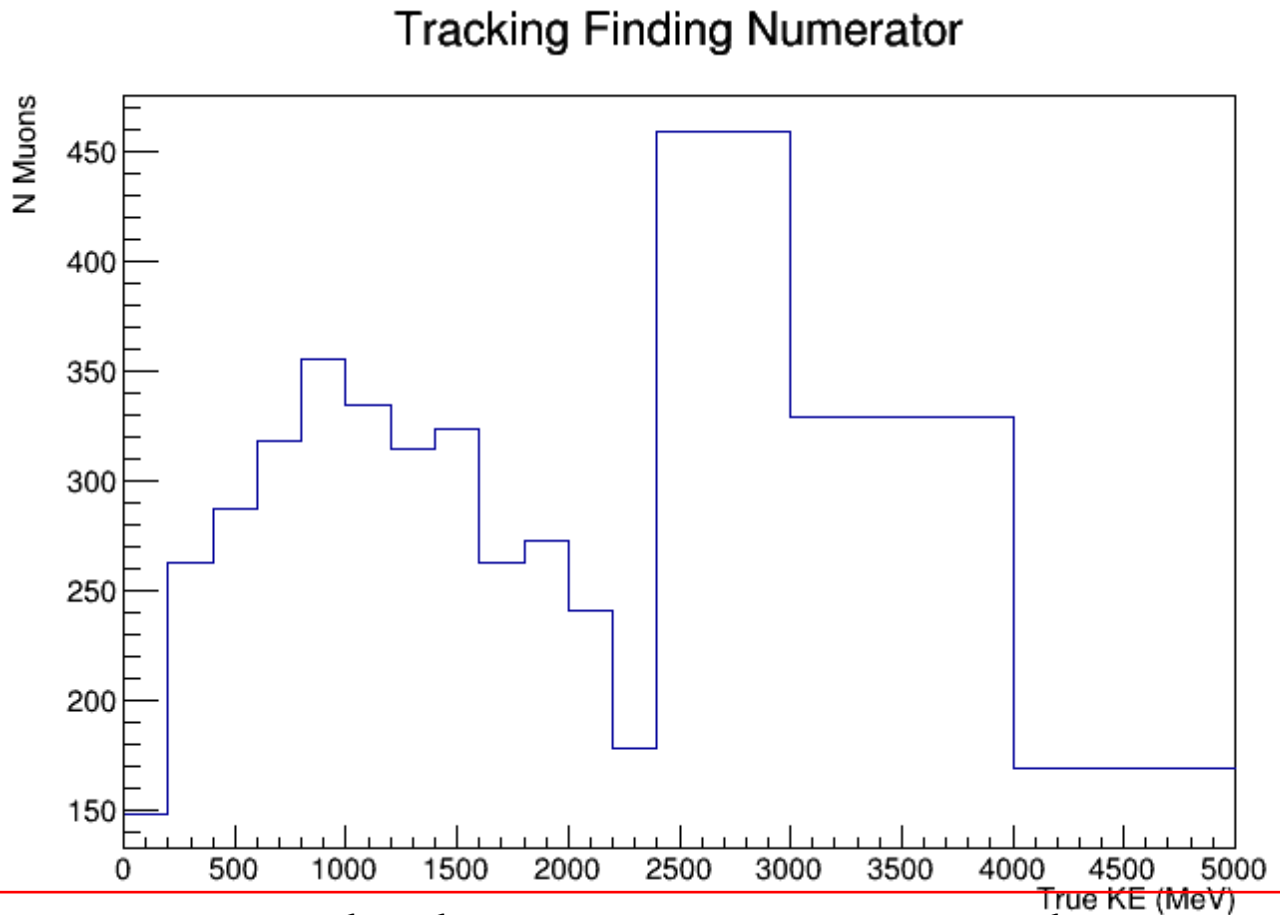
The number of events per bin looks strange but that's because each bin is variable sized. The low stats high E bins have wider bins so more events per bin

One way to make the plots look right is to use hist.Scale(1, "width") to scale per bin, but this isn't used here because we're ultimately dividing

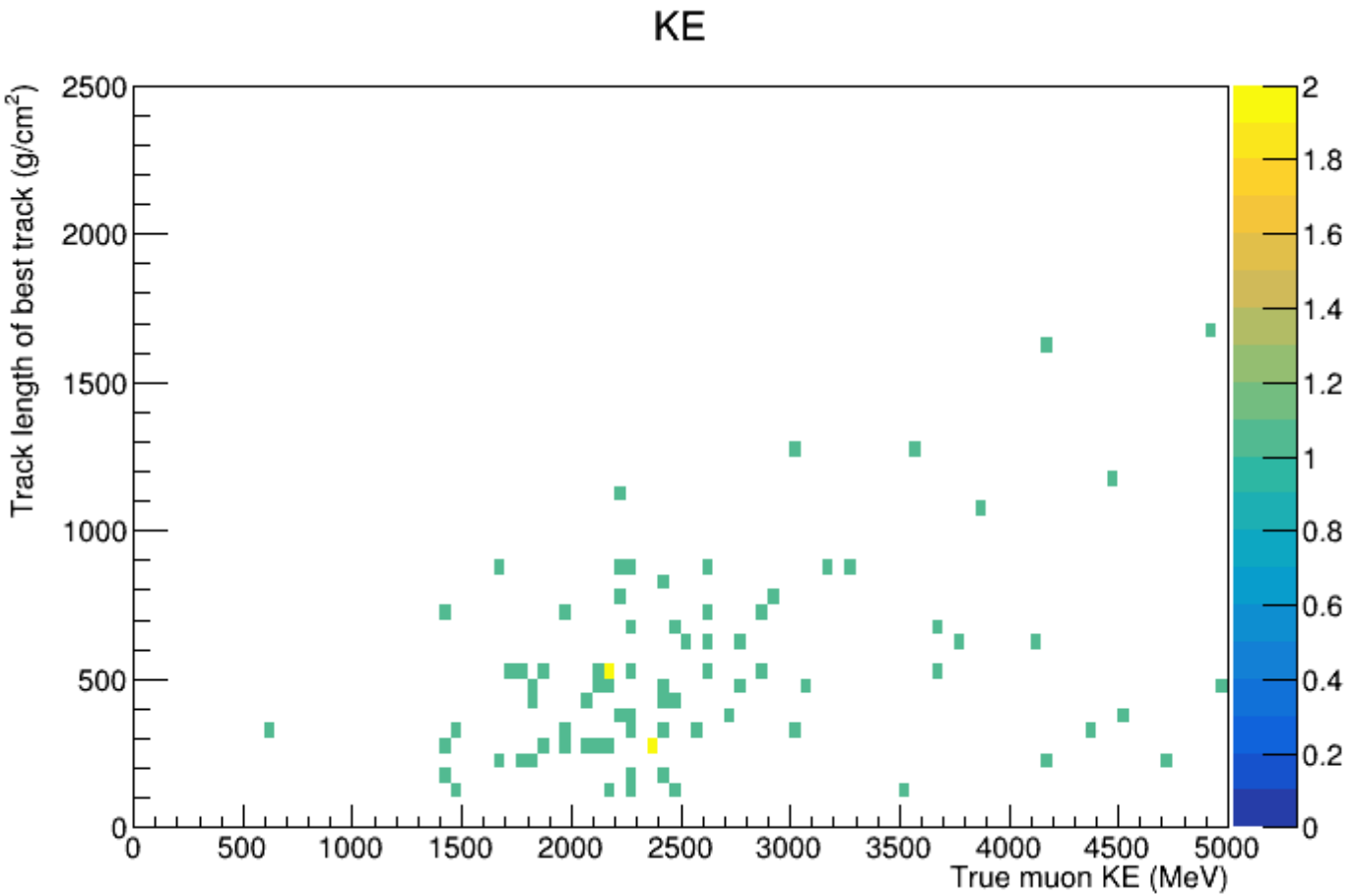
Note that the same denominator is used in eff and eff with cuts since it has all true muons.

$$\text{Eff} = \frac{\text{TMS-starting muons that have one or more reco track and passes cuts}}{\text{all TMS-starting muons}}$$

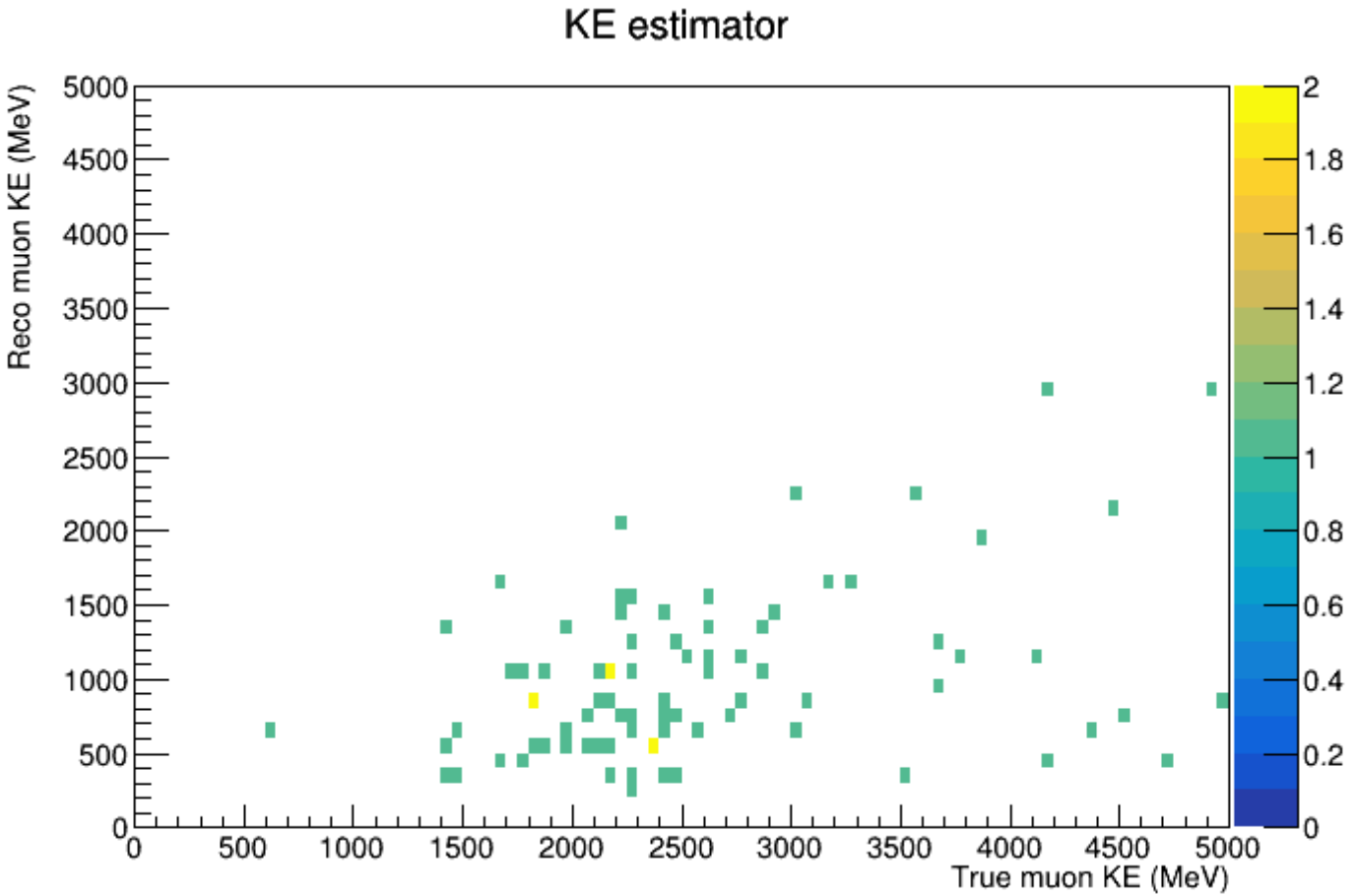
Eff numerator (red box)



$$\text{Eff} = \frac{\text{TMS-starting muons that have one or more reco track}}{\text{all TMS-starting muons}}$$

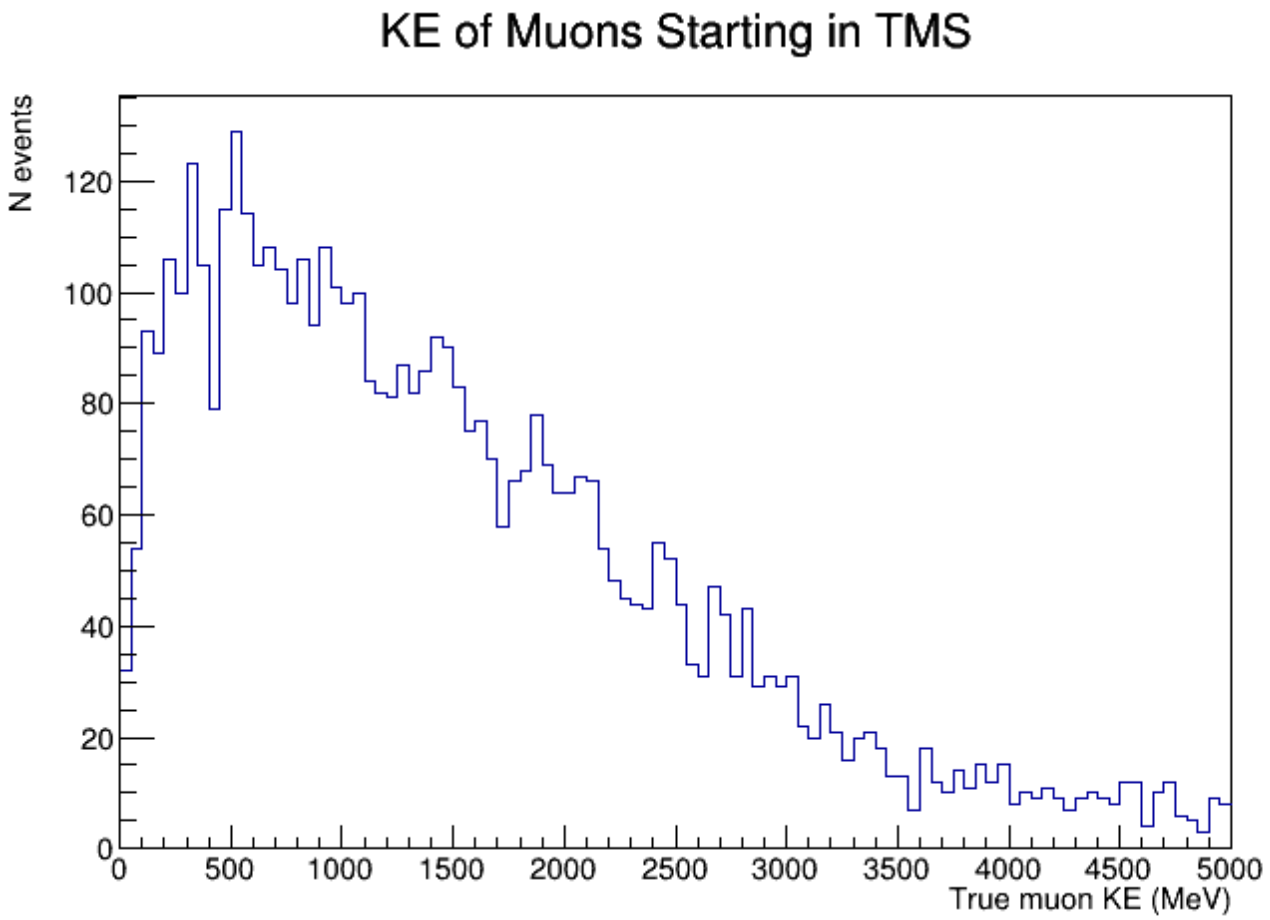


This is true KE vs areal density.
It's using all the muonke.cpp cuts so the number of events is small

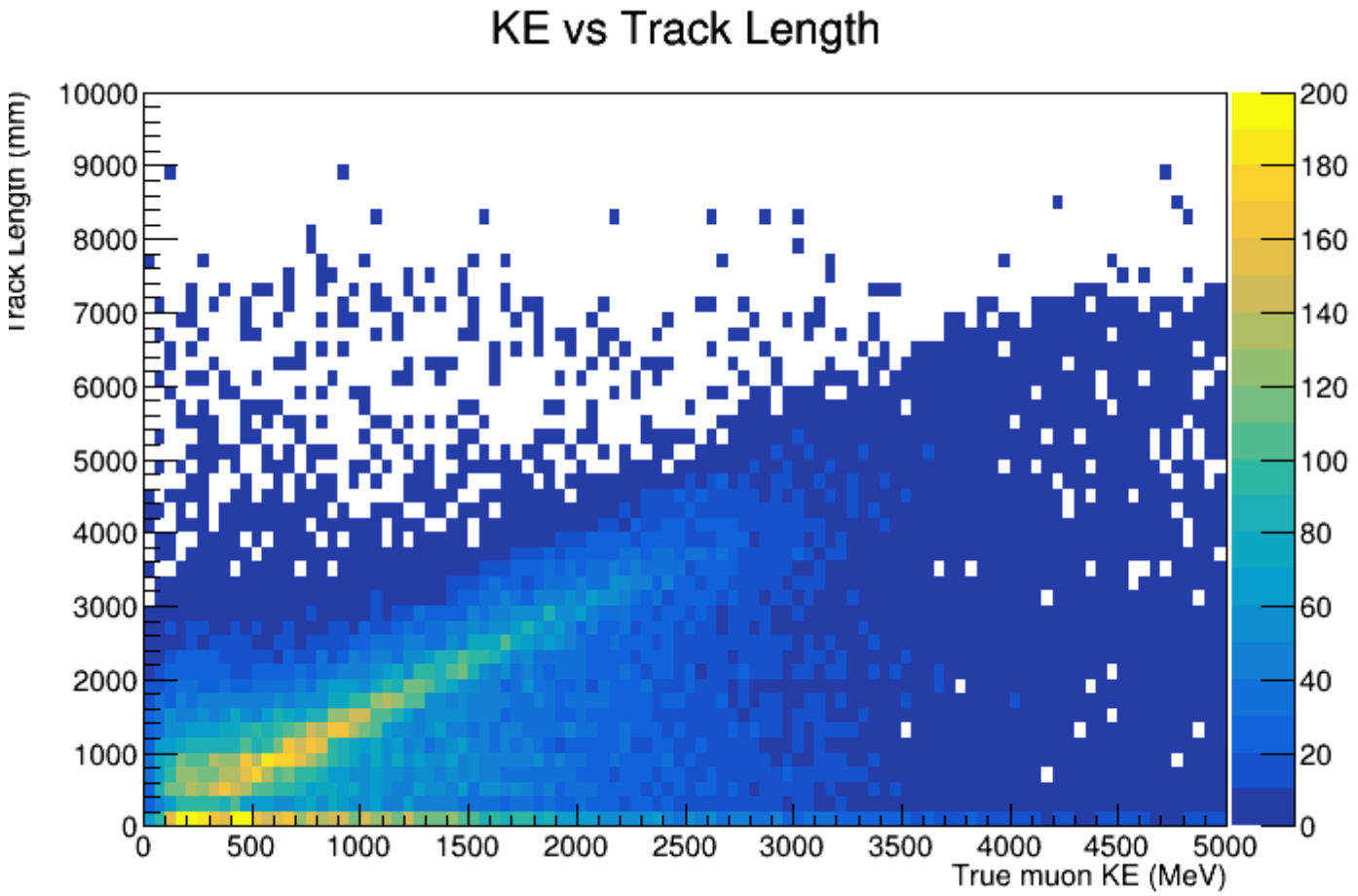


This is true KE vs 82+1.75*areal density, ie reco muon KE
It's using all the muonke.cpp cuts so the number of events is small

KE of muons that start in TMS

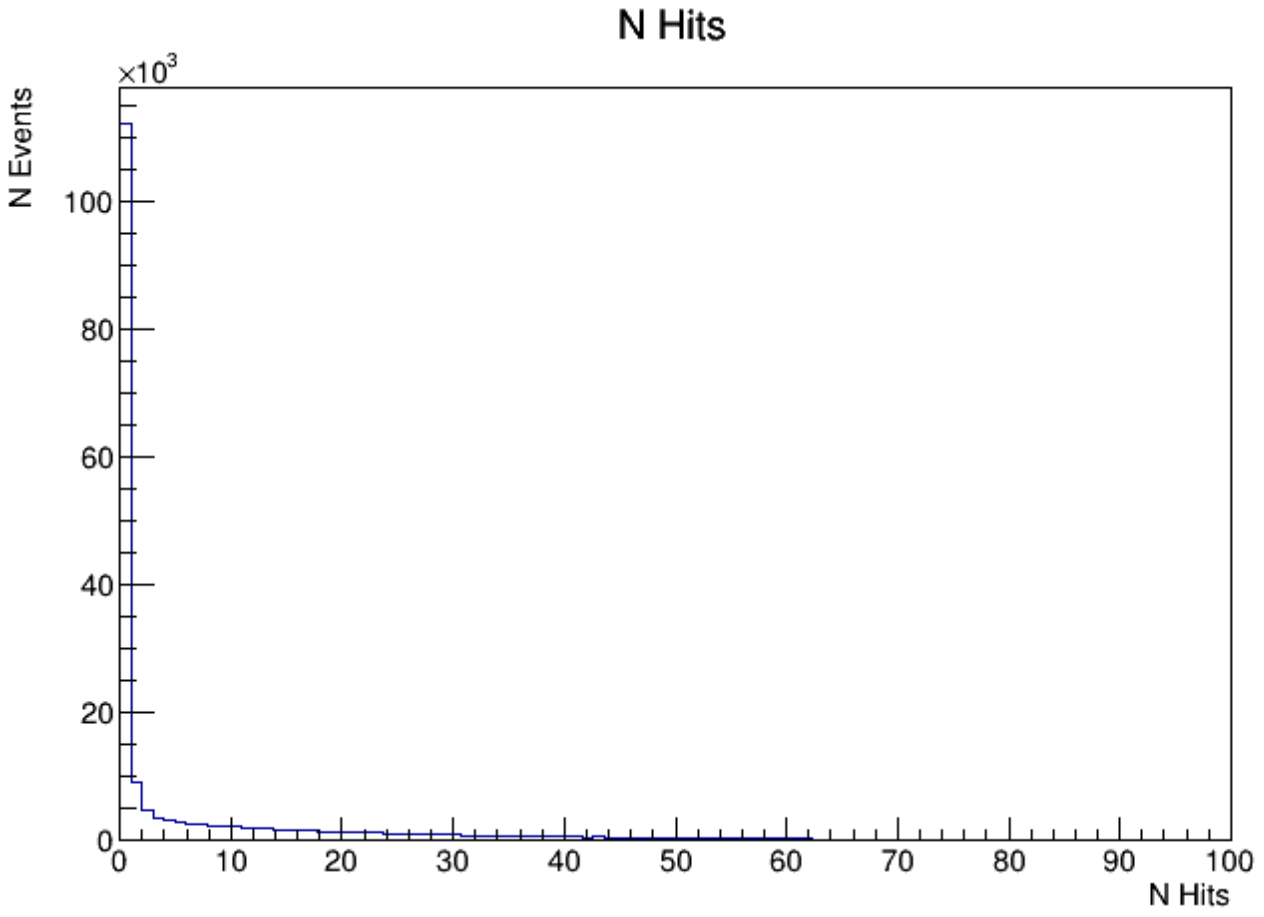


True KE vs reco track length

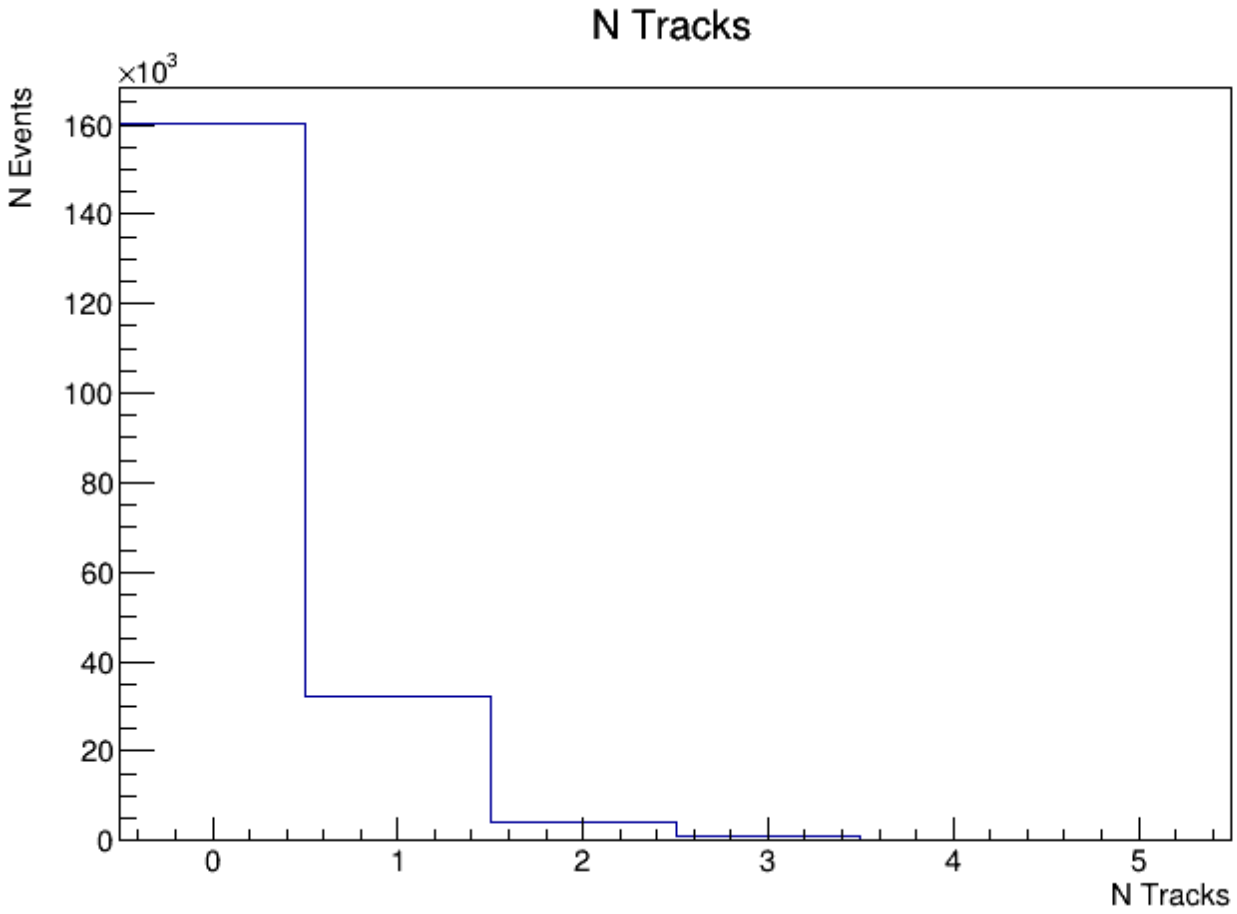


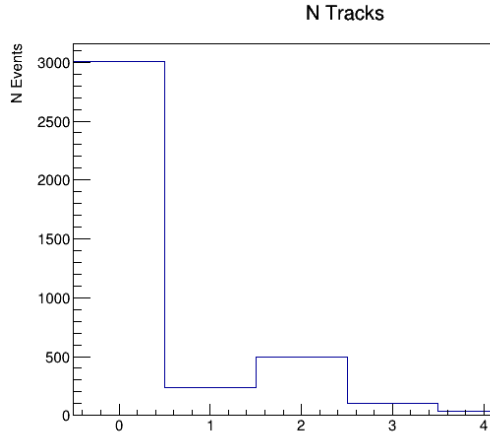
Track length = 0 if there are no tracks which explains the 0 band

Number of hits in an event

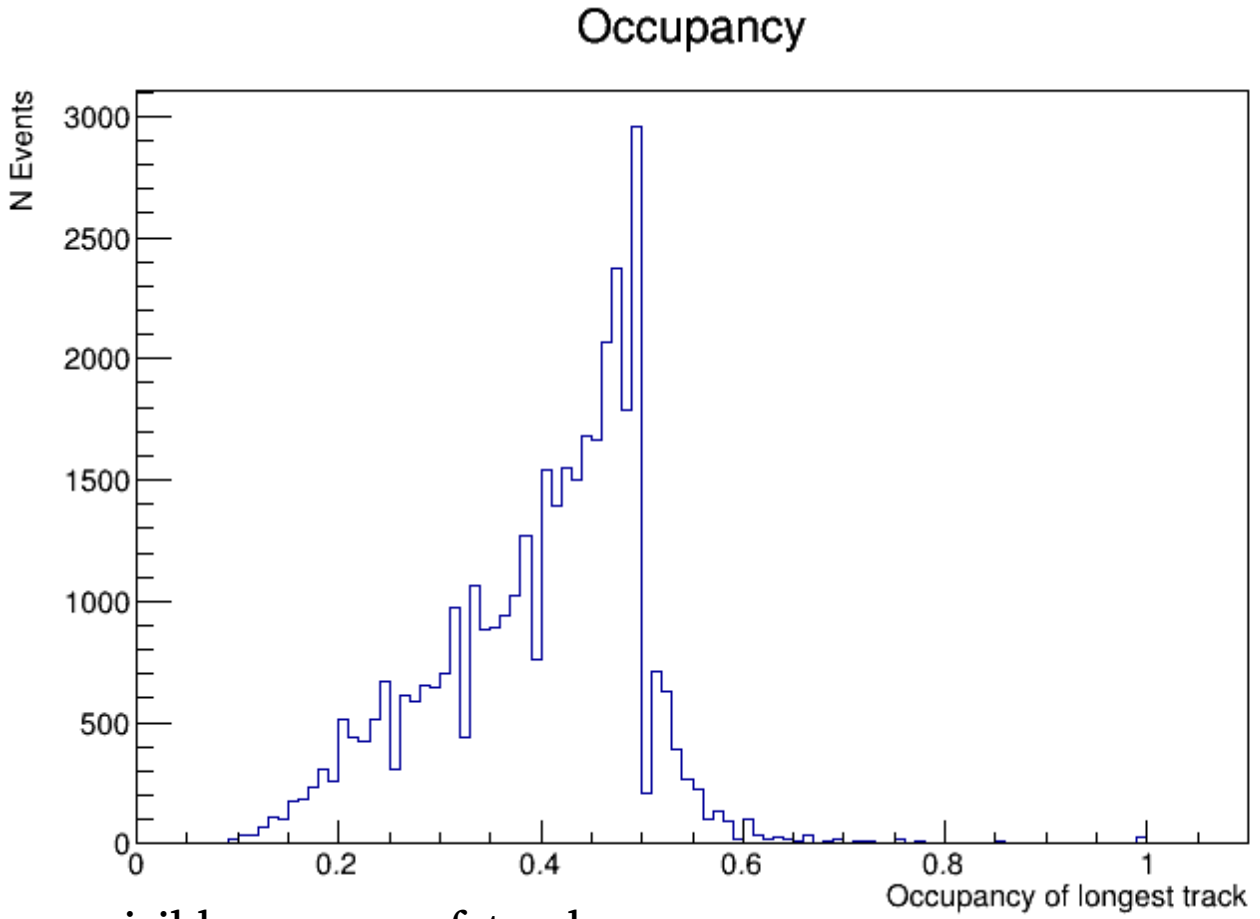


Number of reconstructed tracks in an event





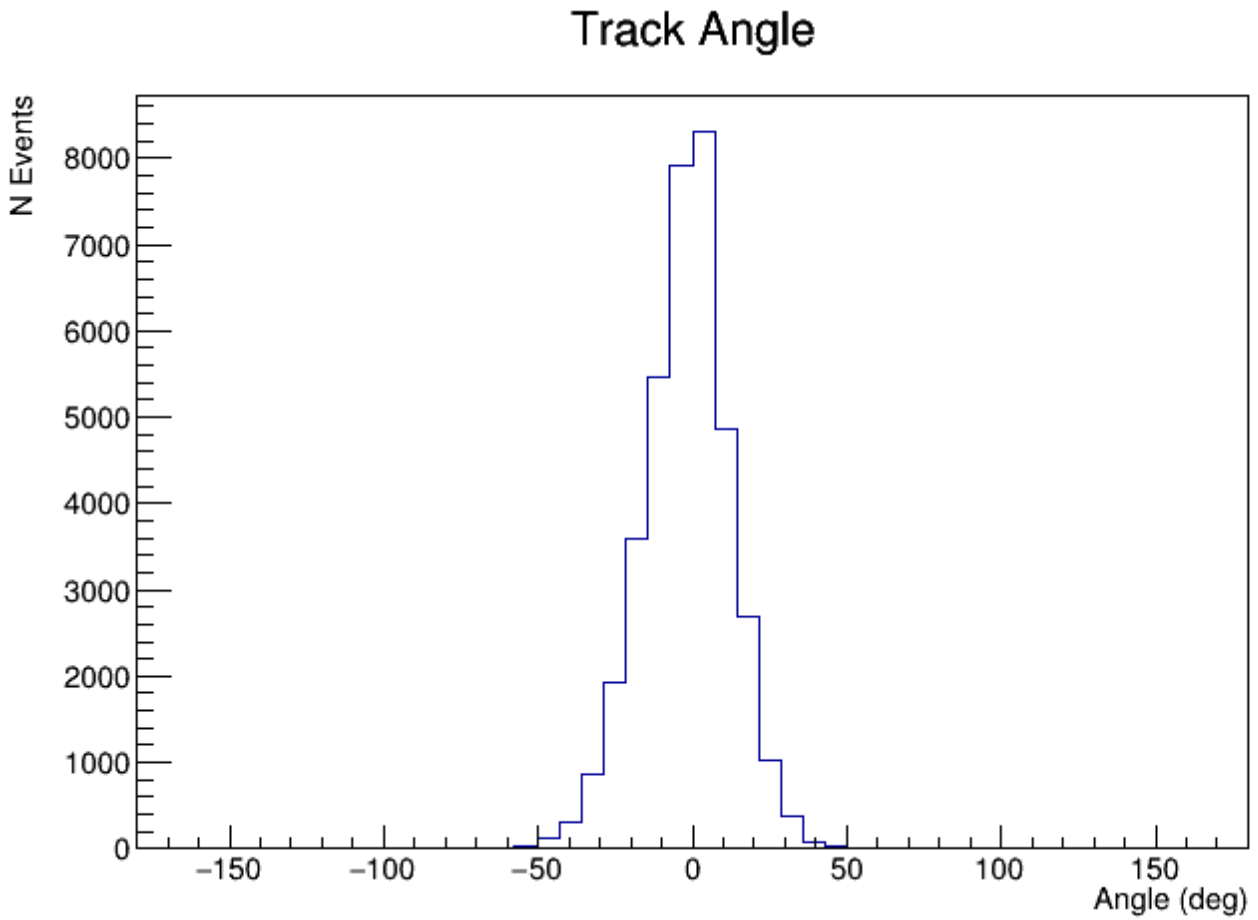
We expect n tracks to peak and zero and then have fewer 1 track events and then even fewer 2 track events.
If there's a peak at 2 tracks, then it's an indication that you're using not using the density fix. See that →



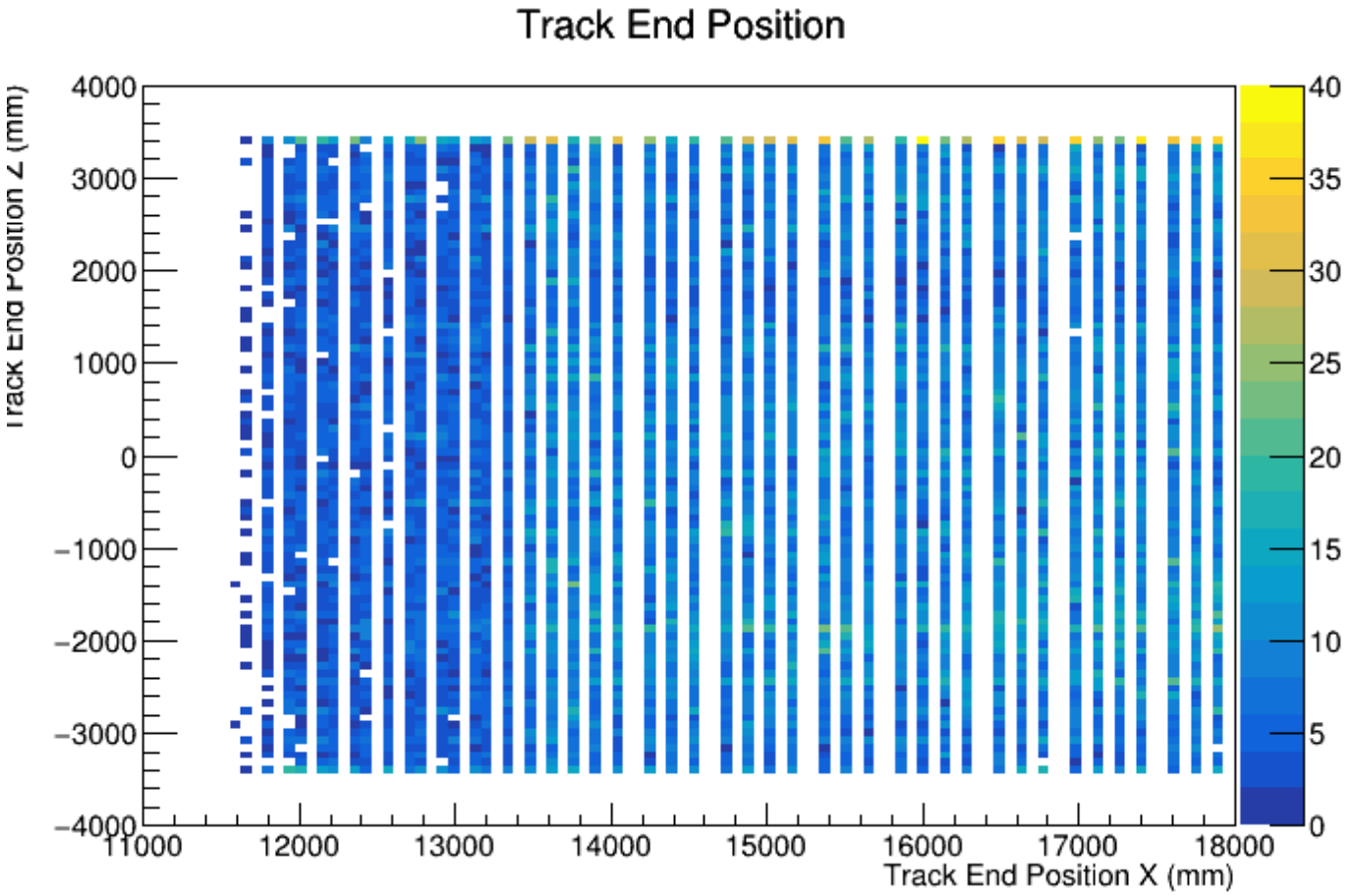
$$\text{Occupancy} = \frac{\text{sum visible energy of track}}{\text{sum visible energy}}$$

Note that now track uses only y planes so you're getting ~50% as much track energy, but no change in total visible energy in slice

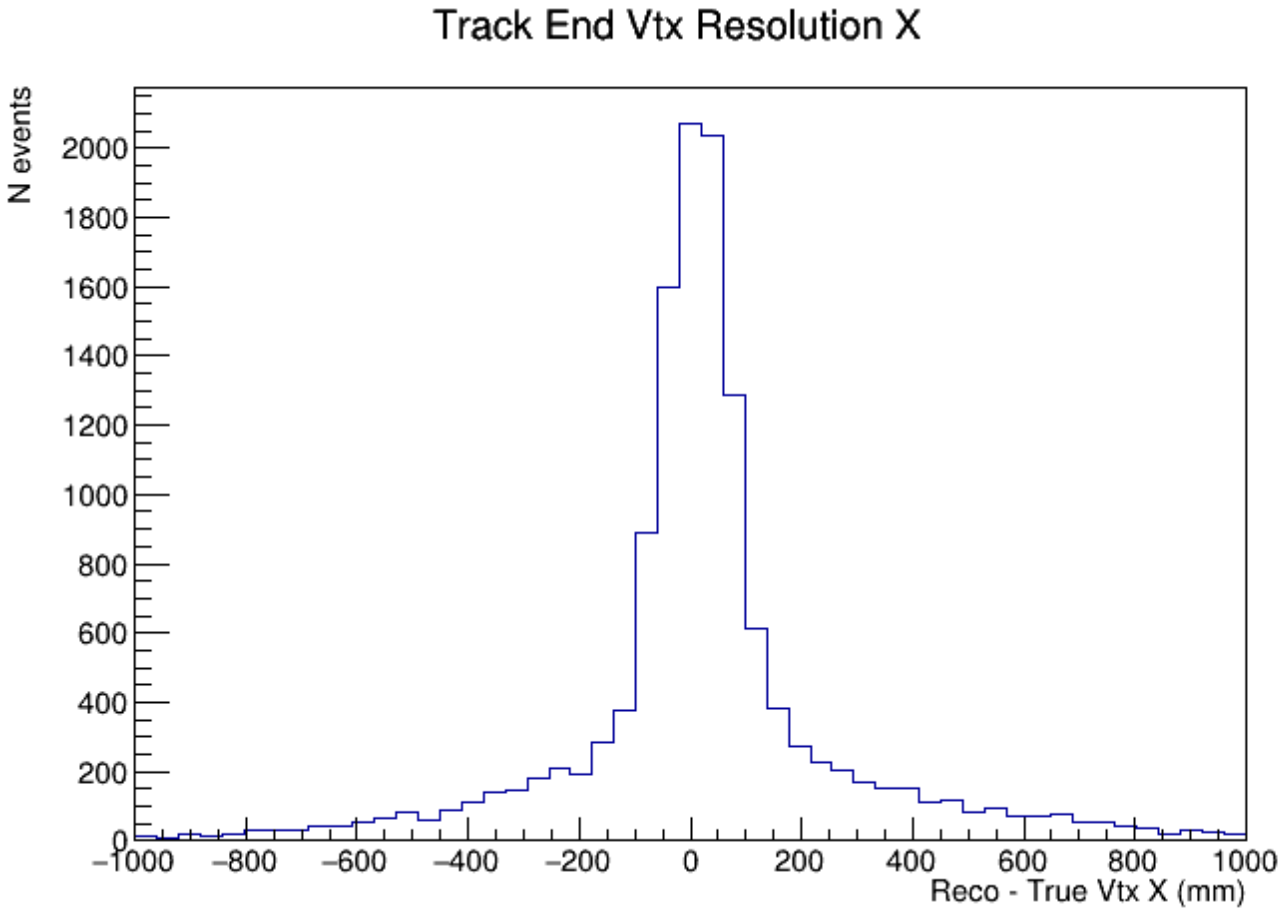
Reco track angle



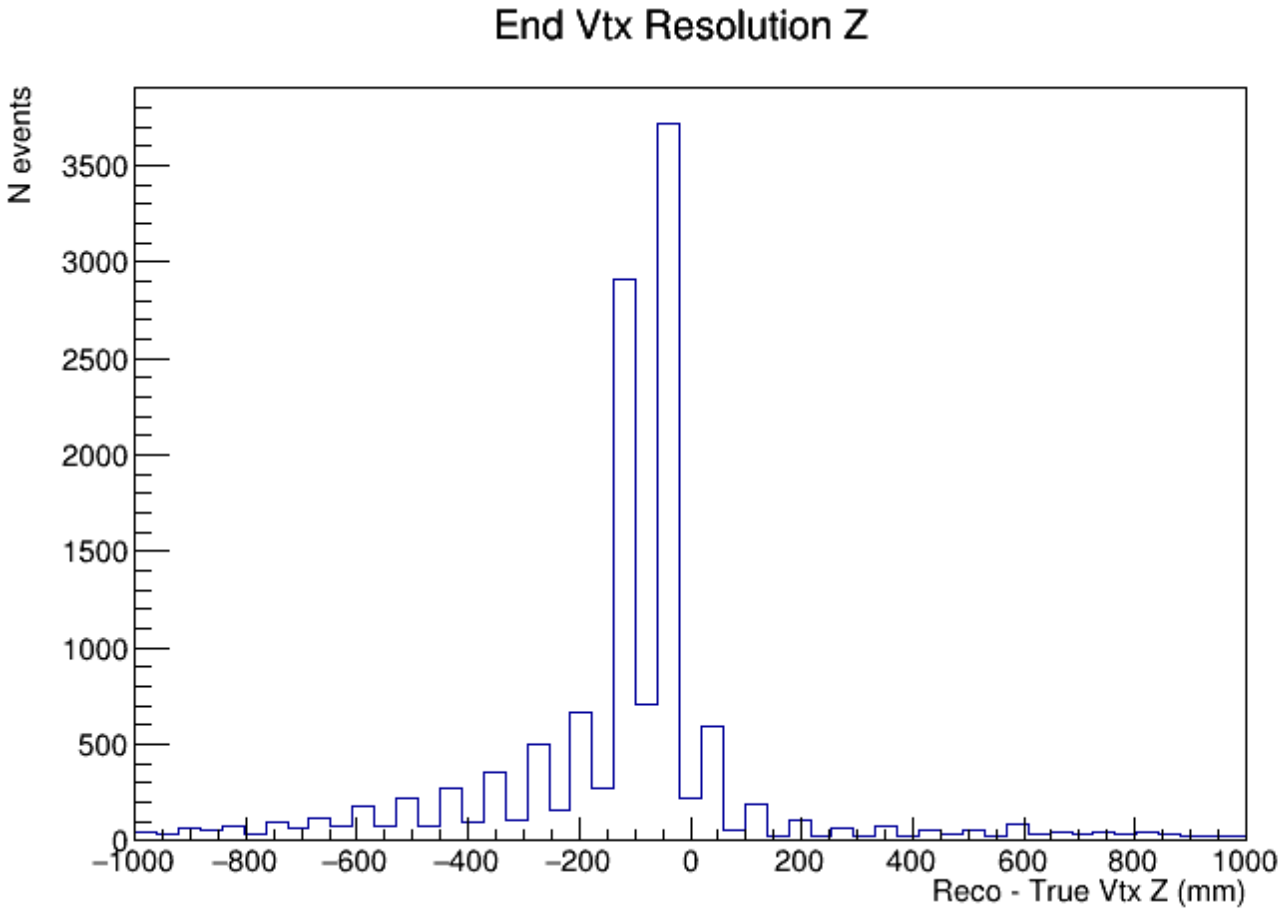
Reco track end position x vs z



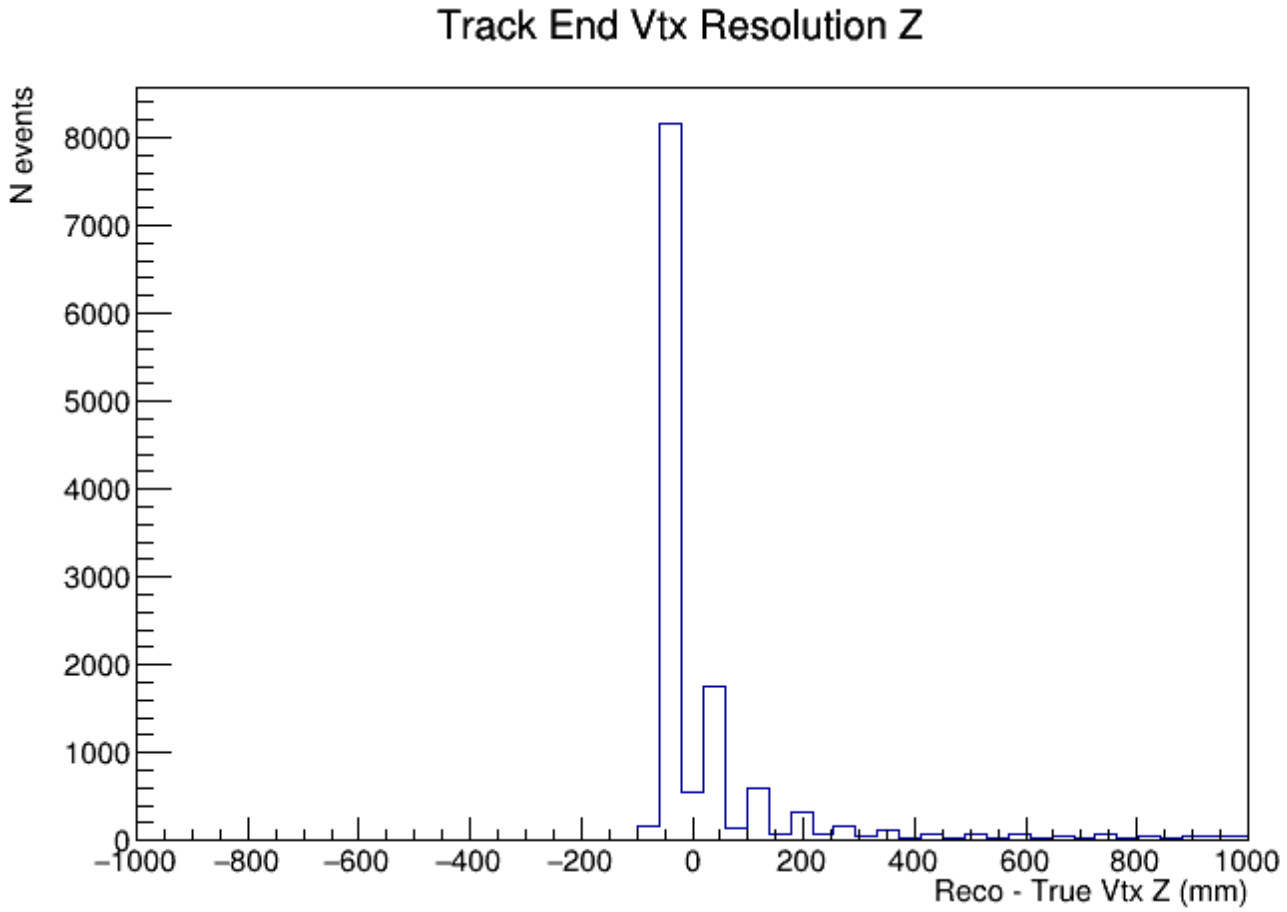
Track end position reco – true X



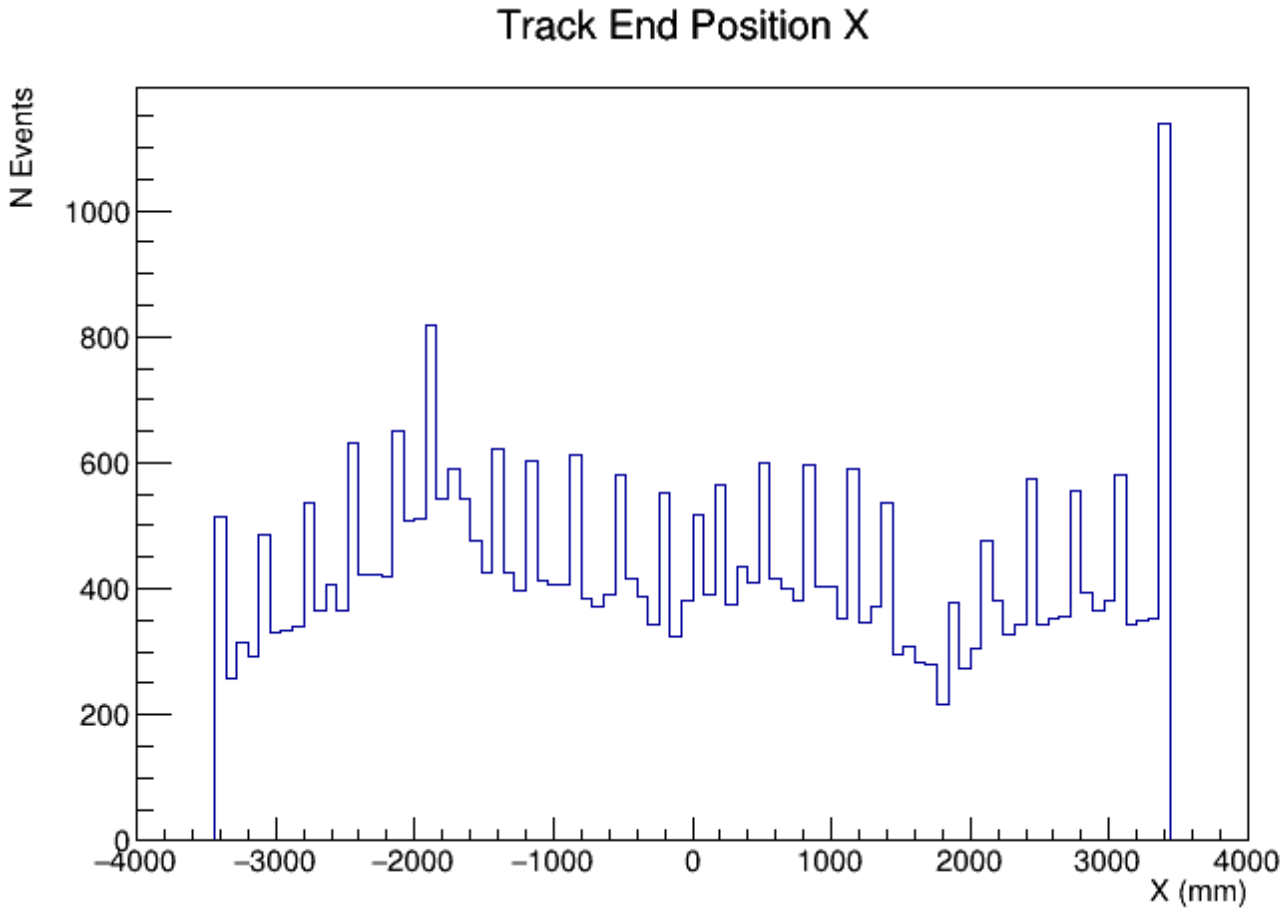
Track end position reco – true z



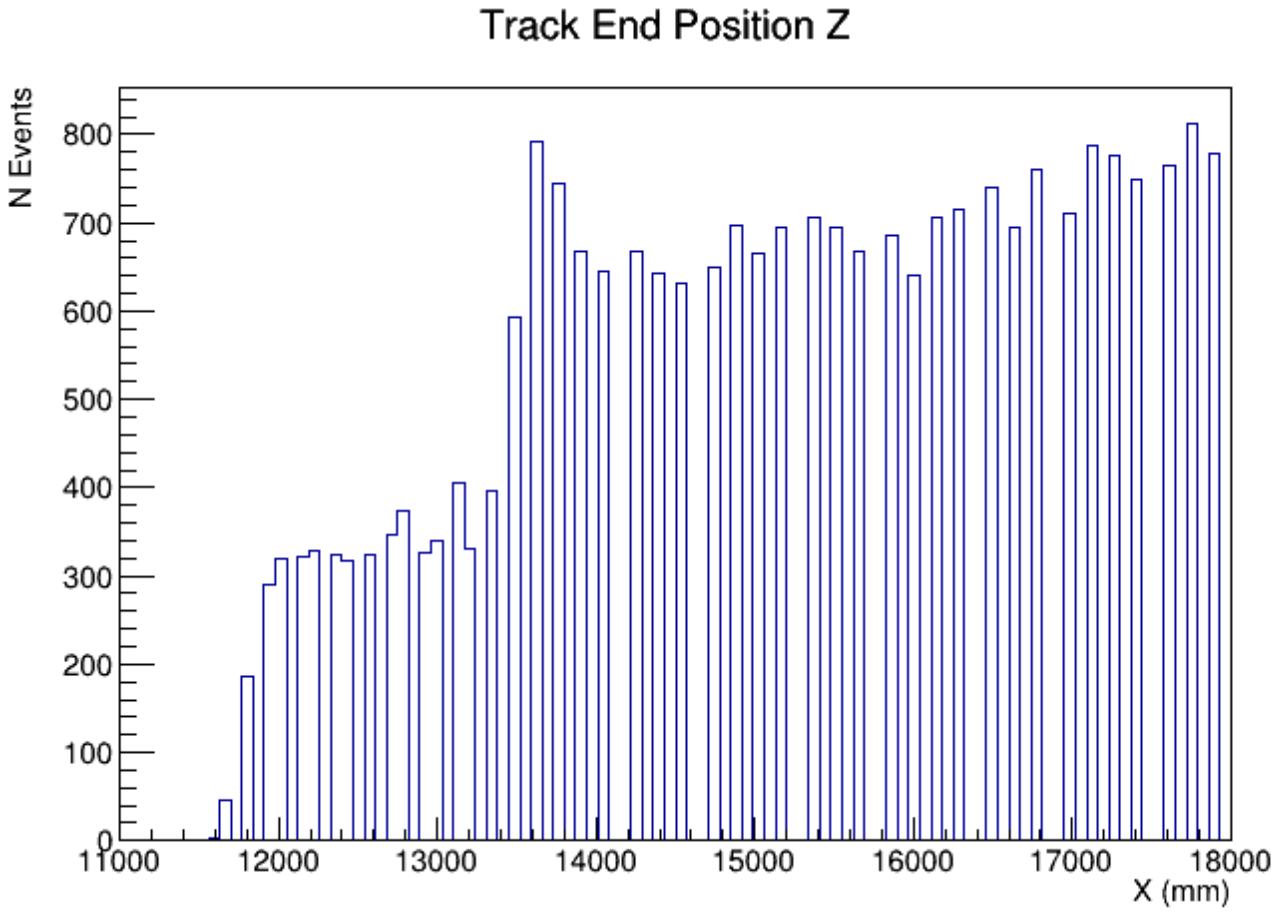
Same but using max dz proxy (can ignore)



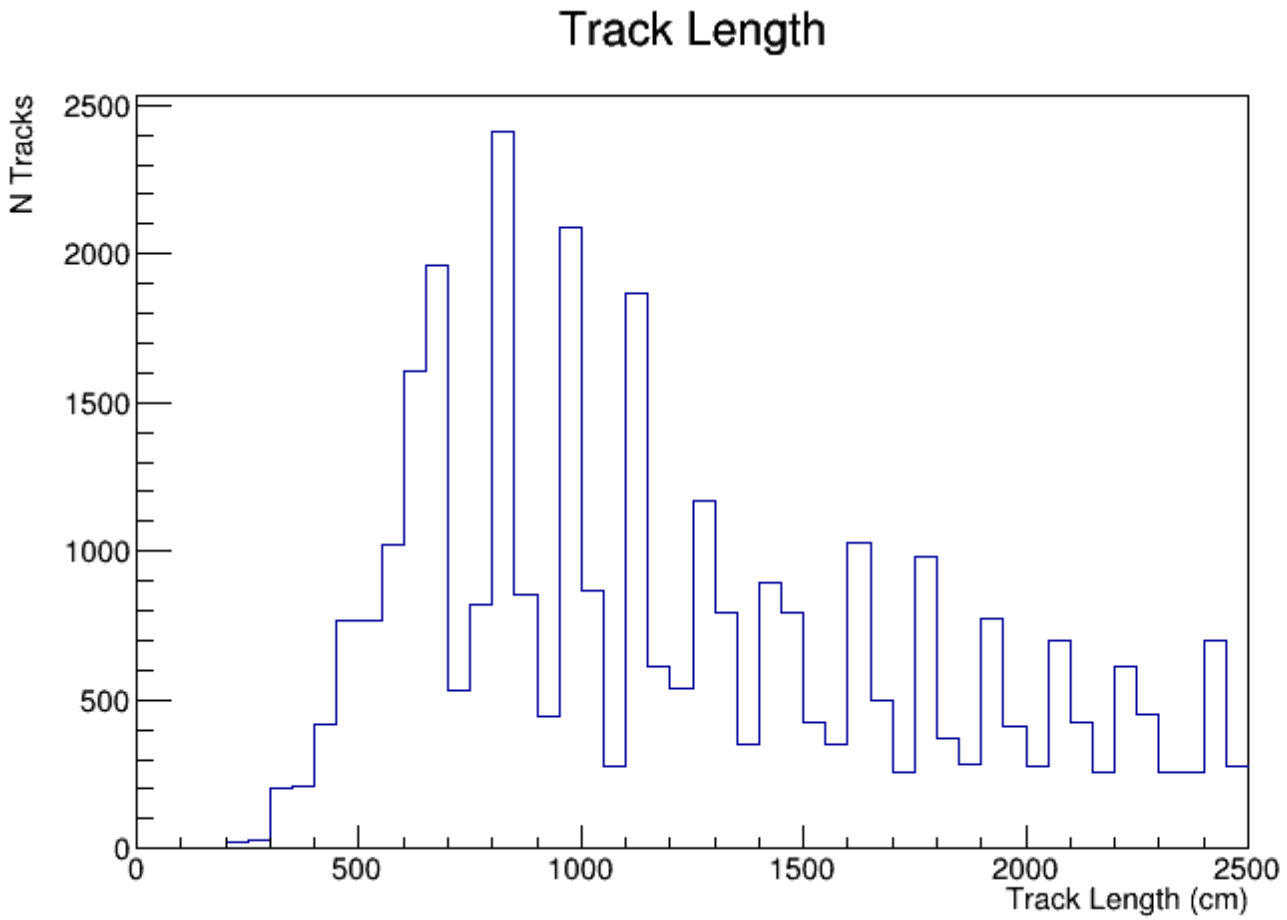
Track end position X



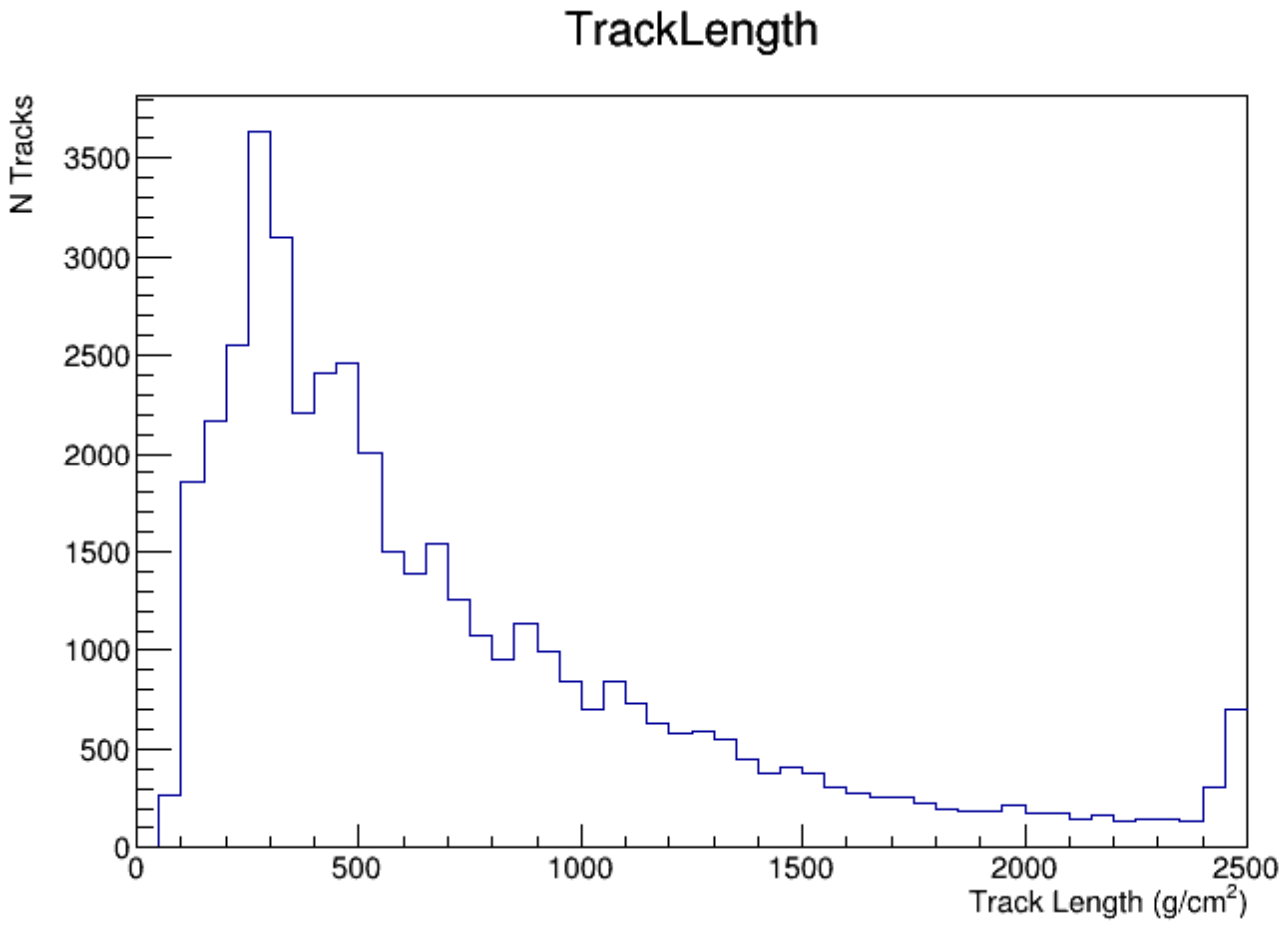
Track end position Z

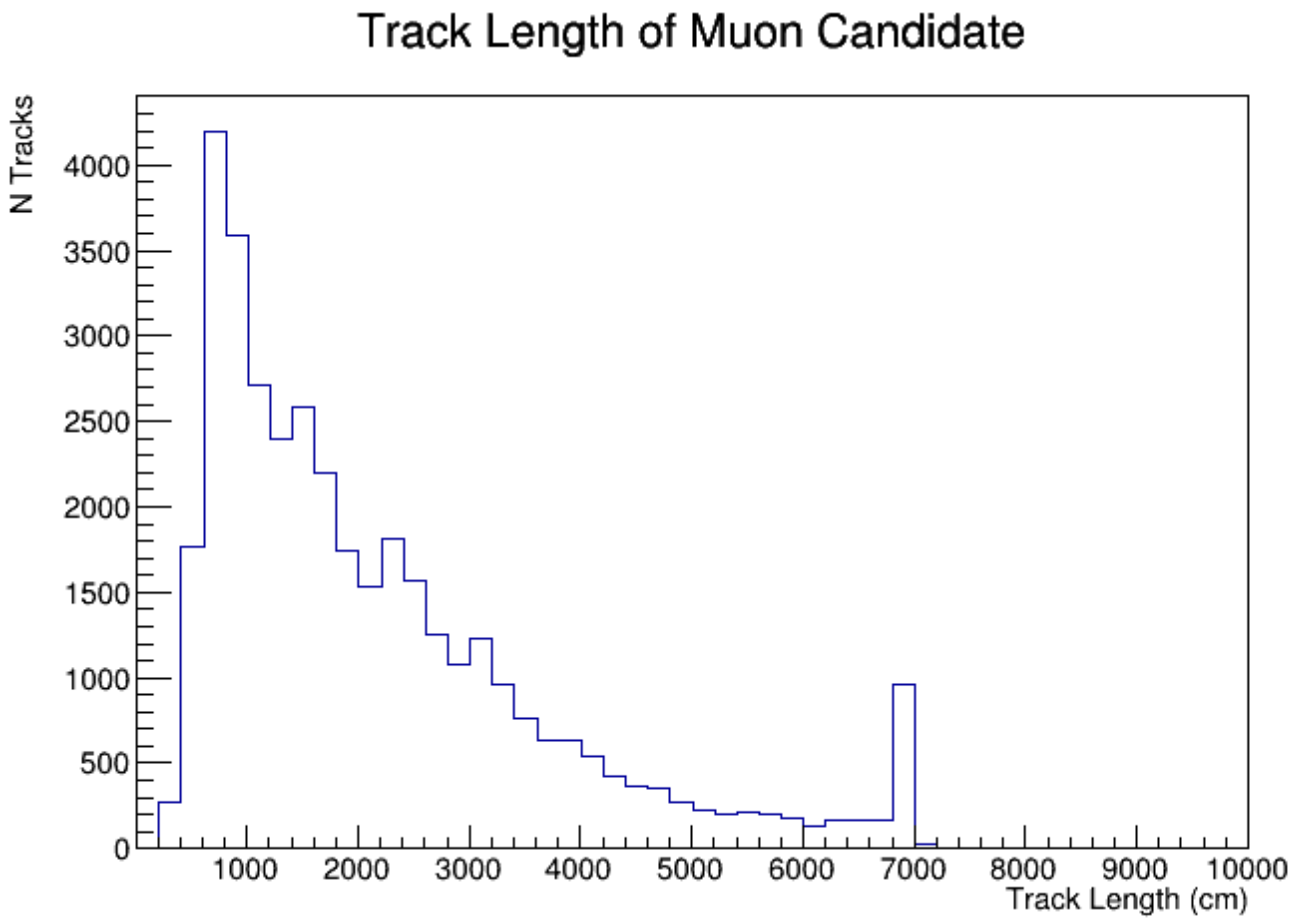


Track length

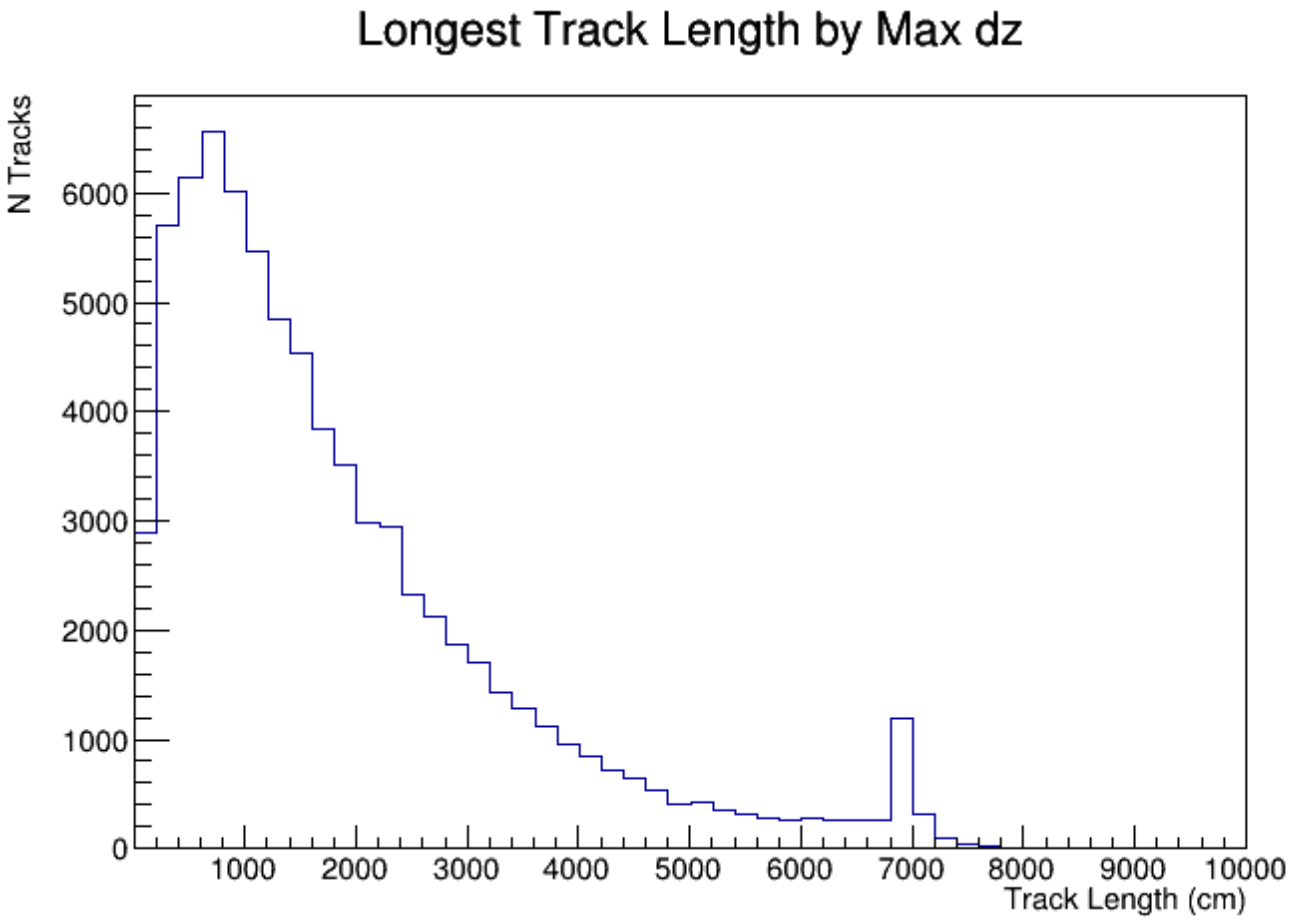


Var called TrackLength is areal density

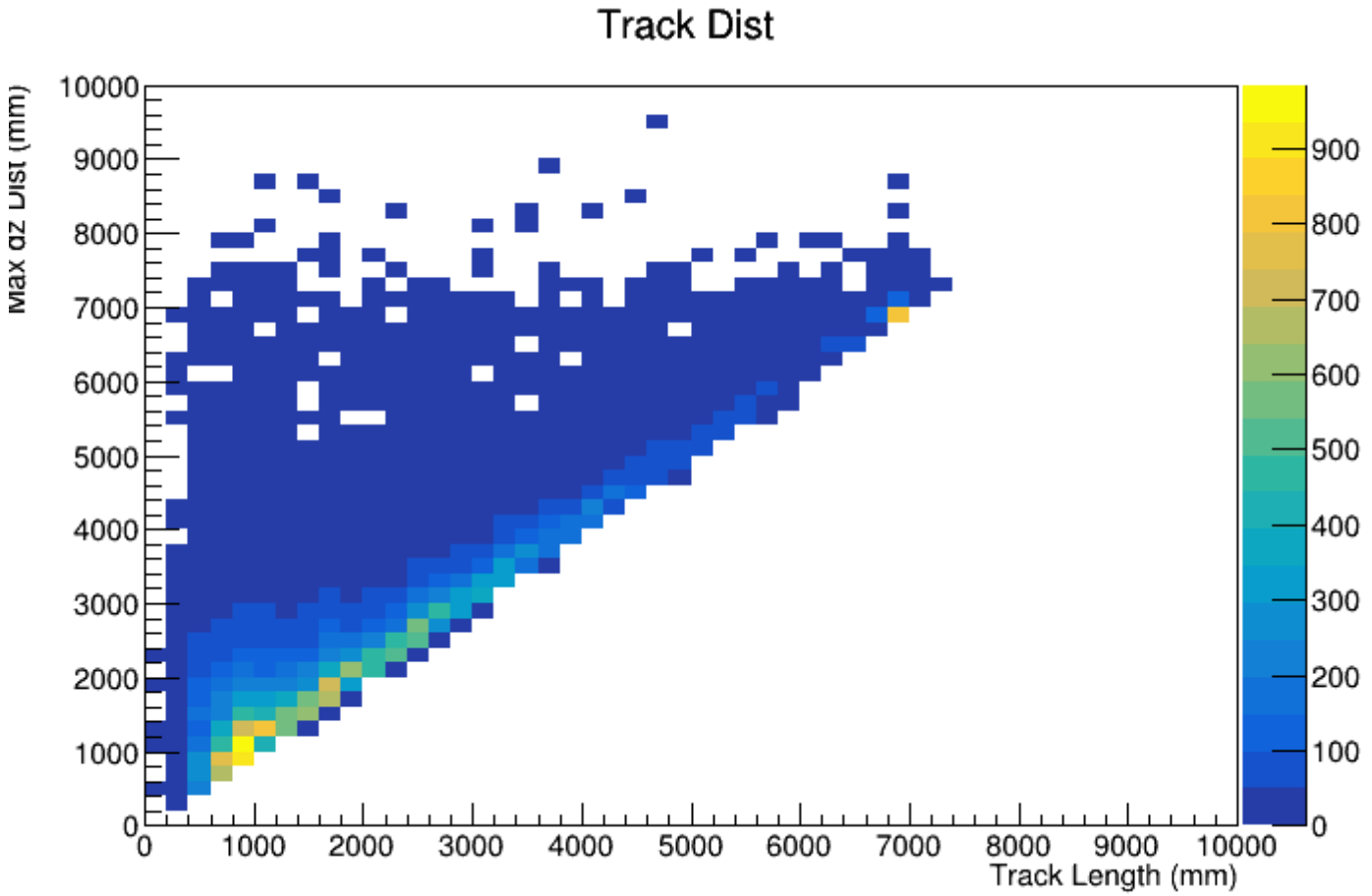


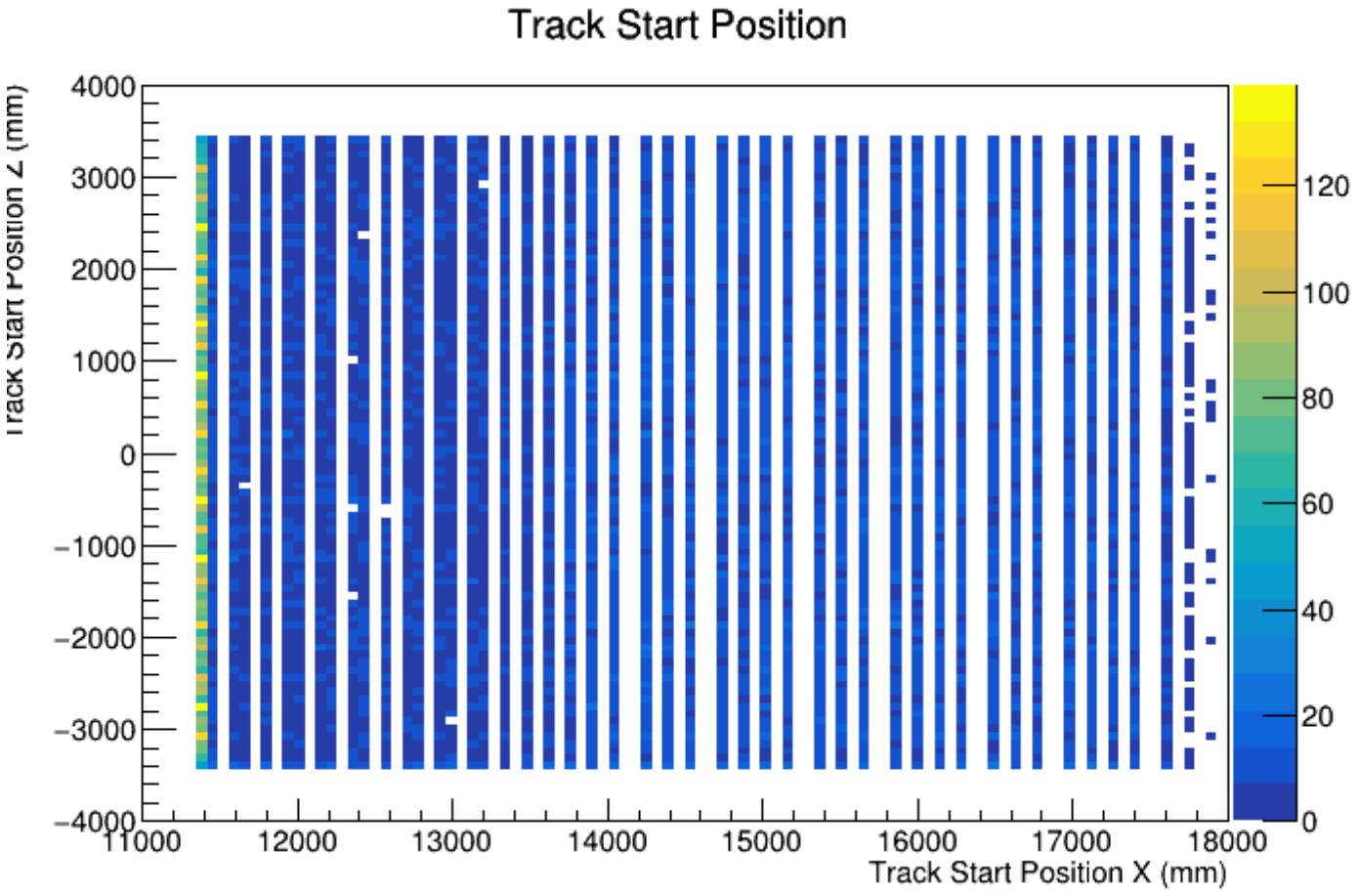

Track Length of Muon Candidate
N Tracks
4000
3500
3000
2500
2000
1500
1000
500
0
1000
2000
3000
4000
5000
6000
7000
8000
9000
10000
Track Length (cm)

Can ignore

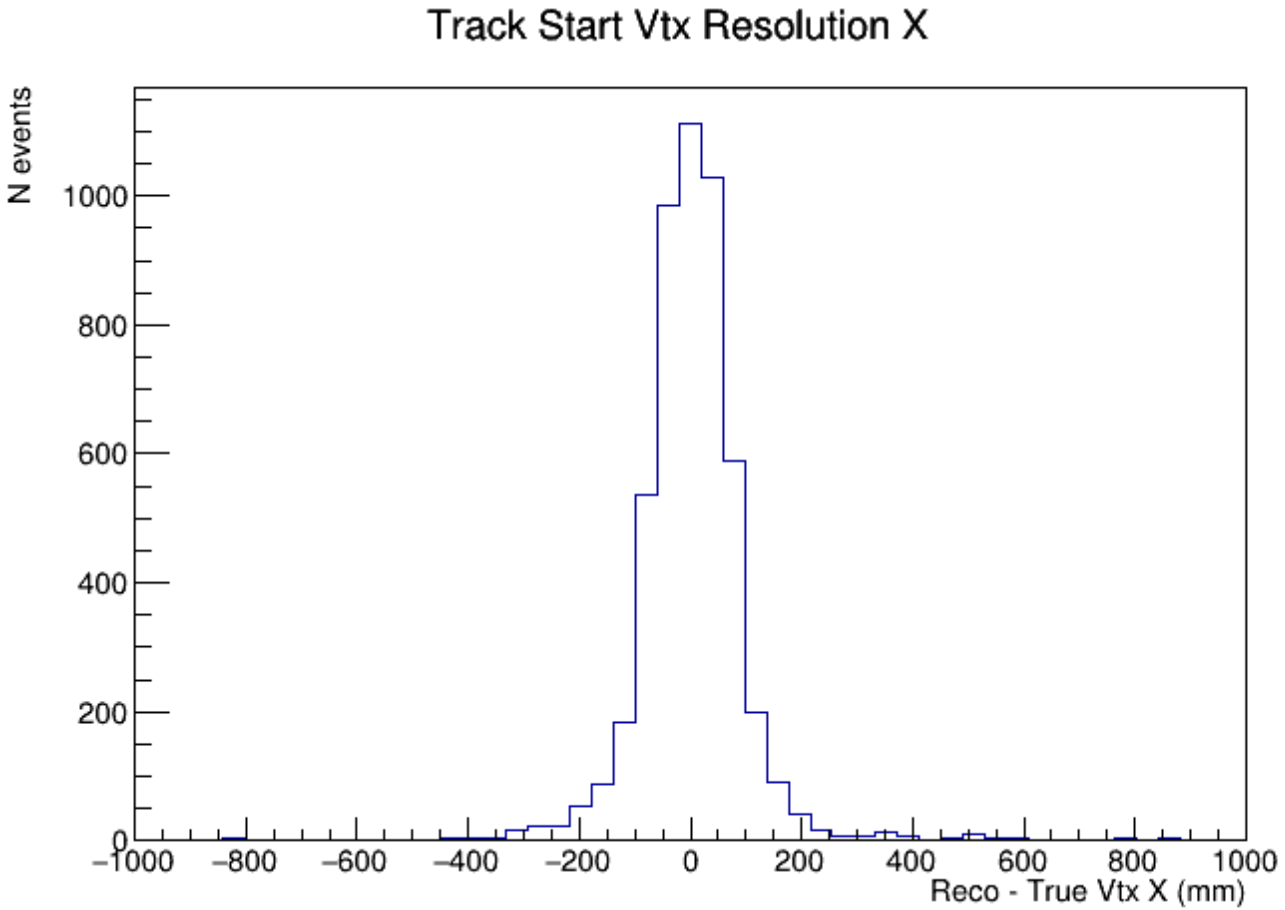

Longest Track Length by Max dz
N Tracks
0
1000
2000
3000
4000
5000
6000
1000
2000
3000
4000
5000
6000
7000
8000
9000
10000
Track Length (cm)

track length vs max dz, Can ignore

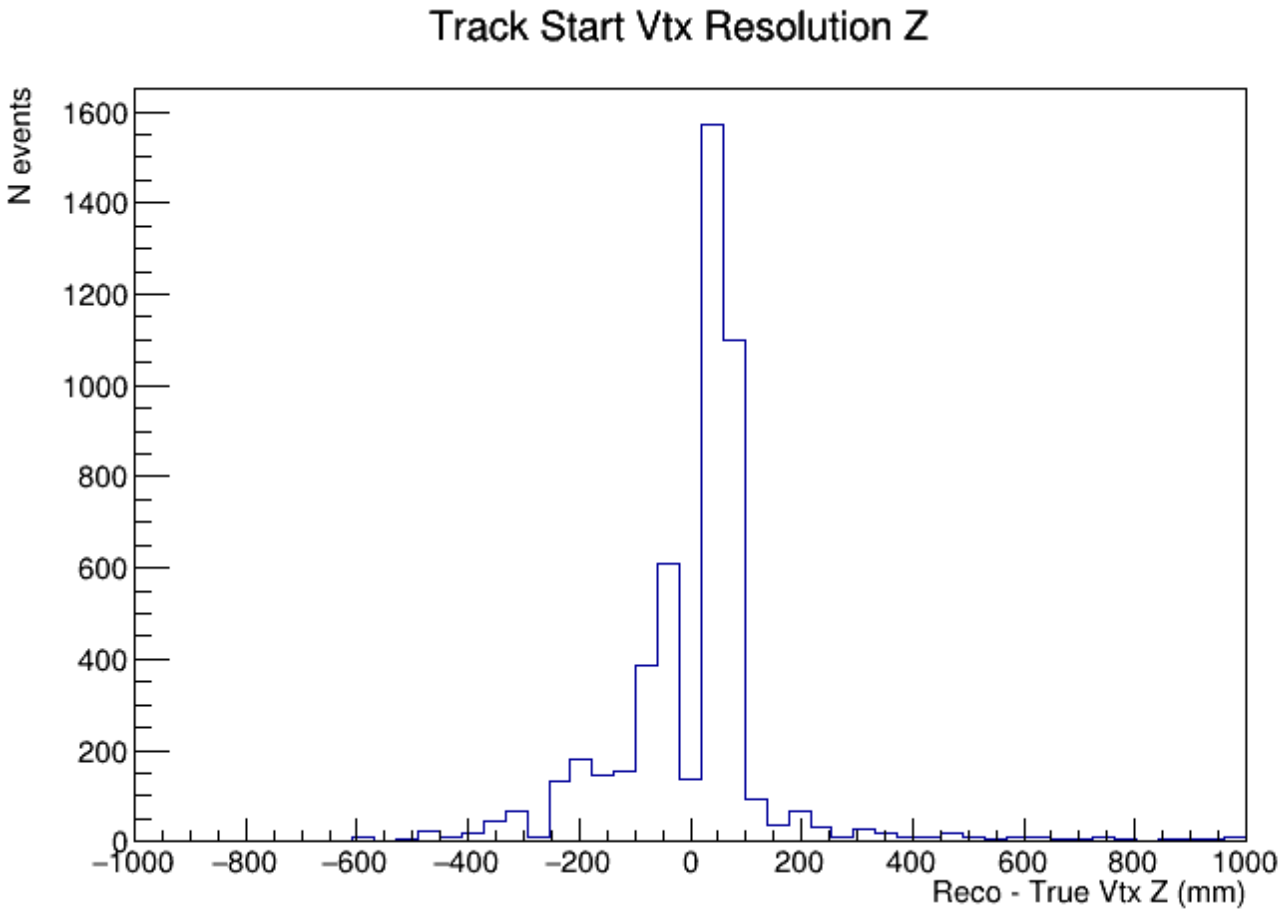



Track Start Position
Track Start Position Z (mm)
4000
3000
2000
1000
0
−1000
−2000
−3000
−4000
11000
12000
13000
14000
15000
16000
17000
18000
Track Start Position X (mm)
120
100
80
60
40
20
0

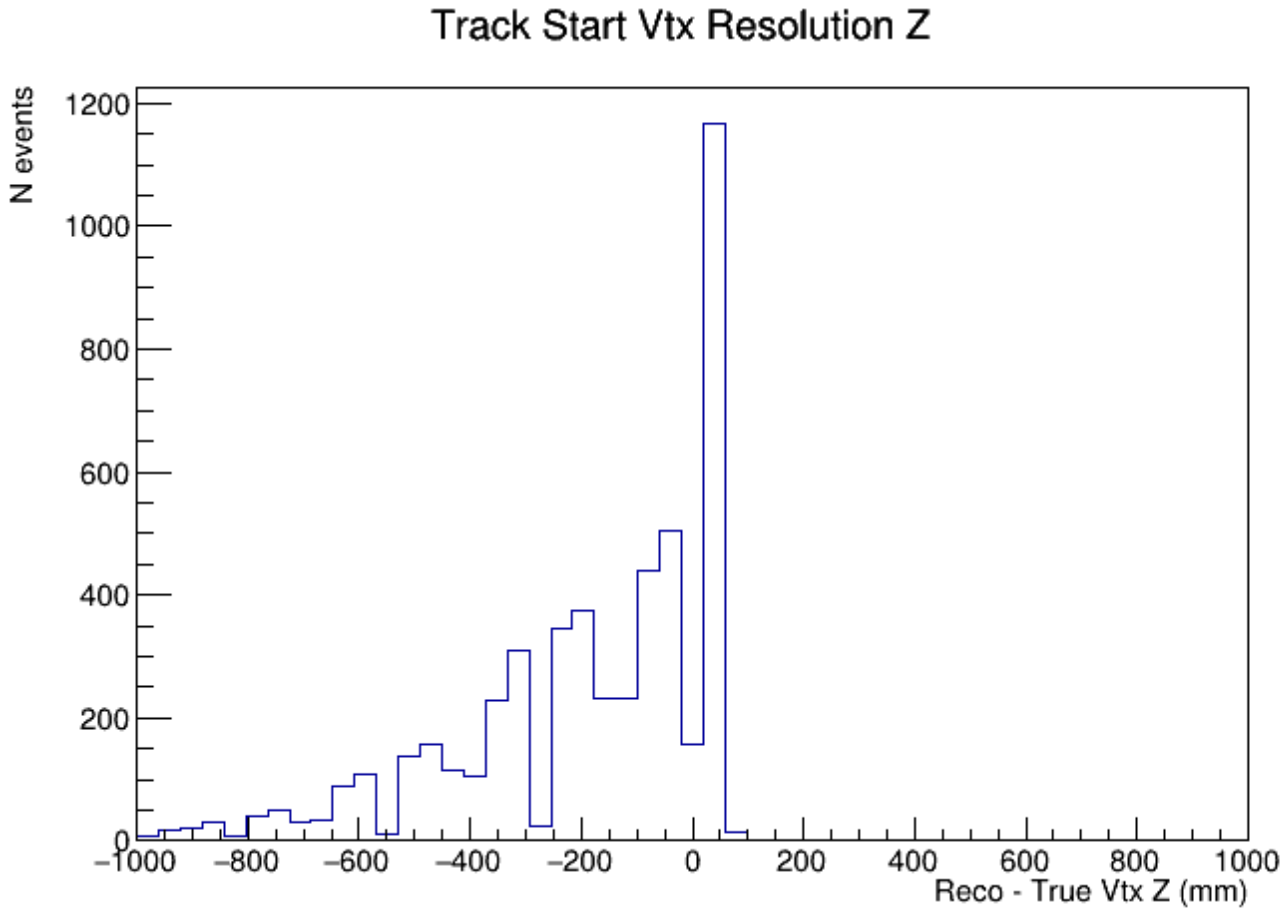
Resolution X

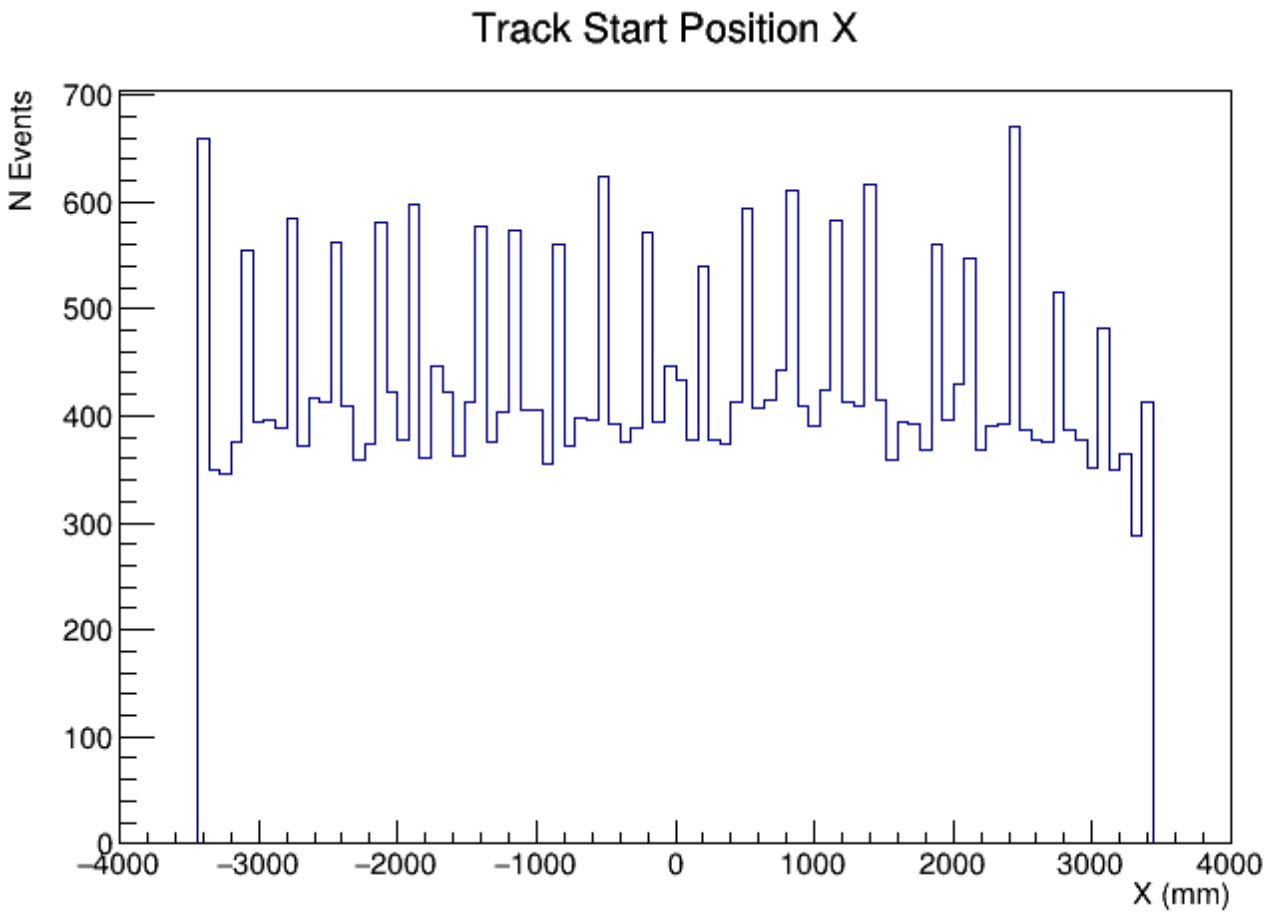


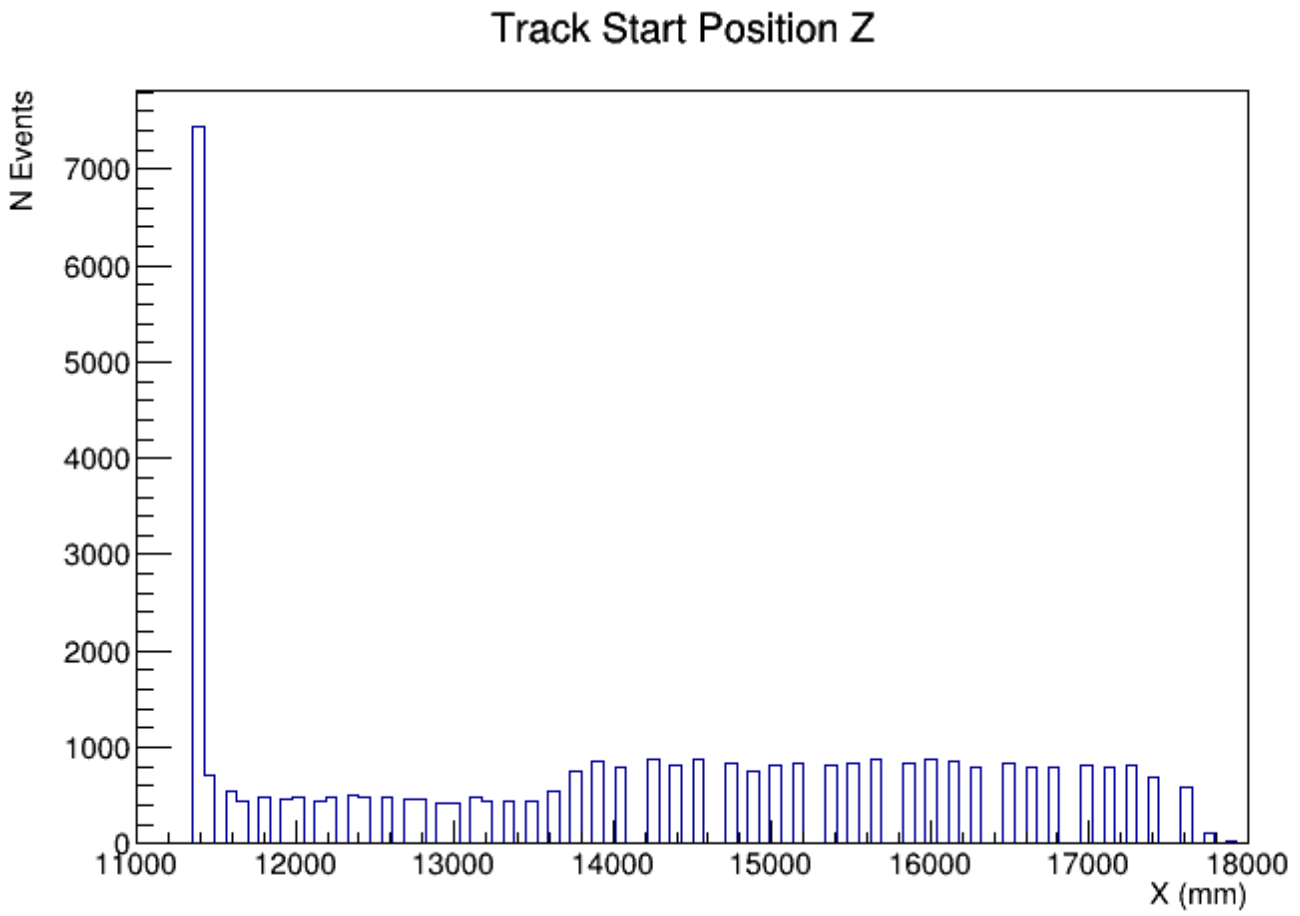
Resolution Z



Resolution using max dz, can ignore

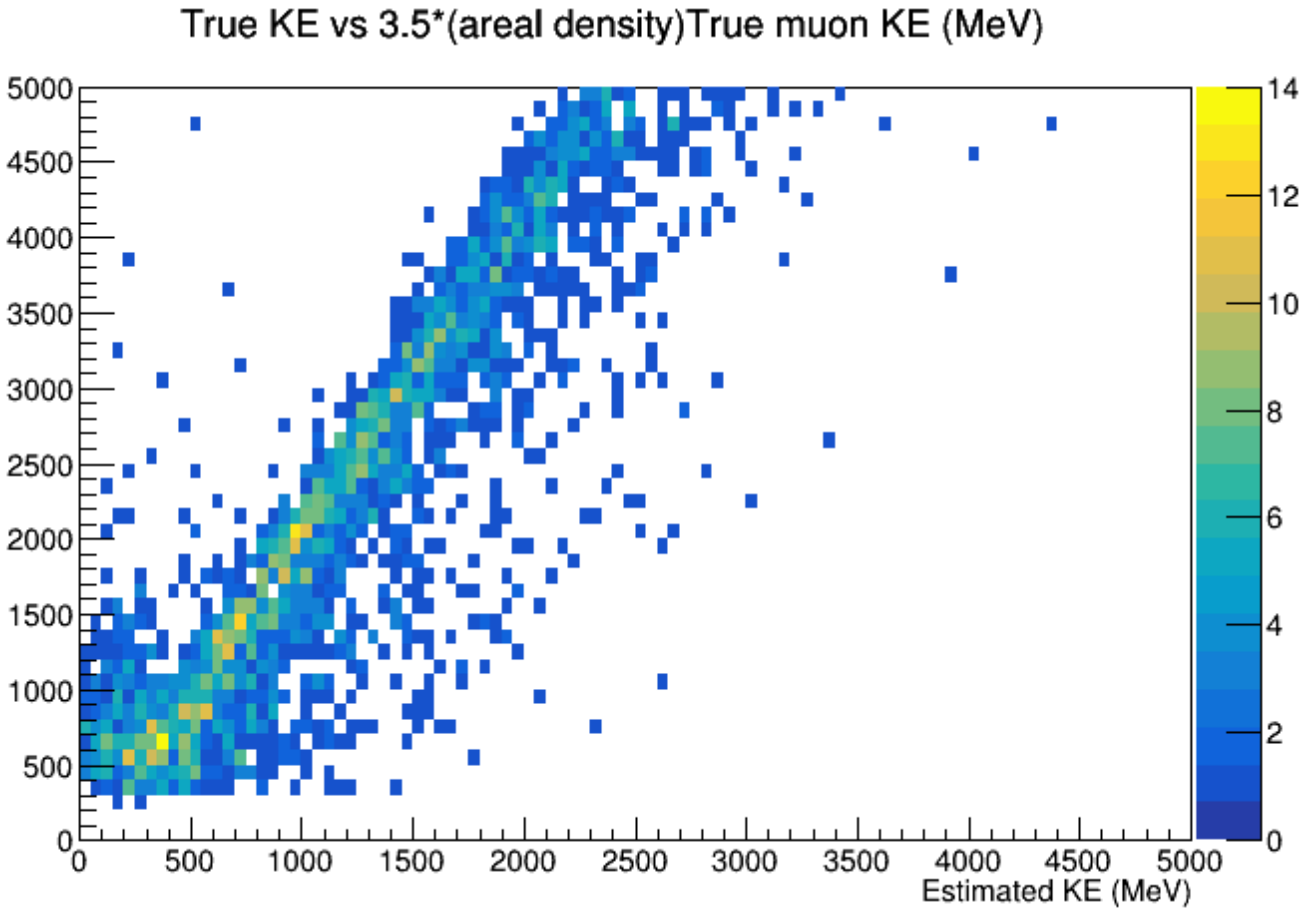



Track Start Position X
N Events
700
600
500
400
300
200
100
0
−4000
−3000
−2000
−1000
0
1000
2000
3000
4000
X (mm)

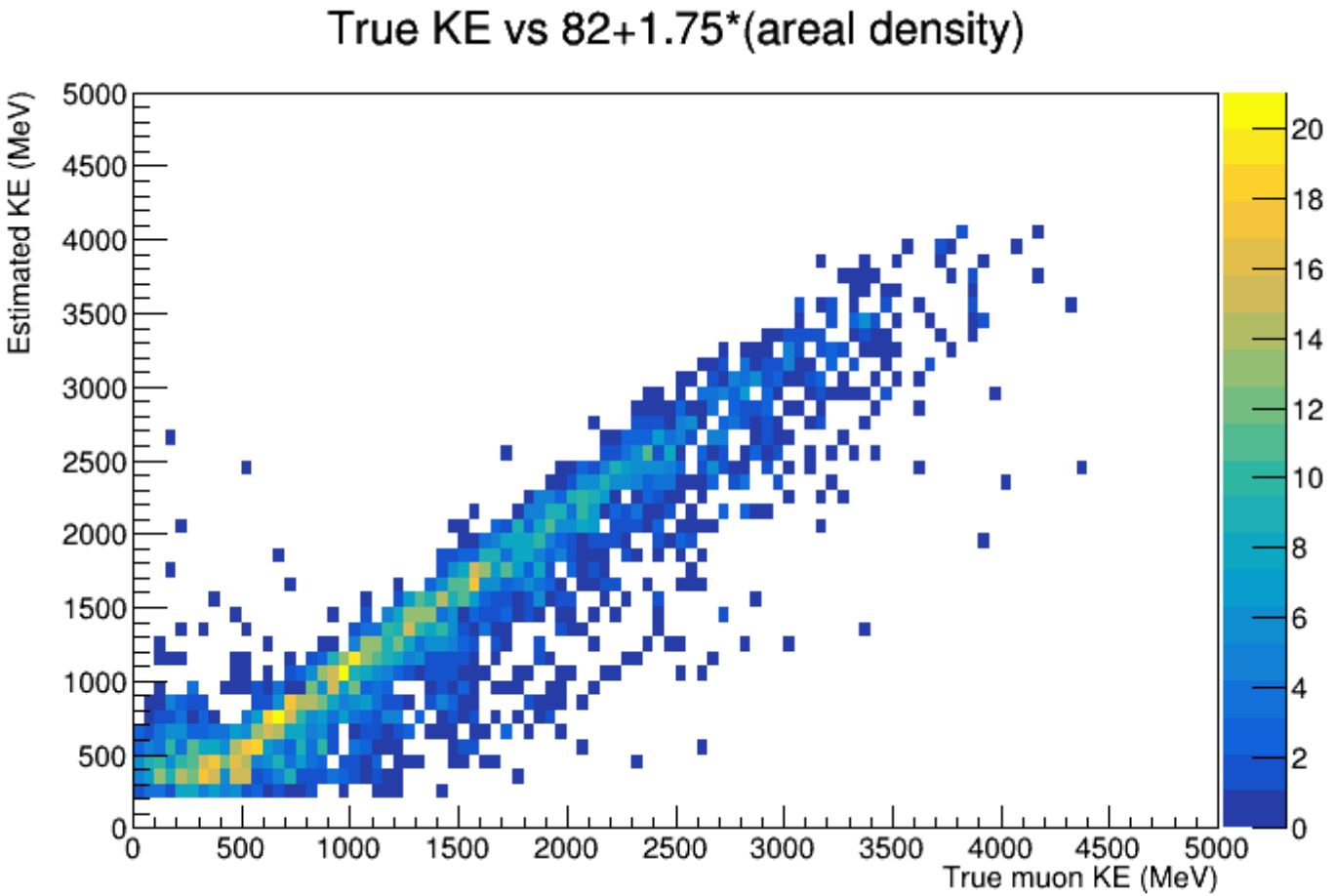


Peak early is muons and other particles entering from LAr detector. The reco track starts at the front of the detector but the actual muon started elsewhere

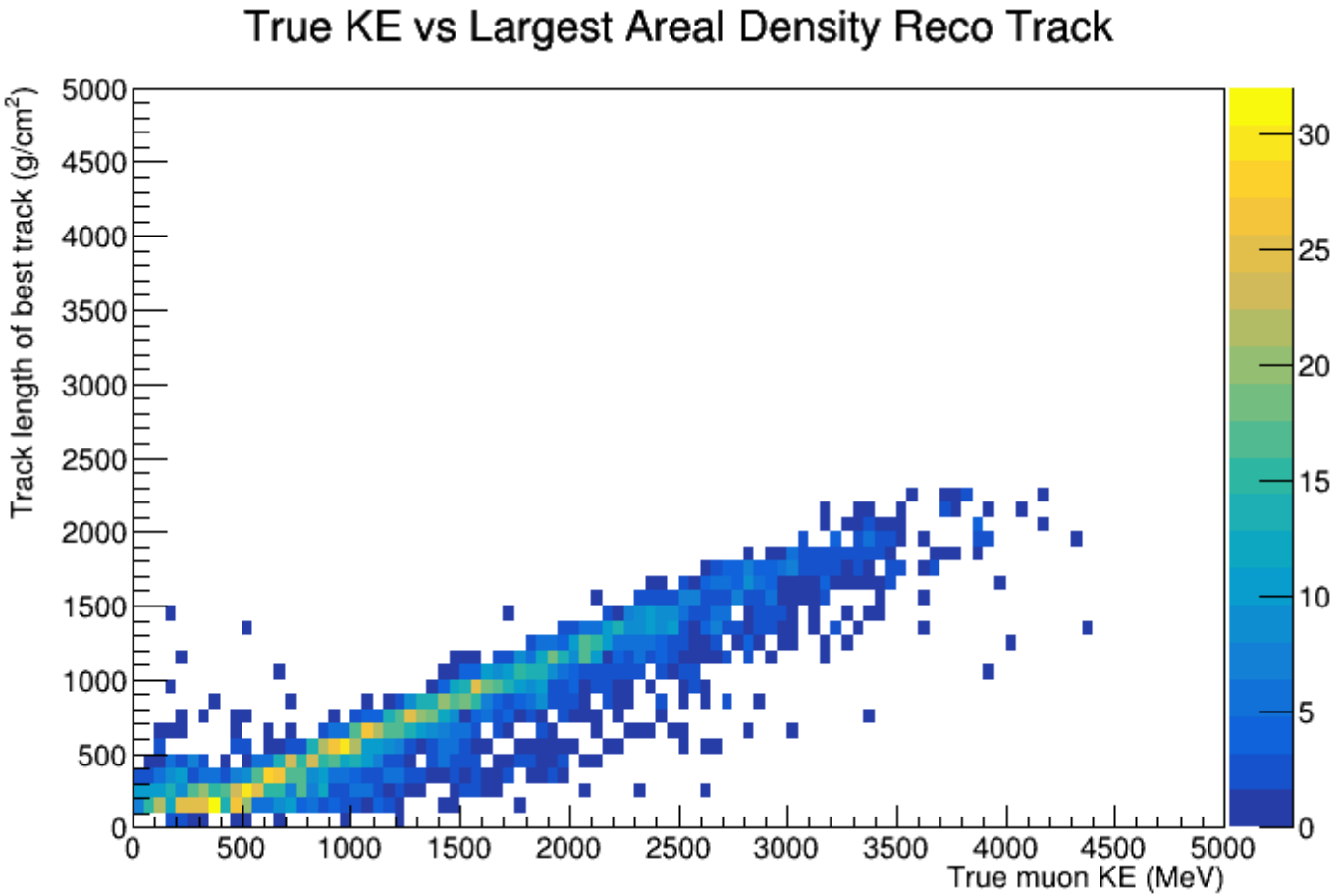
Can ignore



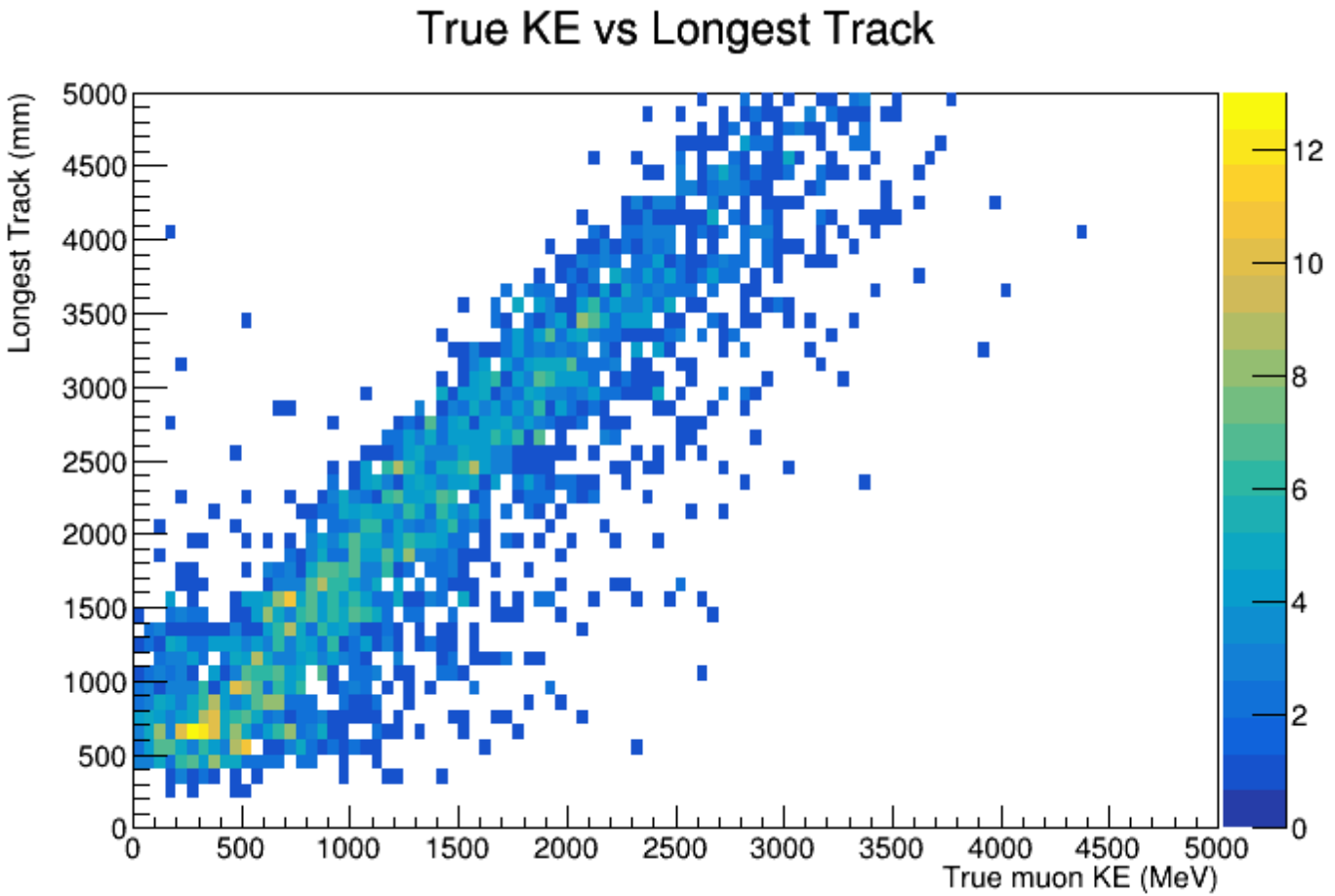
When areal density was too low, one obvious fix is to multiply by a bigger number. That's what this plot was. But that showed clearly that areal density was messed up more than just being too low



This is true muon KE vs KE estimate for all muons that start in TMS

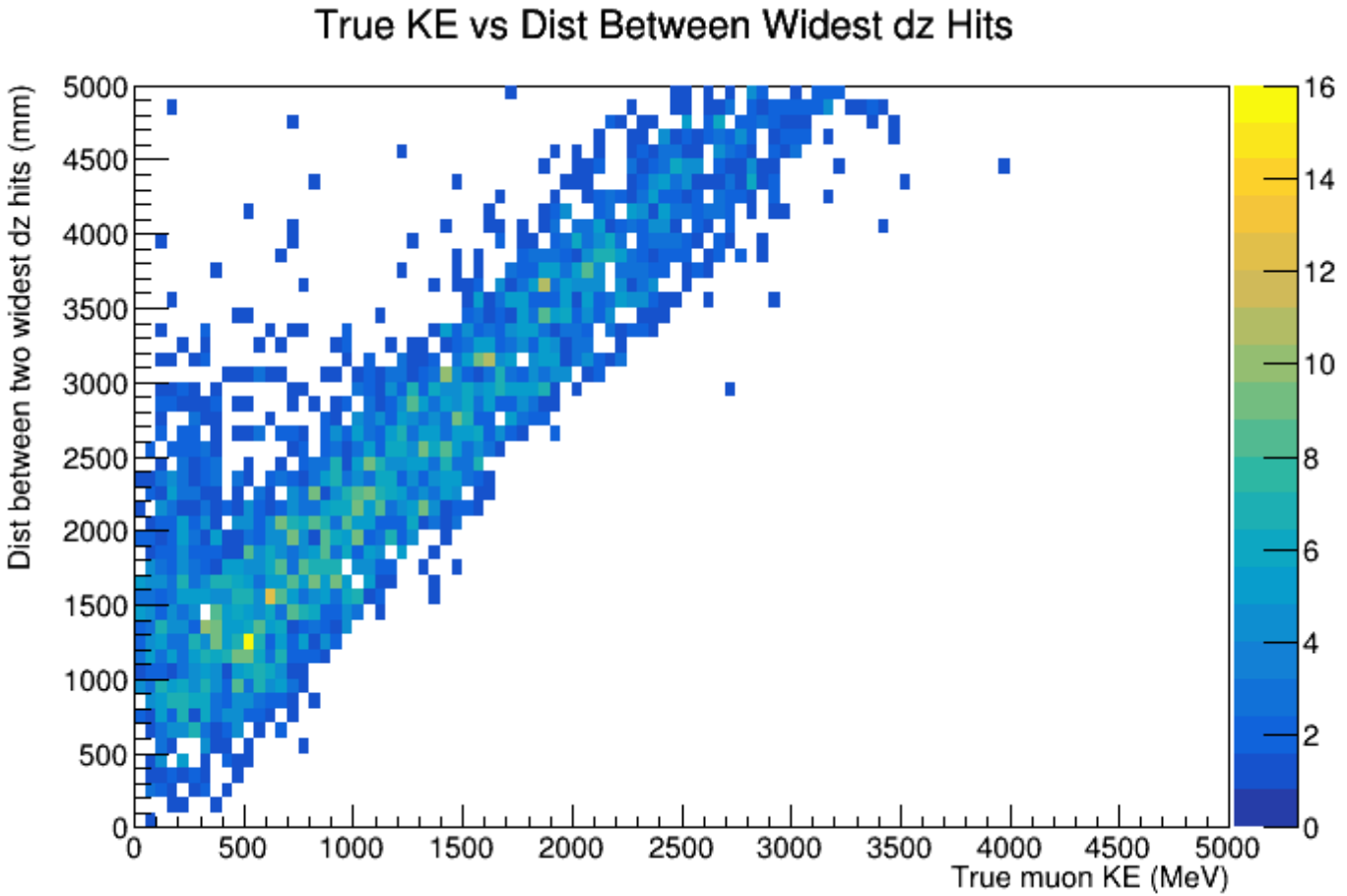


This is true muon KE vs areal density for all muons that start in TMS



This is true muon KE vs longest track for all muons that start in TMS

Can ignore

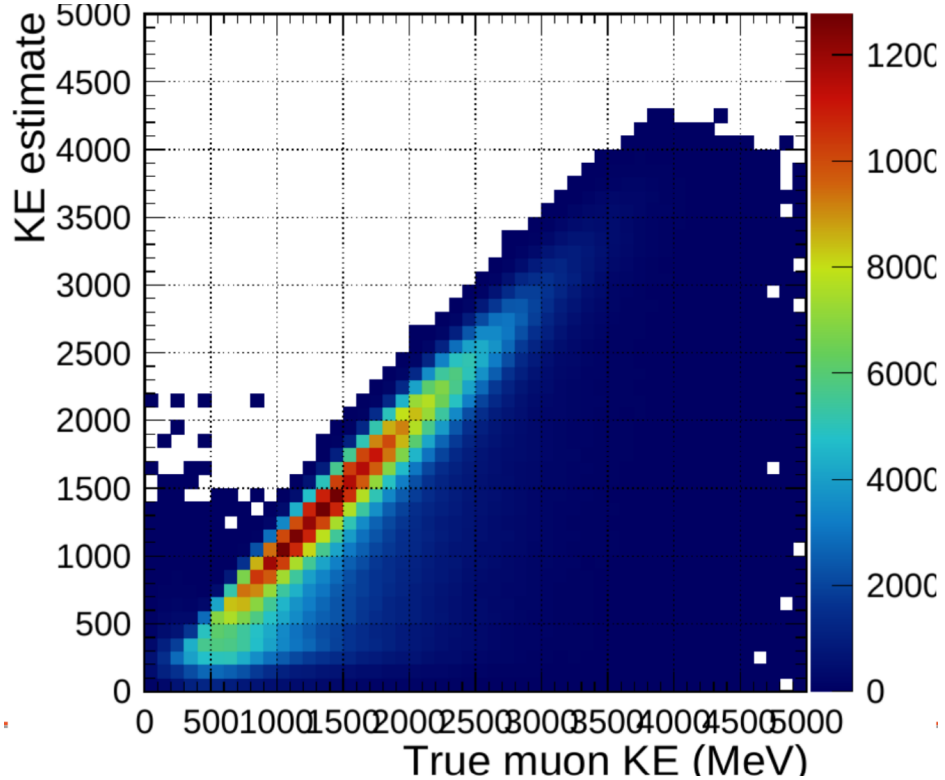


This is true muon KE vs max dz proxy for all muons that start in TMS

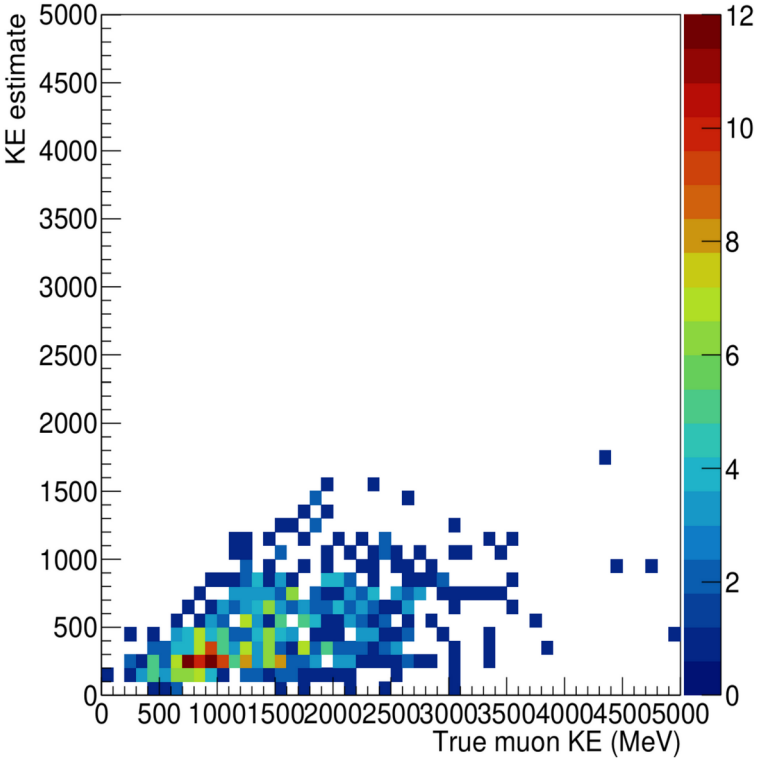
Backup

Comparison of muonke.cpp Script

Clarence 2022



Recent

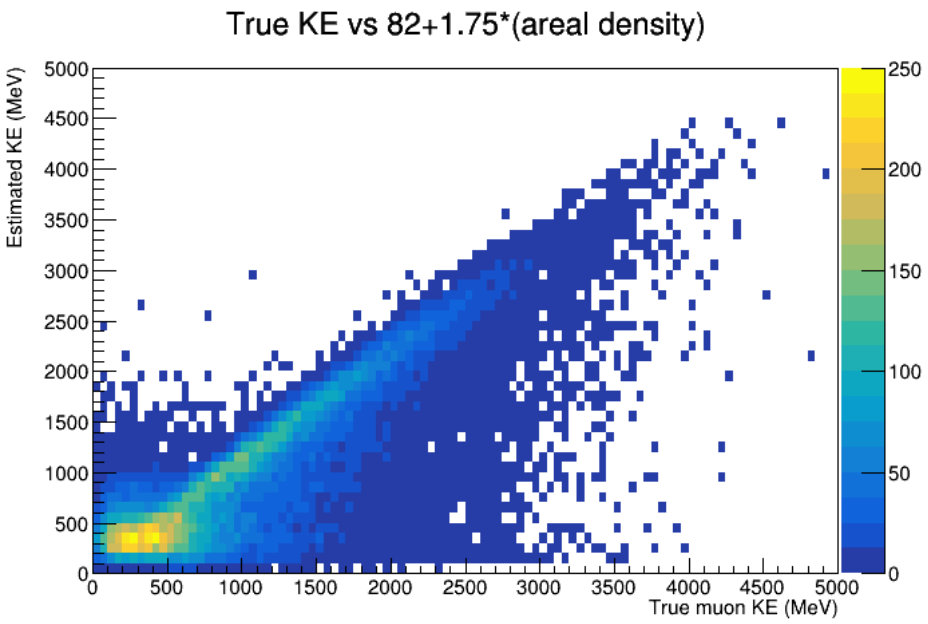


Geometry Versions

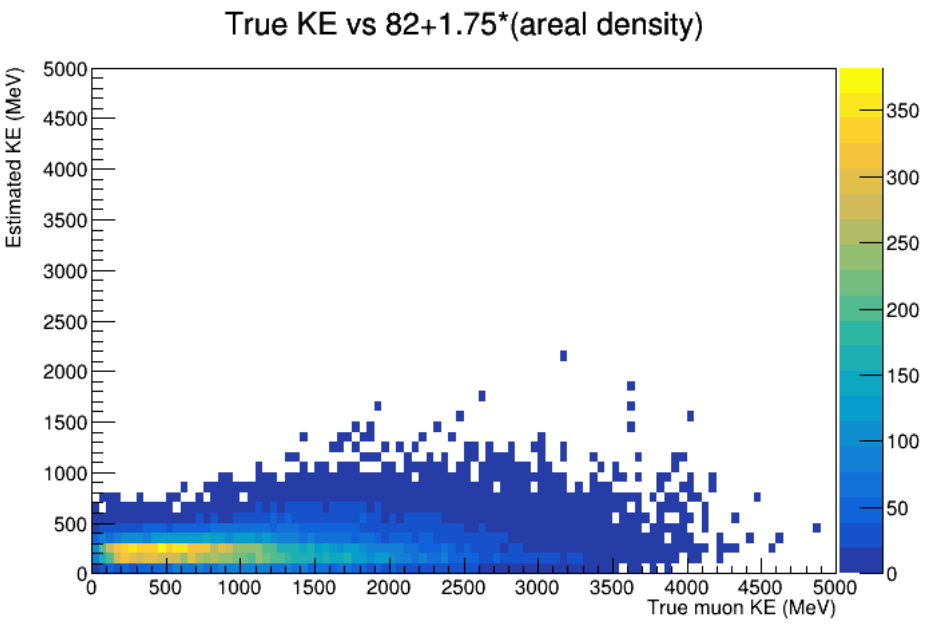
- There were some issues with the geometry used in production
- This lead me to some checking of the other versions
  - https://github.com/DUNE/dunendggd/blob/master/CHANGELOG.md
- TDR_Production_geometry_v_1.0.1
  - First versioned geometry
- TDR_Production_geometry_v_1.0.2
  - Removes overlap between planes
- TDR_Production_geometry_v_1.0.3
  - Slight changes, not sure exactly

Reco vs True Muon KE

Geometry v1.0.1



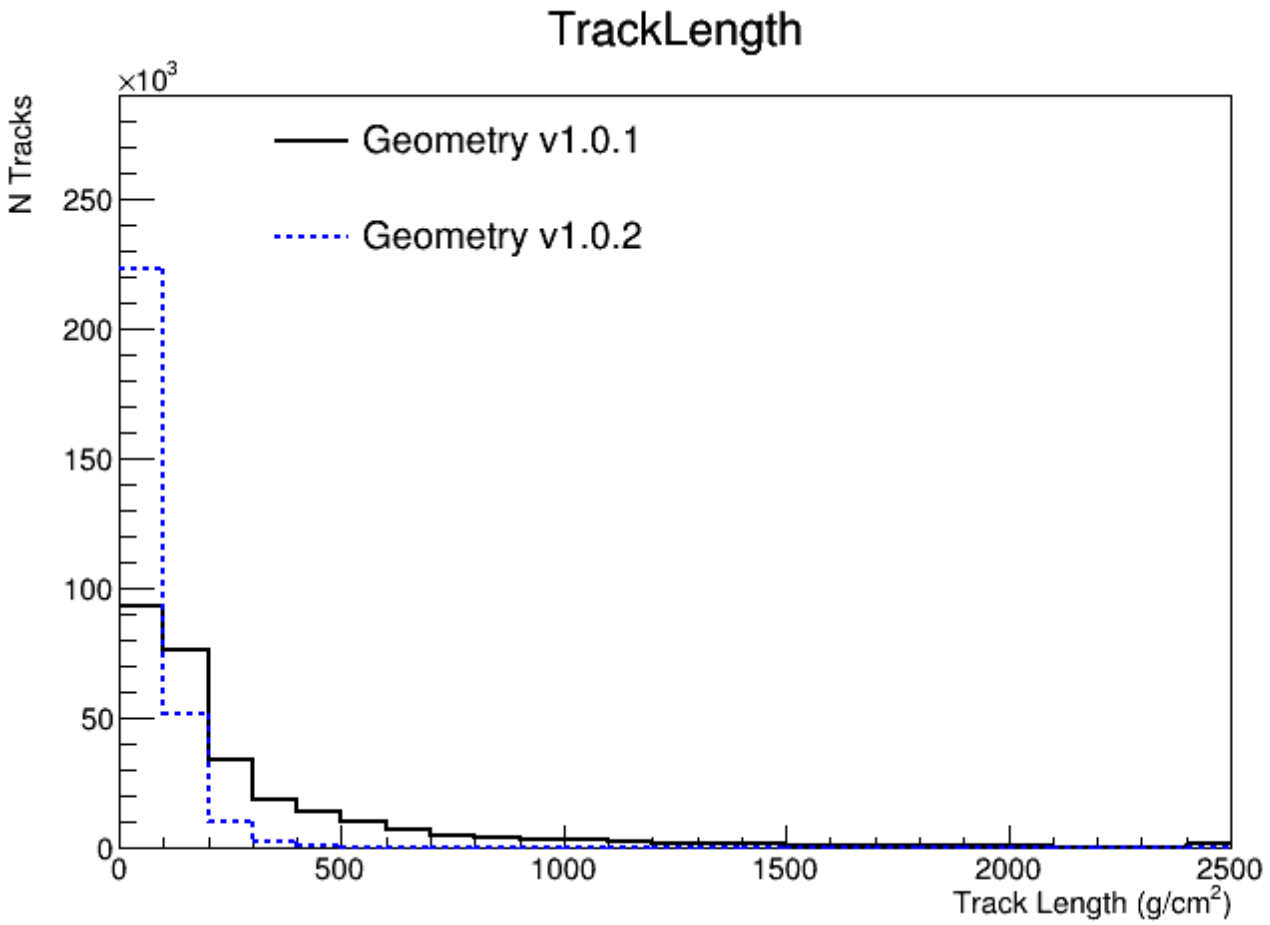
Geometry v1.0.2



Areal density = sum of track segment density * track segment length

Previously found best fit of reco KE = 82 * 1.75 areal density

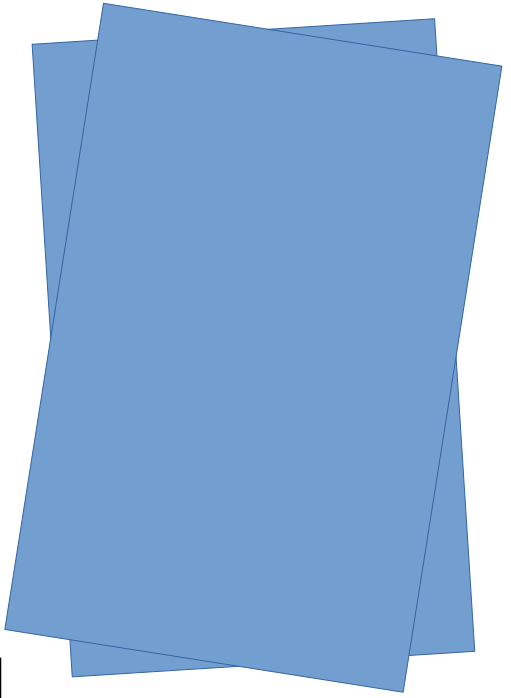
Areal Density Comparison



Exhaustively checked the material densities and other properties in the geometry gdml files. They all look correct and are correct while running.

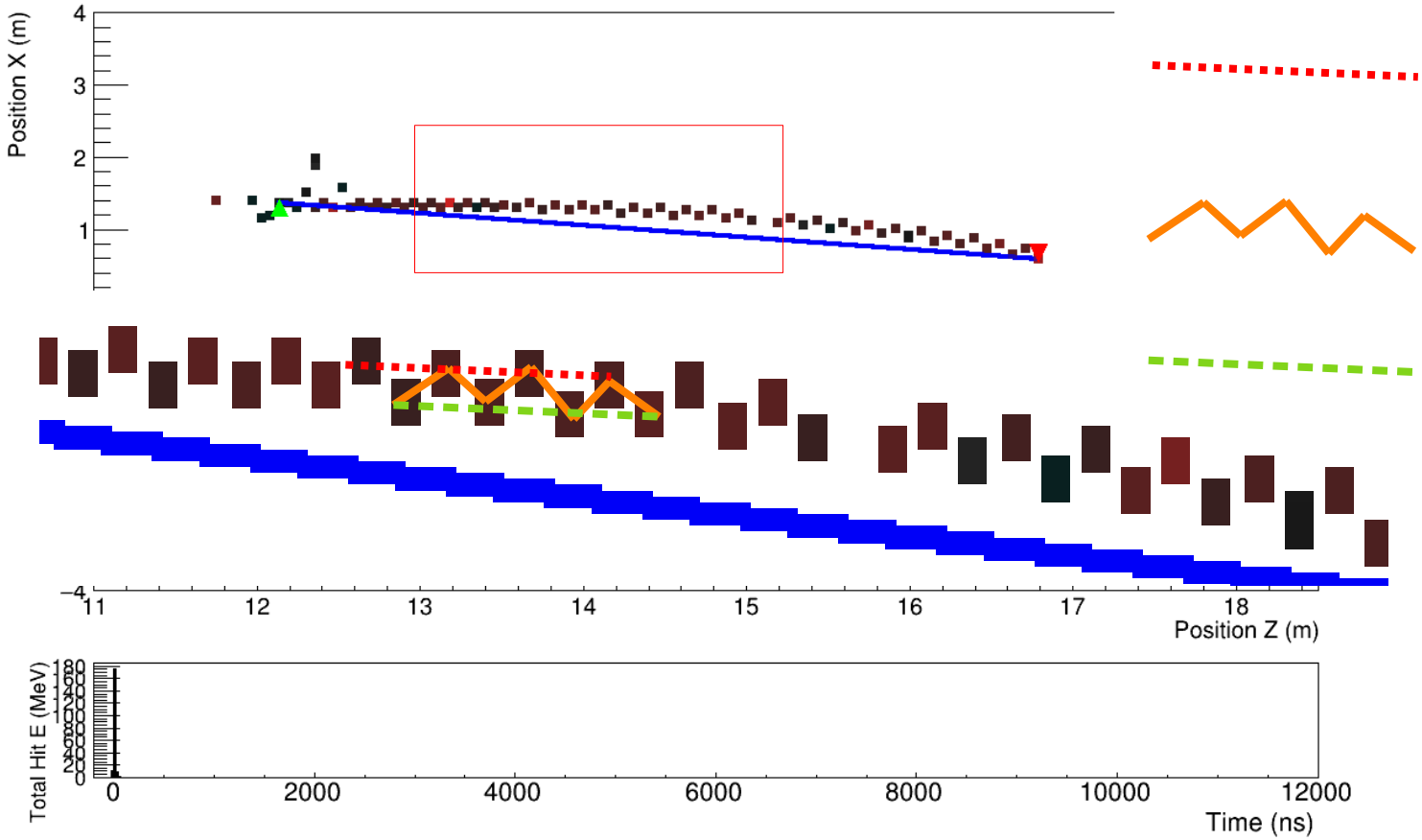
TMS Scintillator Plane Orientations

- Each plane in the stack is either +3 deg rotated or -3 deg rotated
  - Called Y view or V view
  - This gives low resolution 3d reconstruction (see Asa's work)



- In Geom v1, planes were stacked so all hits were Y views only since only first plane was used for hit purposes

Problem 1: Overestimating Areal Density



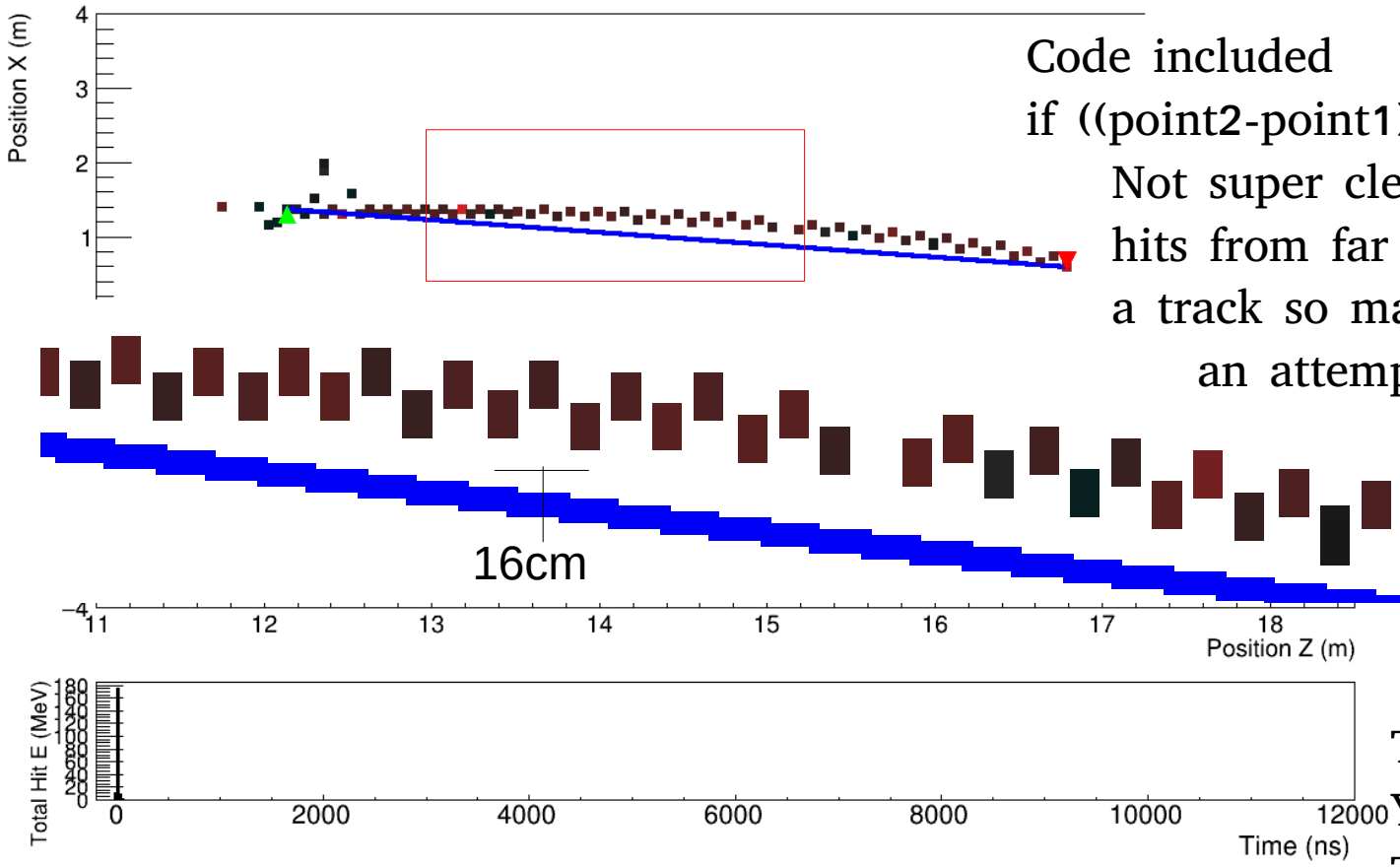
Actual Path
V View

Path Used for Areal Density Estimate

Actual Path
Y View

The zig-zag path is too long. This wasn't an issue in Geom v1 because all hits were y view

Problem 2: Skipping Distance Points



Code included 100mm

if ((point2-point1).Mag() > 100) continue;

Not super clear why, but in some cases hits from far away can be included in a track so maybe this was an attempt to remove the effect of bad reco.

However, many hits are > 100 mm apart if you only consider y or v view

Thin modules 5-6cm apart

Y views are 11cm apart

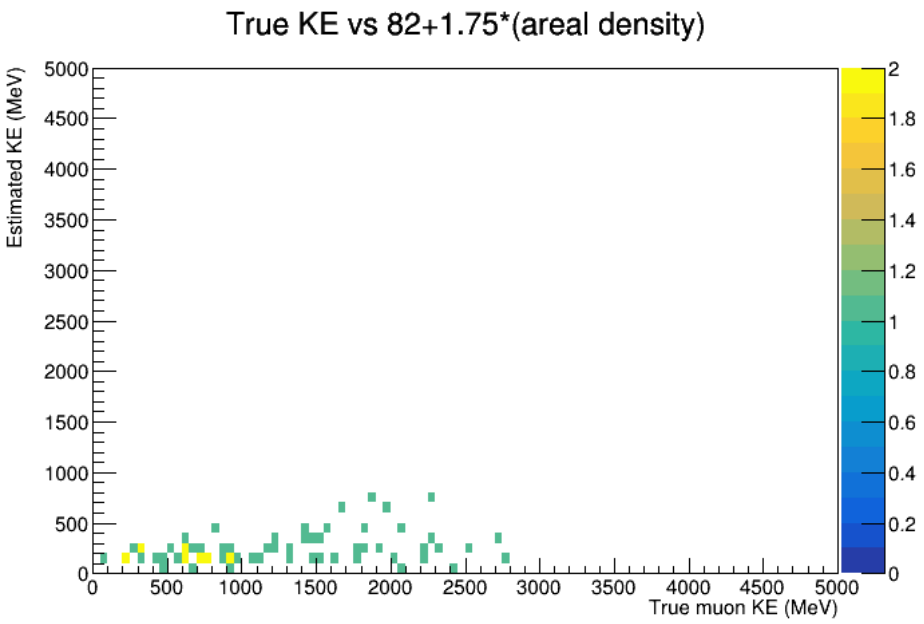
Thick 8cm apart

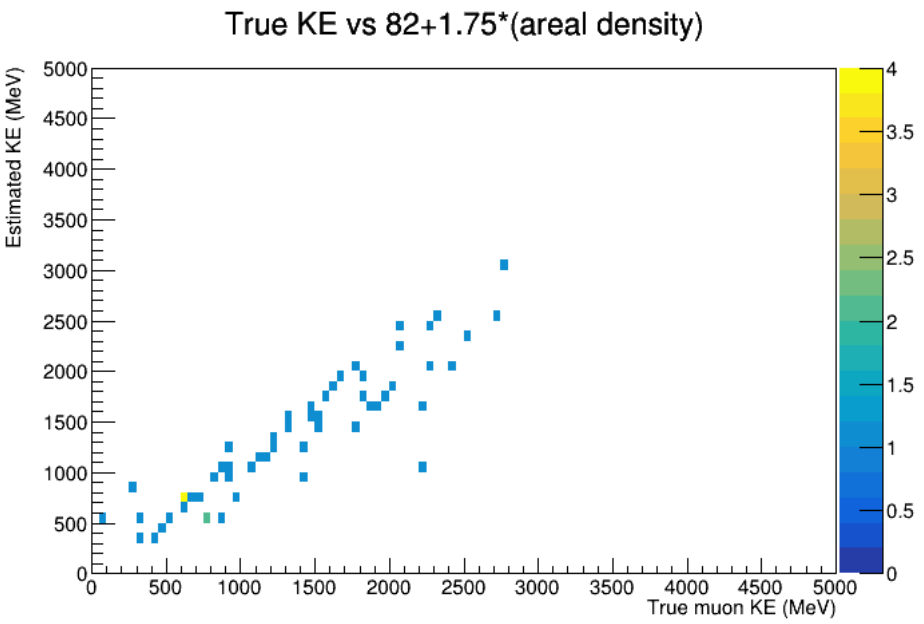
Y views are 16cm apart

Fixes

- Add bar orientation information
  - When creating hits, all were labeled y view. But we can use the geometry node names to find out the view
    - modulelayervol1 are y view
    - modulelayervol2 are v view
  - This does have knock-on effects. Currently, the reco has ProjectHits function everywhere which filters to Y-view only
- Remove that if dist(P1, P2) > 100, continue statement

KE Estimator, Before/After





Conclusion

- All this is hot off the presses
- Will make PR once I clean everything up
  - And some new files