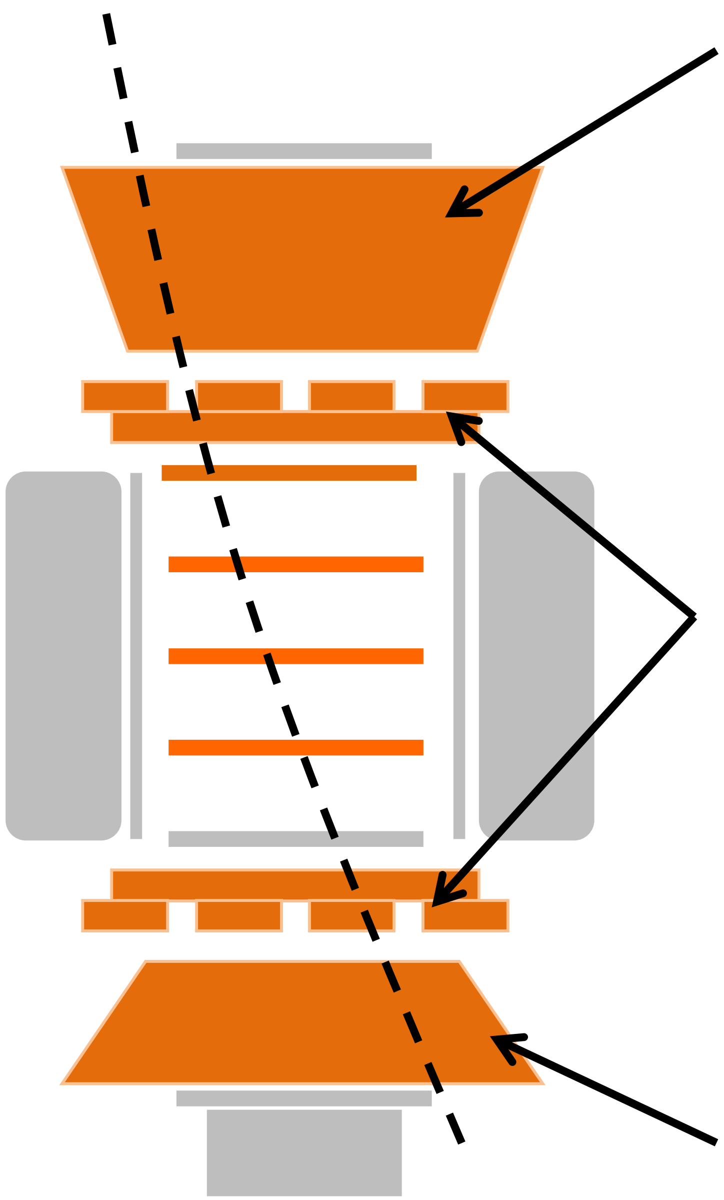
dBar analysis status

23/2/2018 AMS Italy - V. Formato

Overview



TRD → elimination of the large e^- background
separation He/p on top
D/p discrimination (dE/dx for very low energy particles)

Tracker → particle sign +/-
Z = 1
Rigidity for mass estimation (~ 10% up to 20 GV),

TOF → β for mass estimation (4% at β = 1 and Z =1, β_{thr} ~ 0.4)
Z = 1

RICH NaF → β for mass estimation (~0.4% at β = 1 and Z = 1, β_{thr} ~ 0.75)
RICH Aerogel → β for mass estimation (~0.1% at β =1 and Z = 1, β_{thr} ~ 0.96)
Z = 1

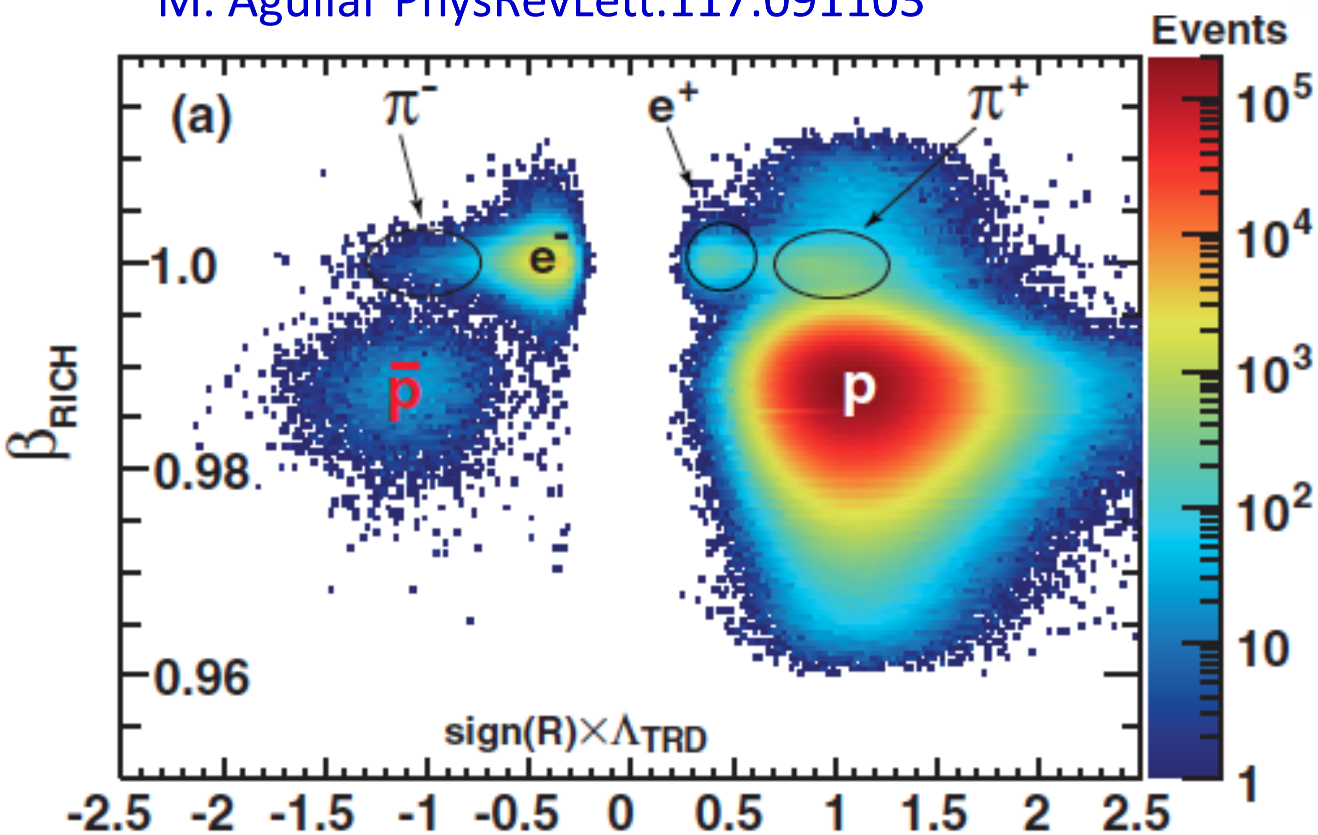
First attempt

Our first attempt on data analysis is based on the selections used for the antiproton paper

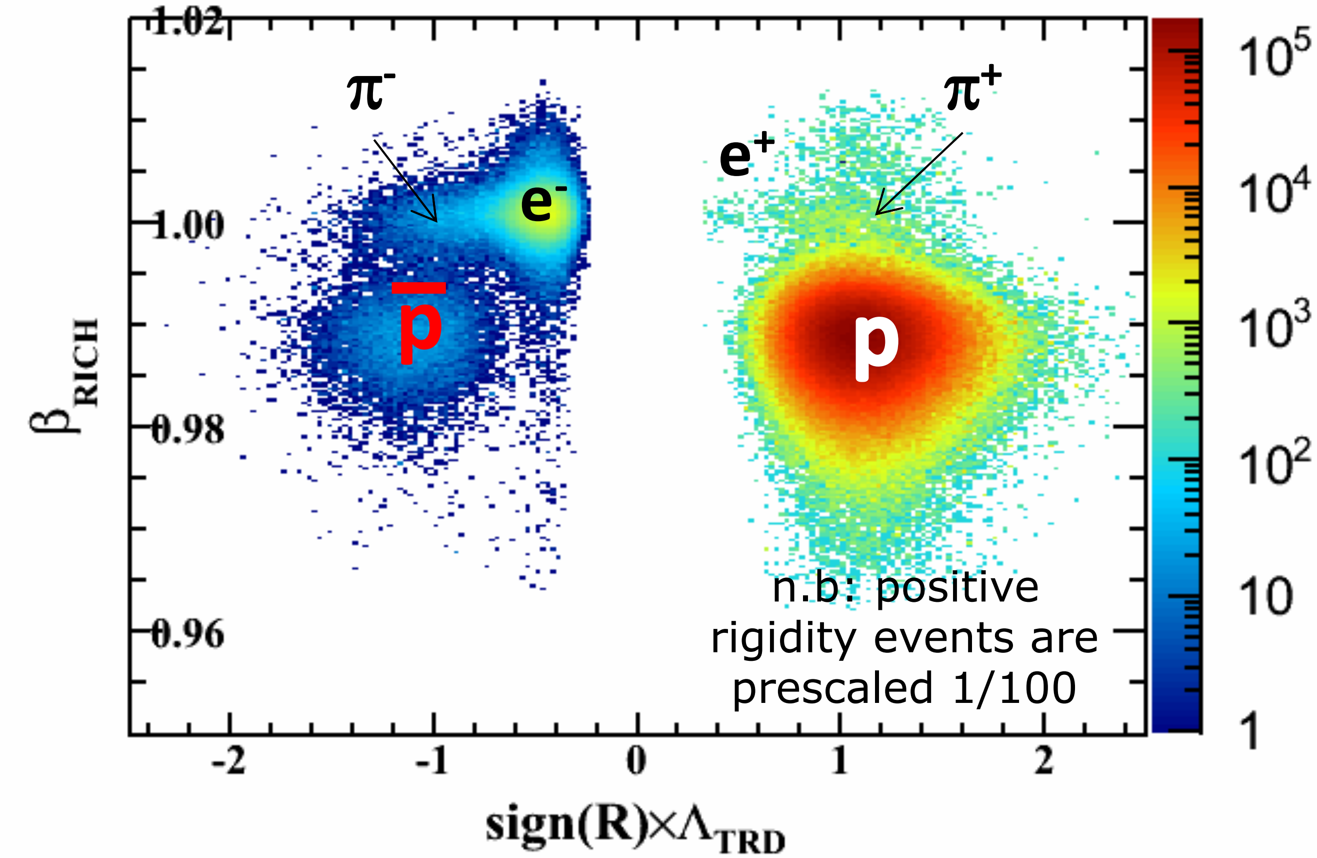
Data period: 4 years

5.4<R<6.5

M. Aguilar PhysRevLett.117.091103



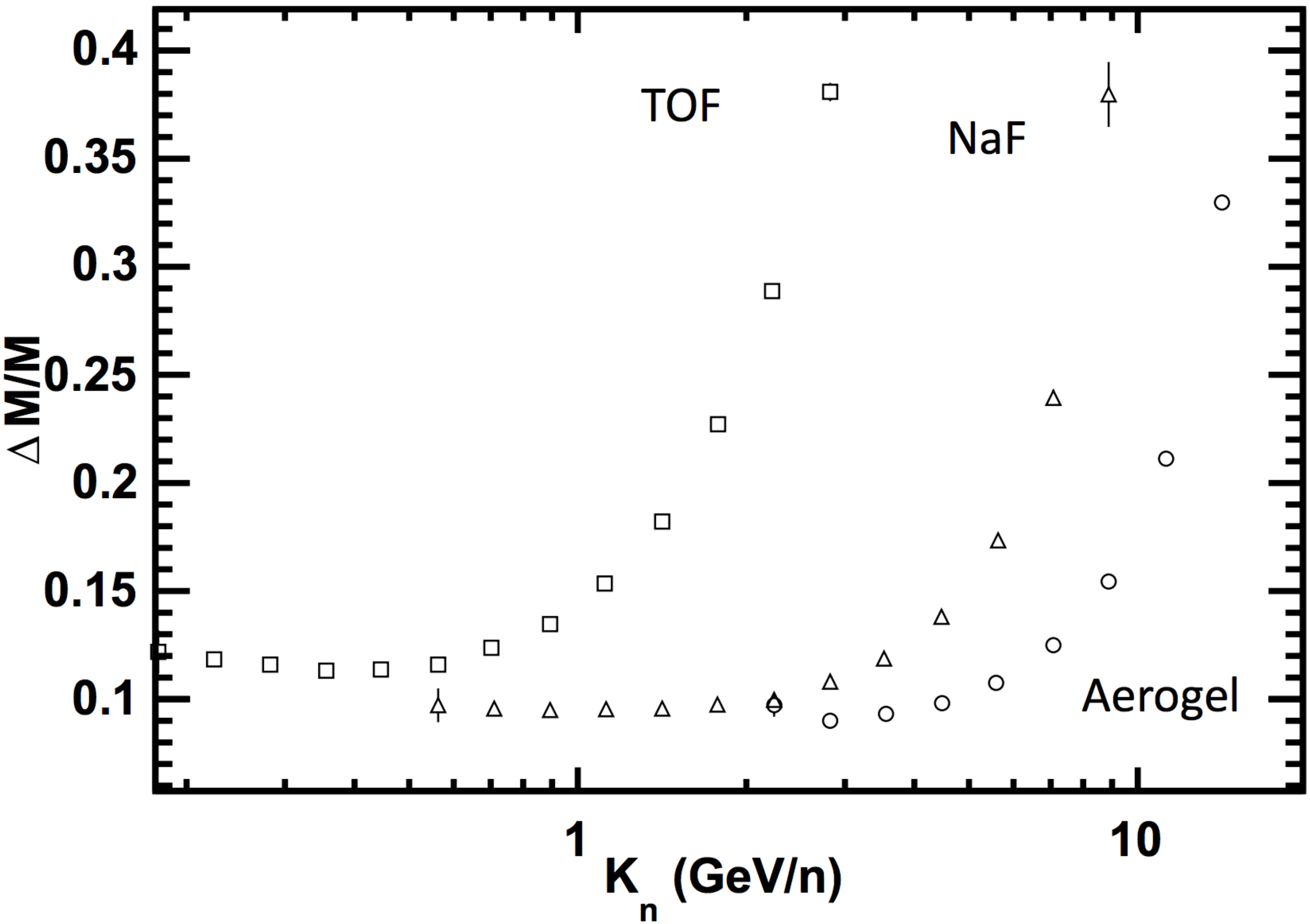
This analysis



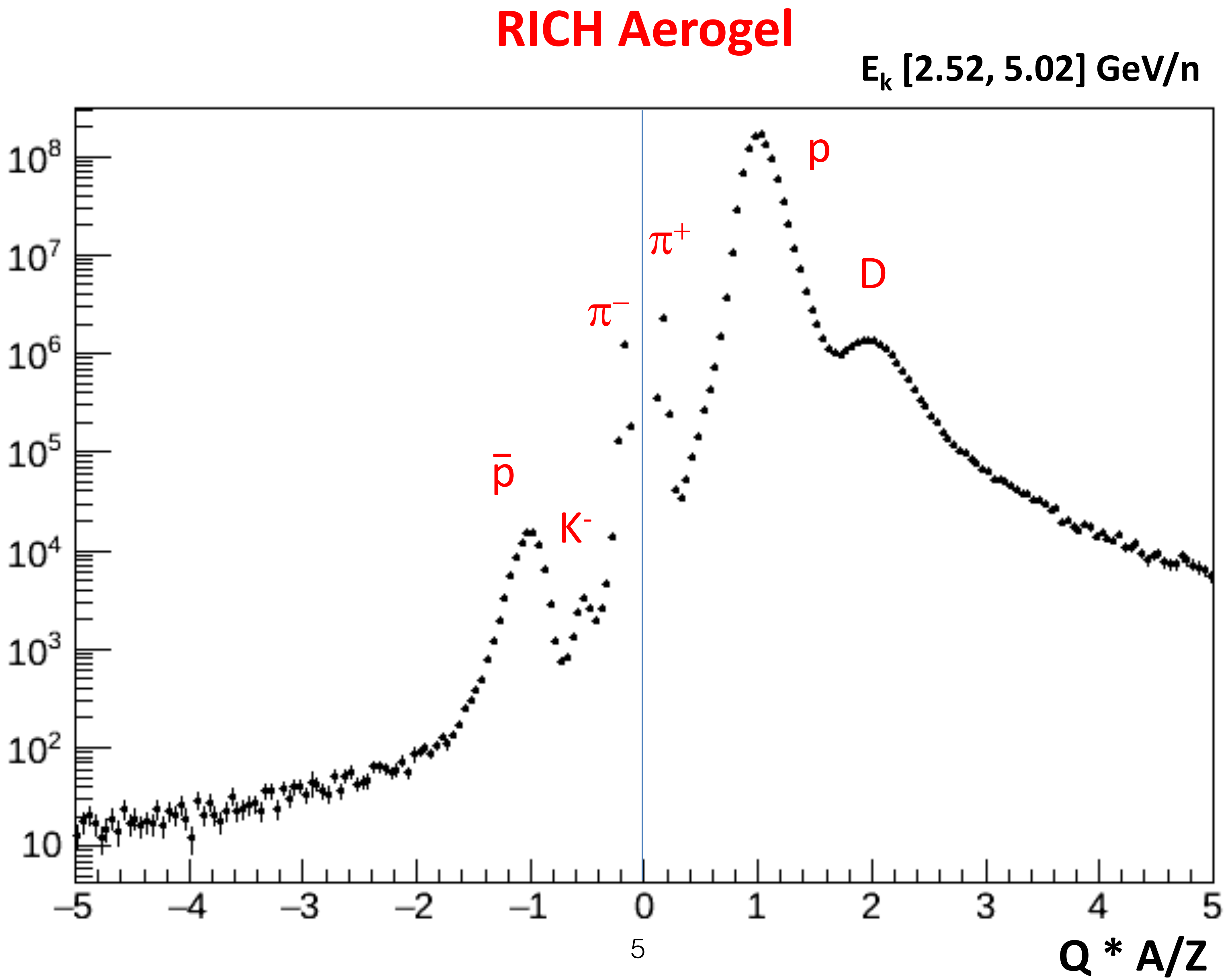
Counting strategy:
Beta RICH cut (R dependent)
Project X and subtract remaining e-like with ECAL Template

Counting strategy:
TRD likelihood cut (0.6)
Project Y and fit the Mass distribution
Mass window: +- 3 mass sigma

Mass resolution

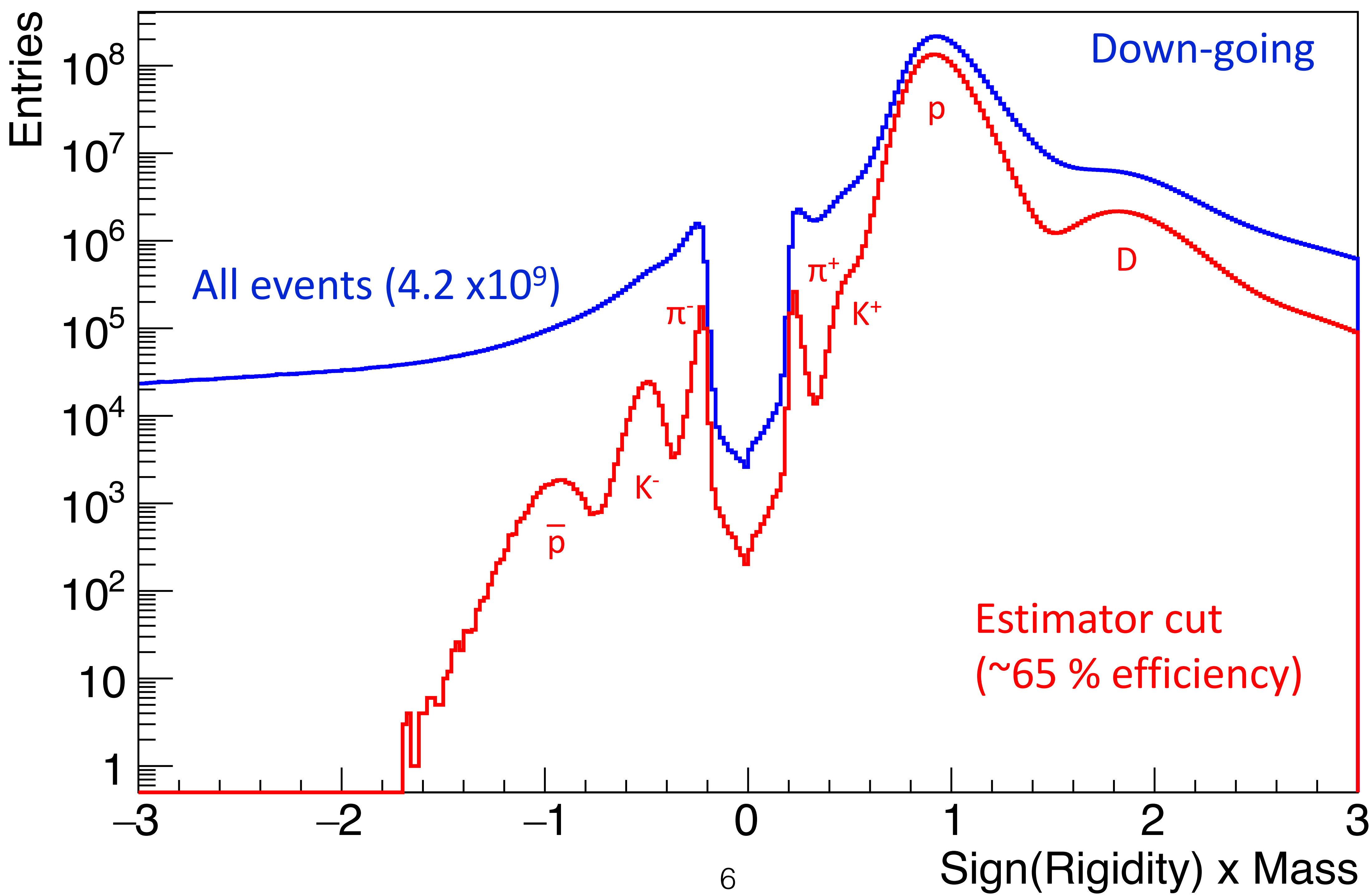


First attempt



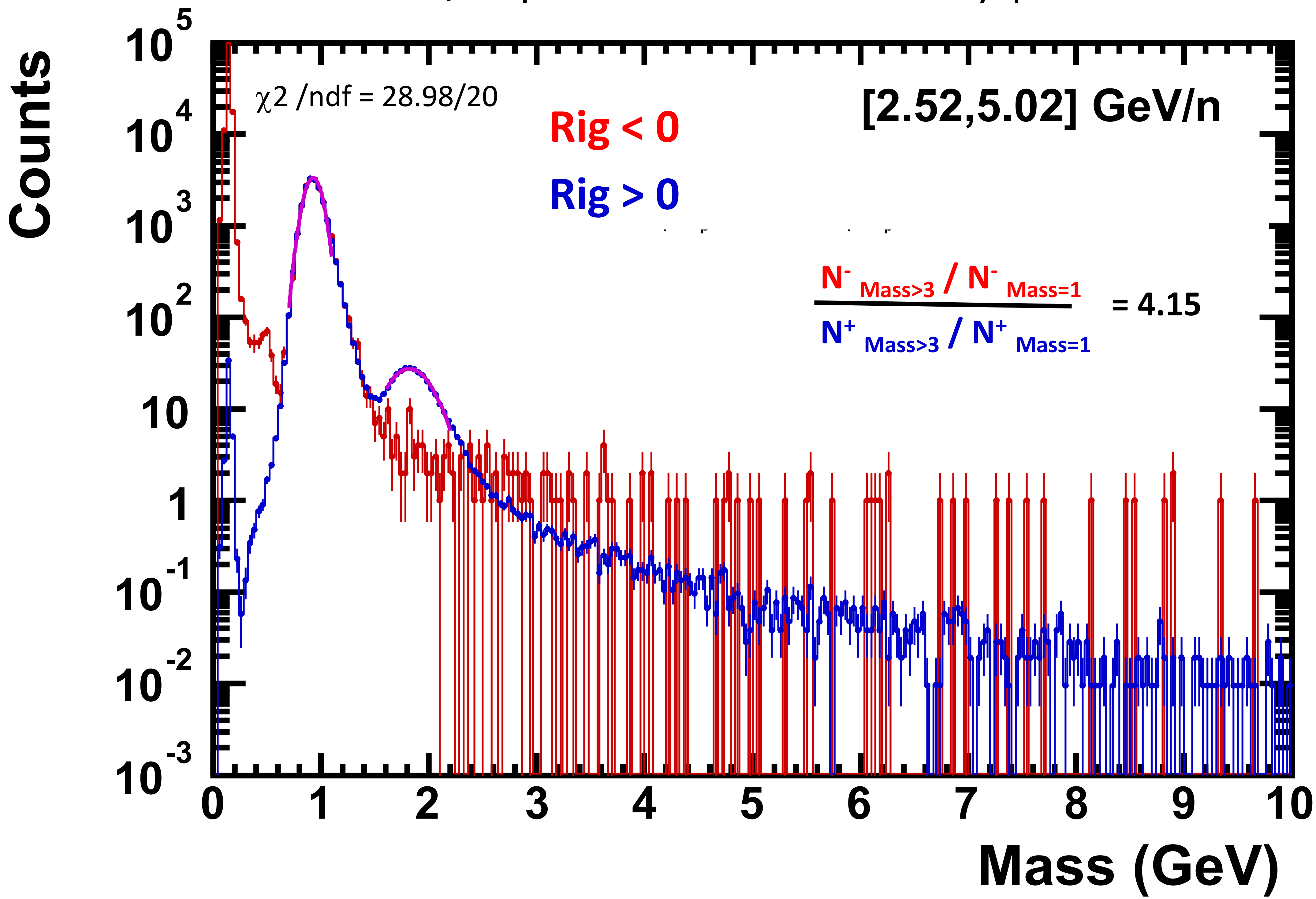
Tof Mass quality likelihood

Mass tails in the Tof region are mostly due to a bad beta measurement. There is already a mass quality estimator from Sada that can clean the distribution quite well



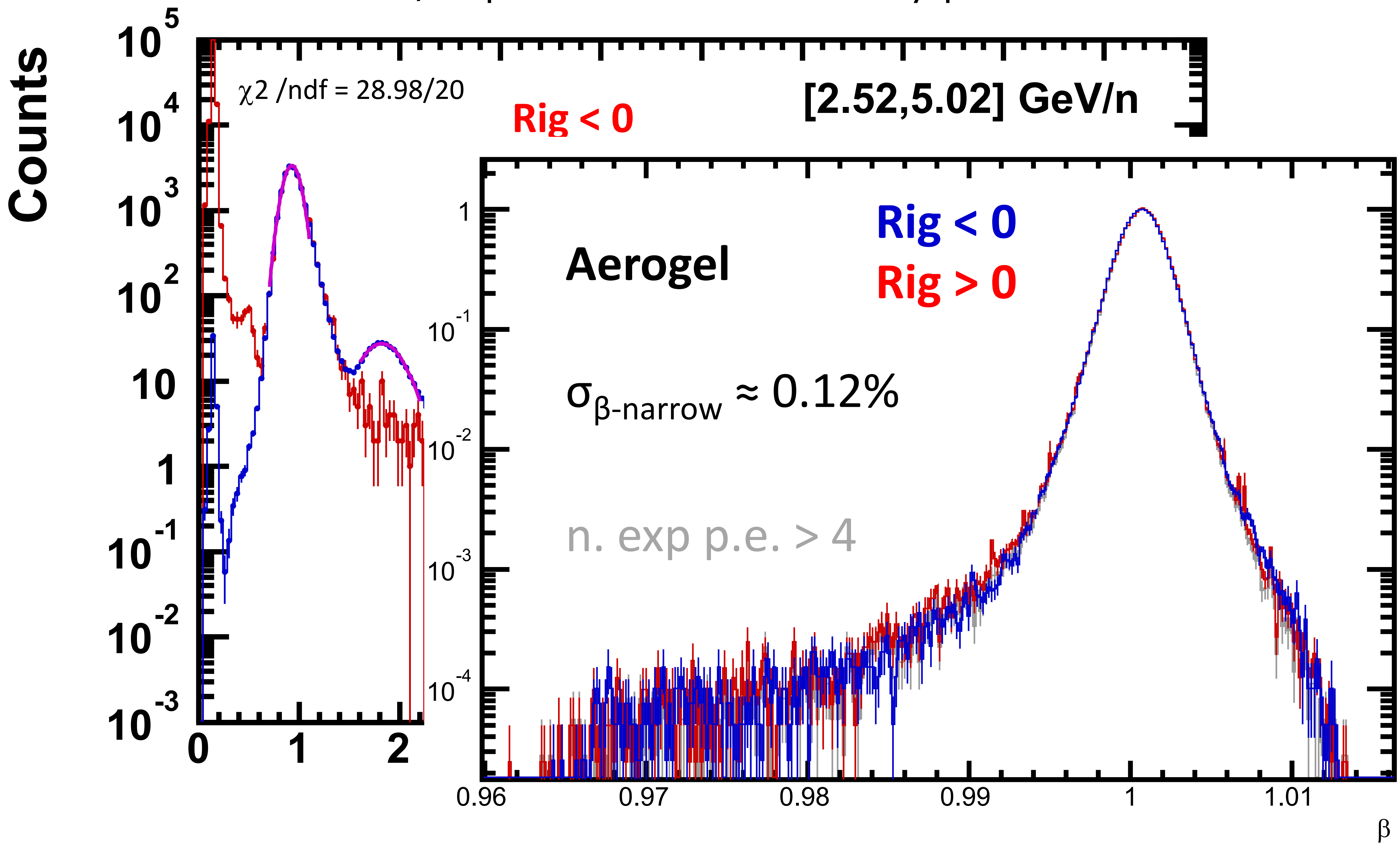
Rich mass quality?

However tails in the mass distribution for the Agl region are most likely due to the tracker resolution and/or photons from secondary particles in the RICH



Rich mass quality?

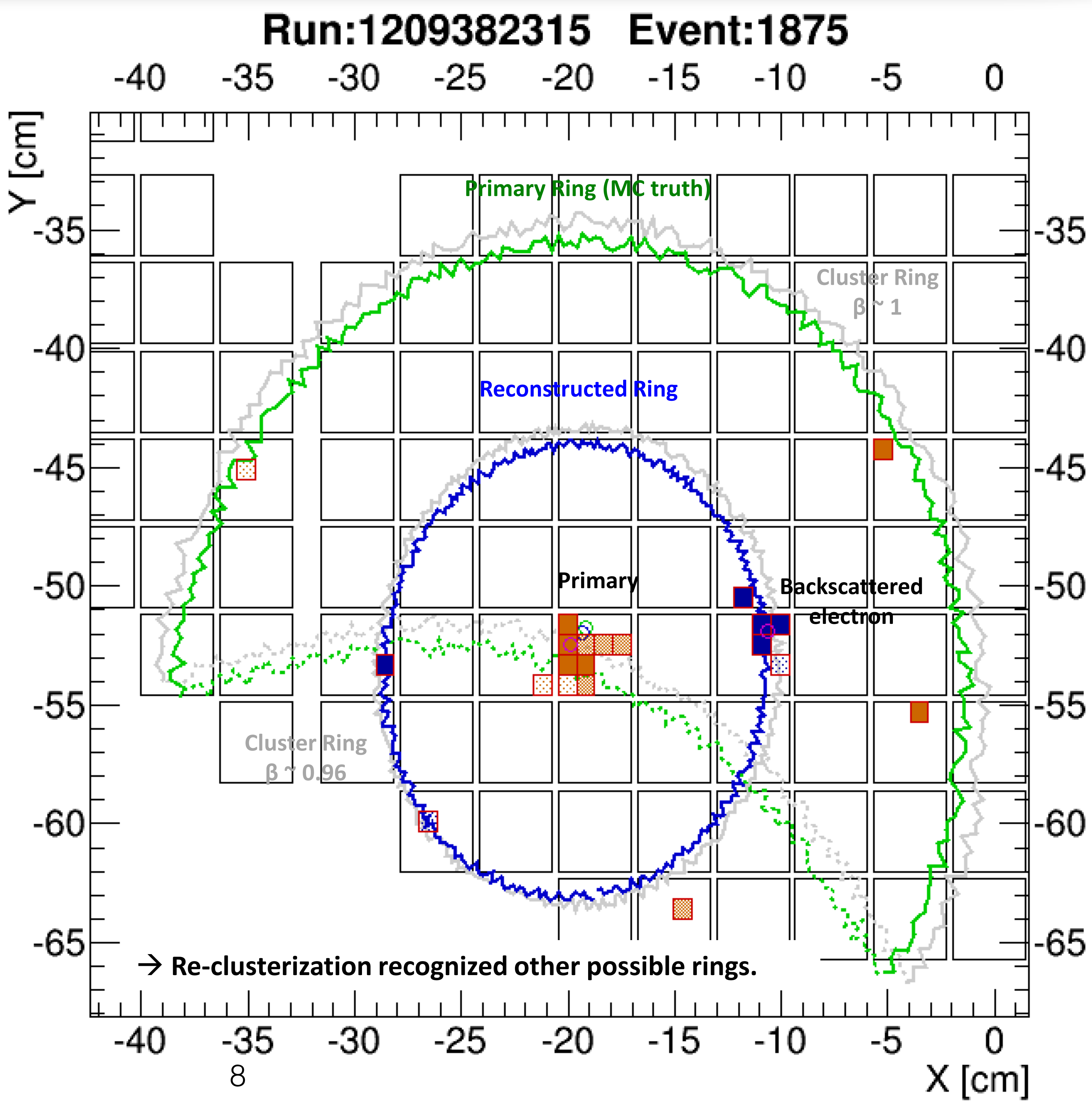
However tails in the mass distribution for the Agl region are most likely due to the tracker resolution and/or photons from secondary particles in the RICH



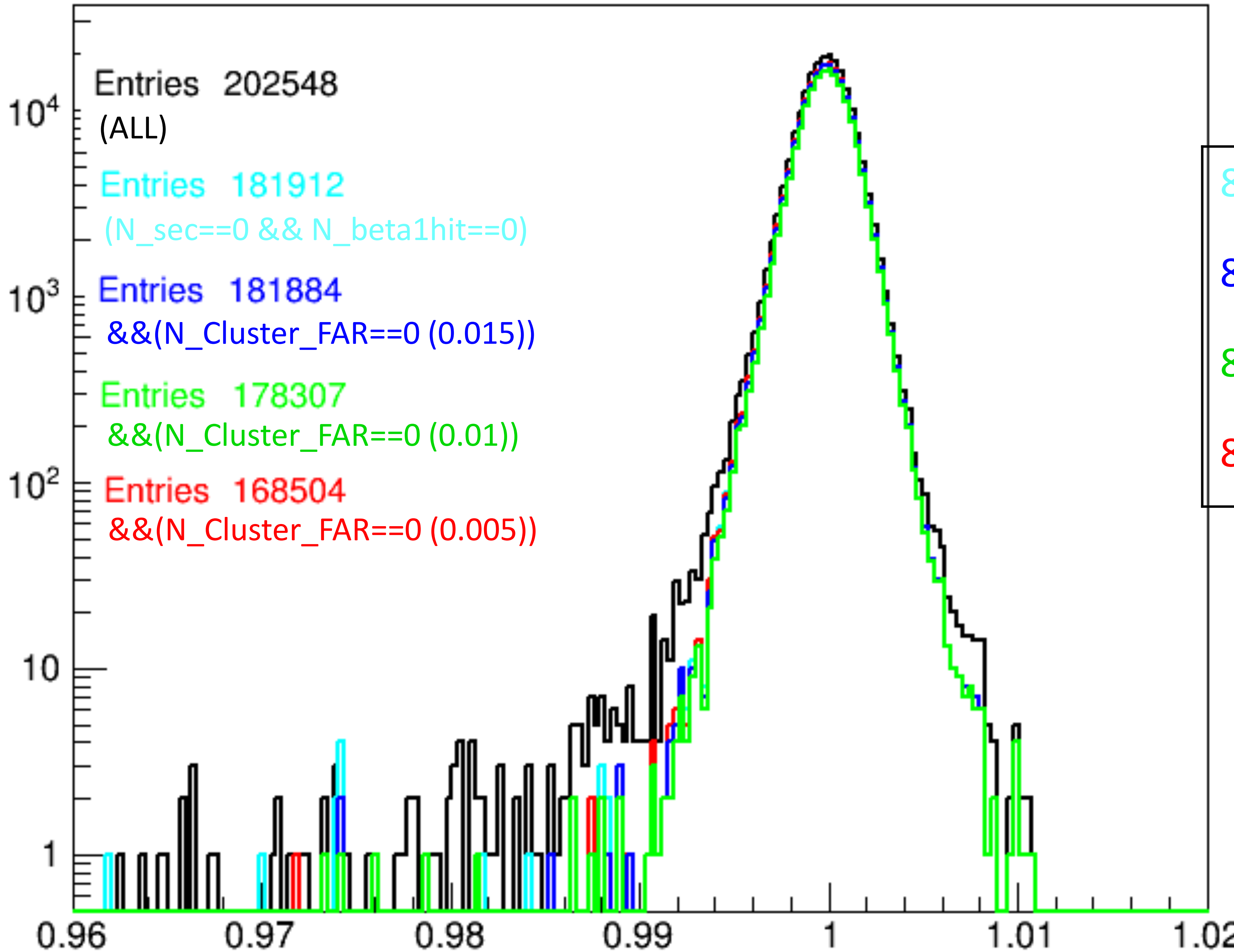
Rich mass quality?

Agl tails are due to wrong clustering of hits in the PMTs. There are quite a few different cases but most of them can be caught with the following criteria:

- No secondary “hotspot”
- No unused hits with beta~1
- Hits re-clustering doesn’t give drastically different results



Rich mass quality?

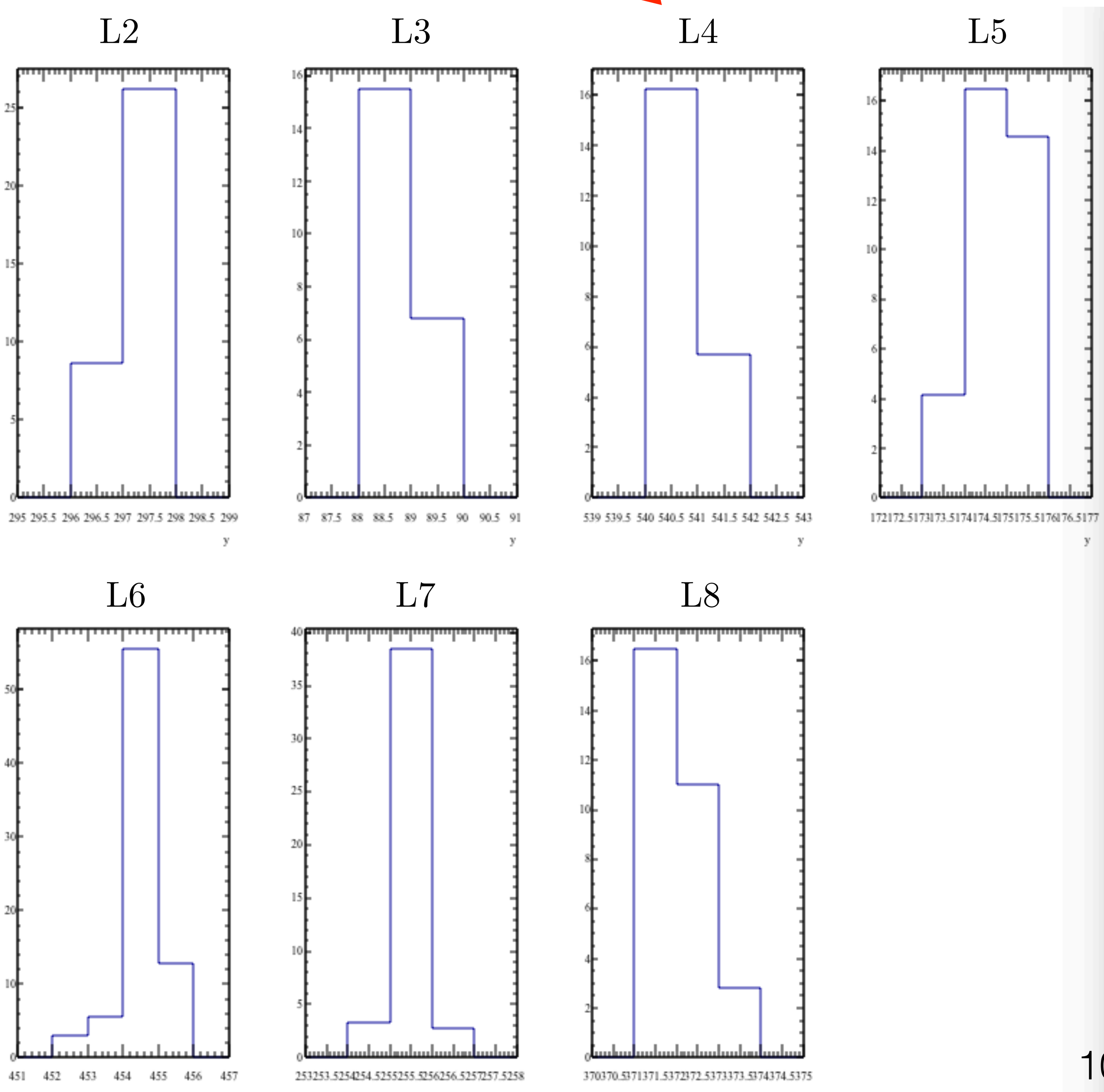


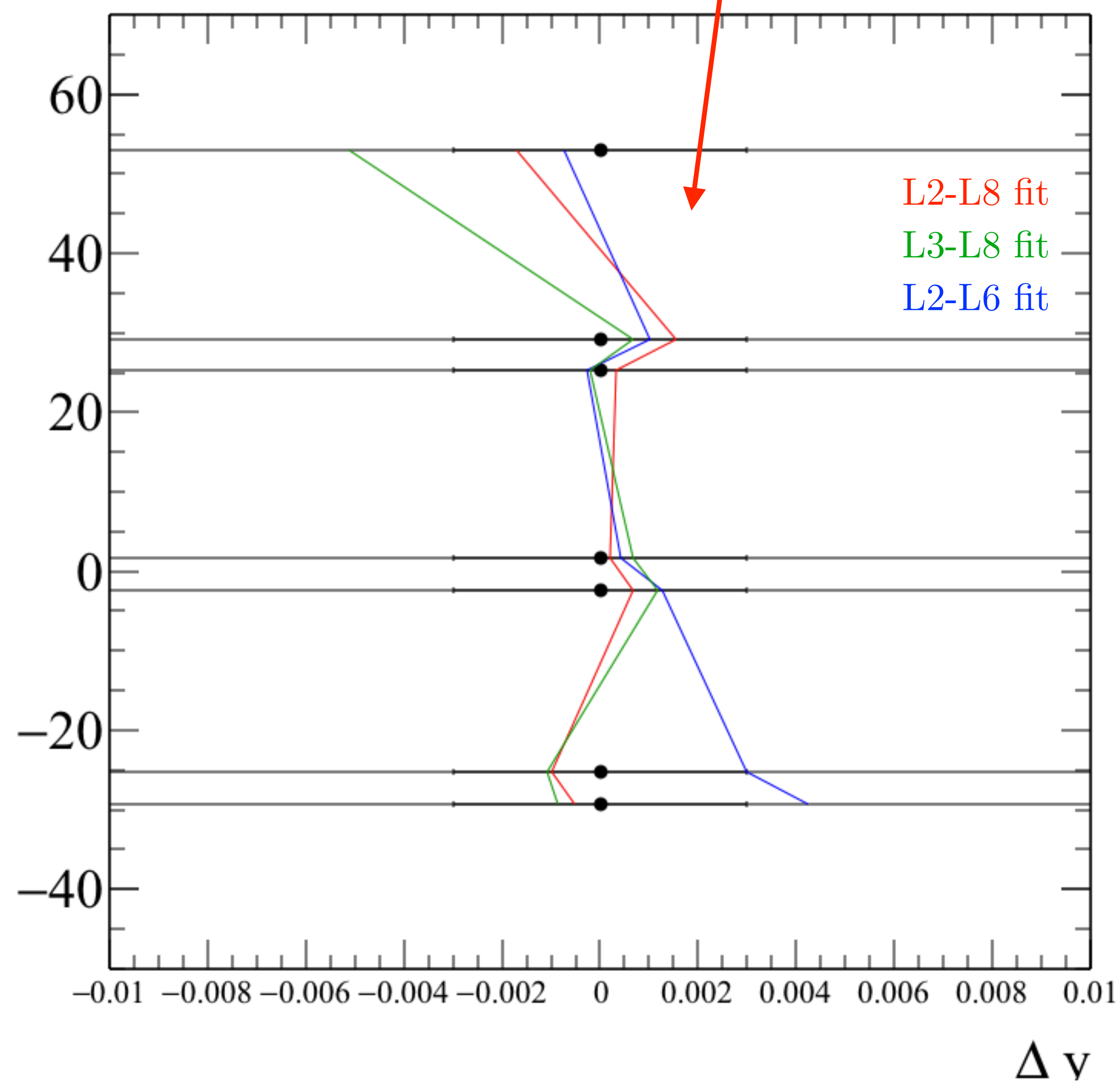
Peak eff.	Tail eff.
89.8%	22%
89.8%	15%
88.8%	12%
83%	8.5%

(n.b. The beta~1 criterion can actually lead to a bias at low energy, currently investigating other possibilities)

Multiple scattering?

Currently working on a similar concept for the tracker. Trying to identify common patterns on negative events with high mass. Currently looking cluster shape and fit consistency while dropping hits (especially on L2 and L7/8)





Selections

Tof cuts

$\beta > 0.3$

N. BetaH clusters = 4

1 BetaH cluster per plane

$0 <$ Tof χ^2(time) < 15

Good association with track

$0 <$ Tof χ^2(coord) < 10

All 4 clusters on track

Extra edep on Tof $<$ 20 MeV

Charge Z=1

RICH cuts

Good RICH status

No bad PMTs on ring

At least 2(1) expected ph.el. for Agl(NaF)

β compatibility Ciemat/LIP

At least 3 hit PMTs

RICH prob. > 0.01

No hotspots

Re-cluster consistency

No β =1 unused hits

TRD cuts

Track inside TRD

TRDK charge Z=1

TRDK likelihood hits $> 6.57 * \log_{10}(E_k) + 3.43$

In-house vertex reconstruction TRD segments < 10

Tracker cuts

1 TrTrack

Tracker Y pattern L2 && (L3 || L4) && (L5 || L6) && (L7 || L8)

At least 6 Y hits, 4 XY hits

Minimum hit charge in Inner Tracker > 0.5

Inner Tracker charge Z=1

Choutko/Chikanian $\chi^2_x < 15$

Choutko/Chikanian $\chi^2_y < 10$

Choutko/Chikanian rigidity sign compatibility

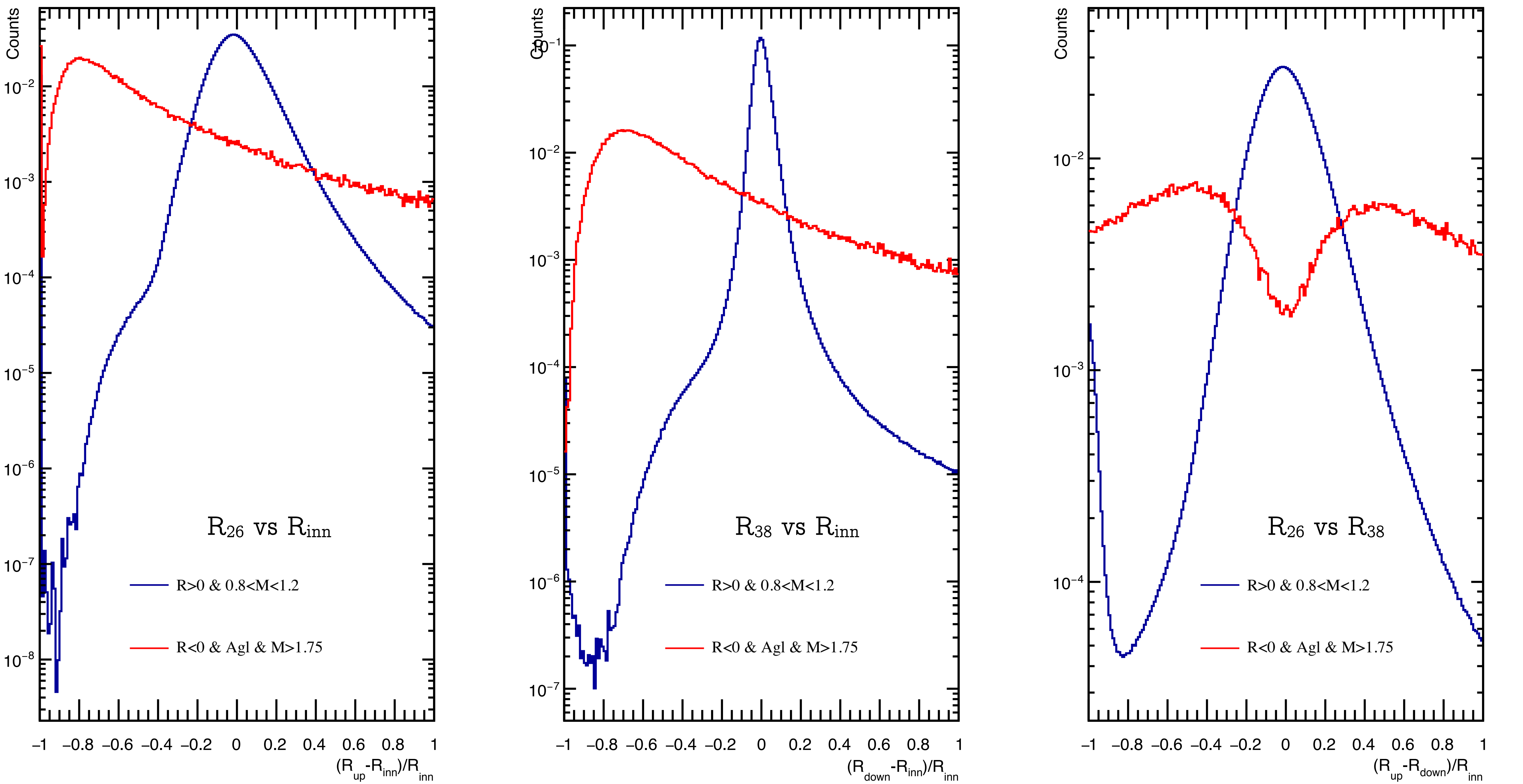
Good L2 hit charge and status

Edep off-track on L2 $<$ 2 MeV

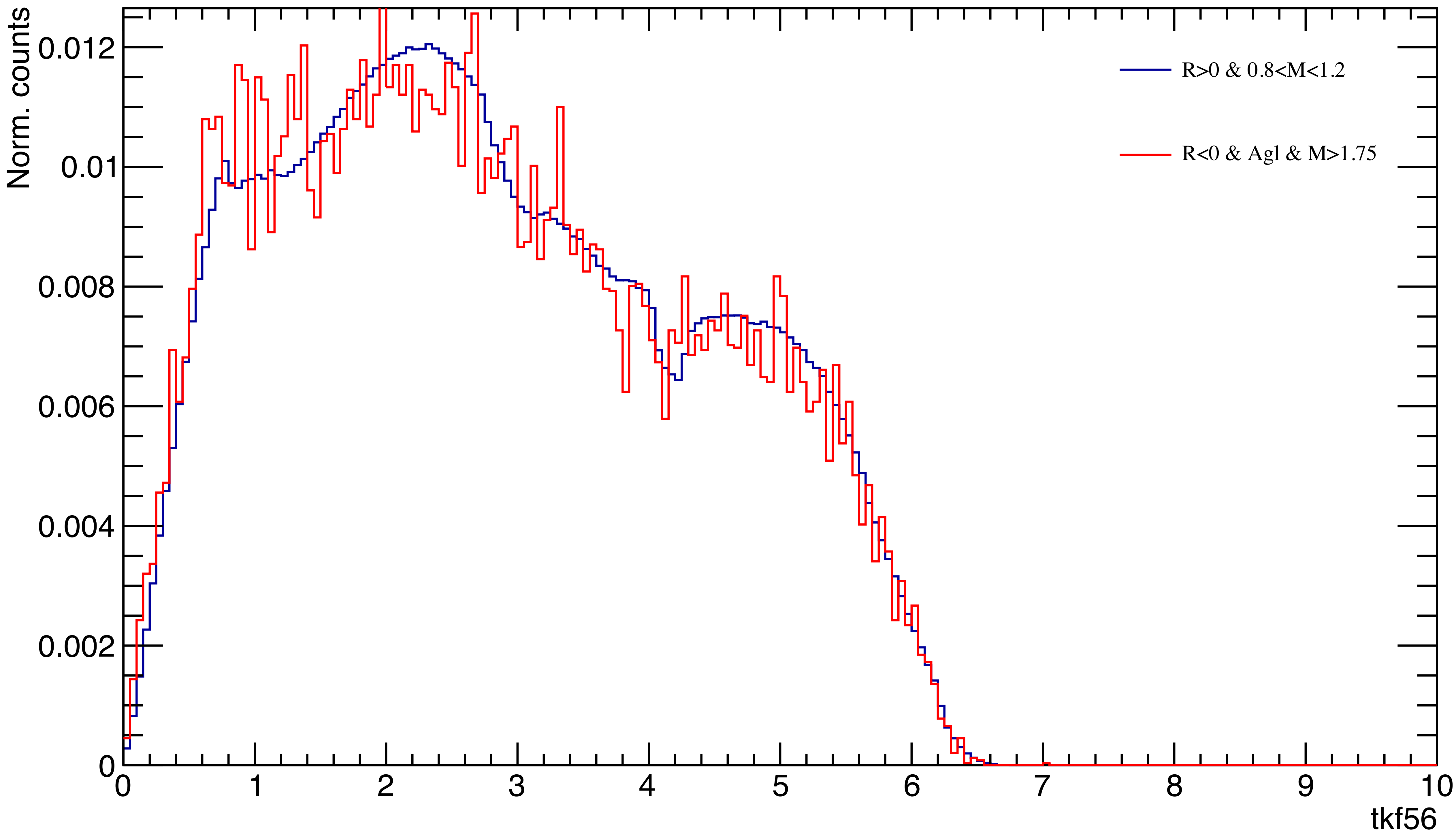
Edep off-track whole inner tracker $<$ 5 MeV

L2-6/L3-8 rigidity compatibility

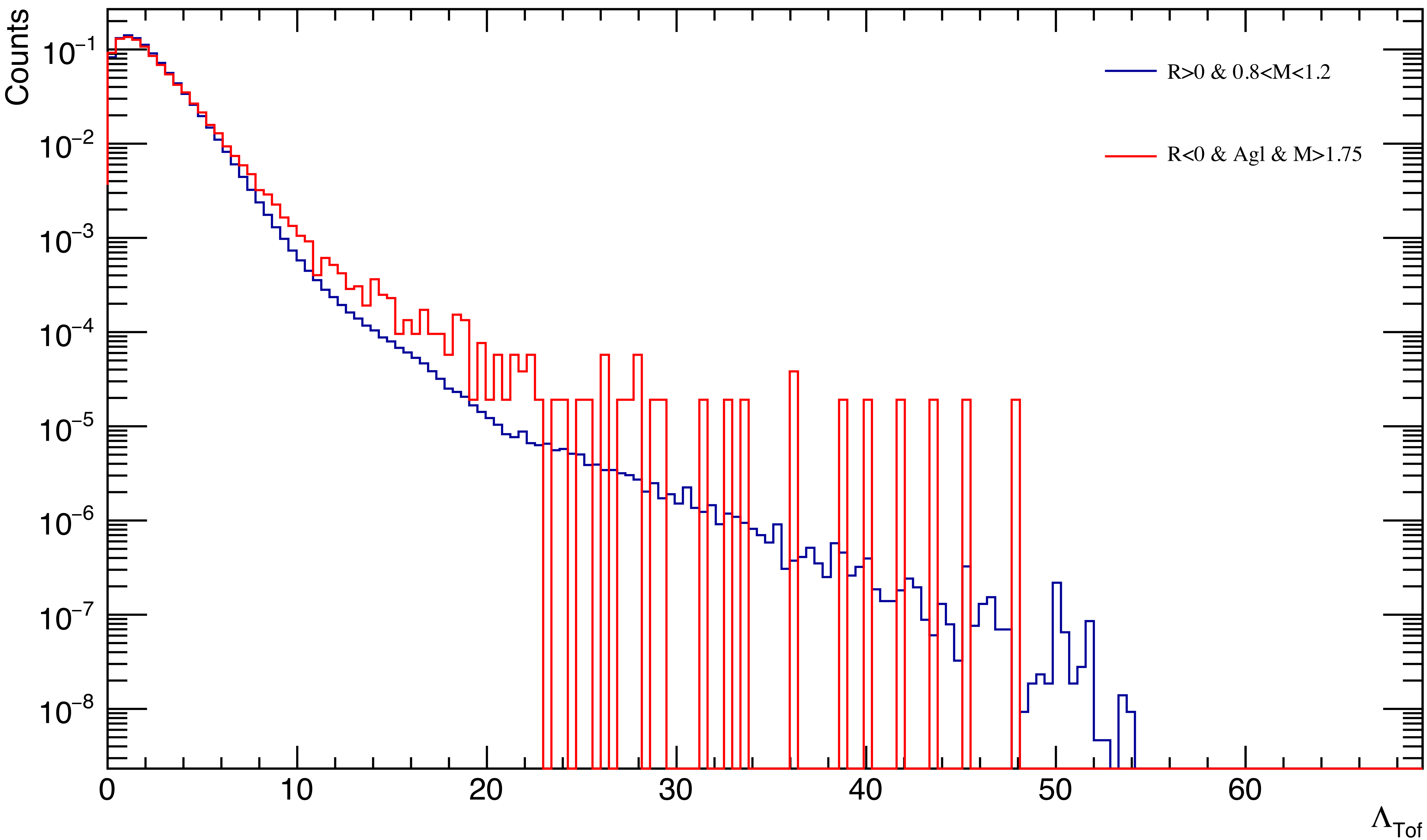
Rigidity consistency



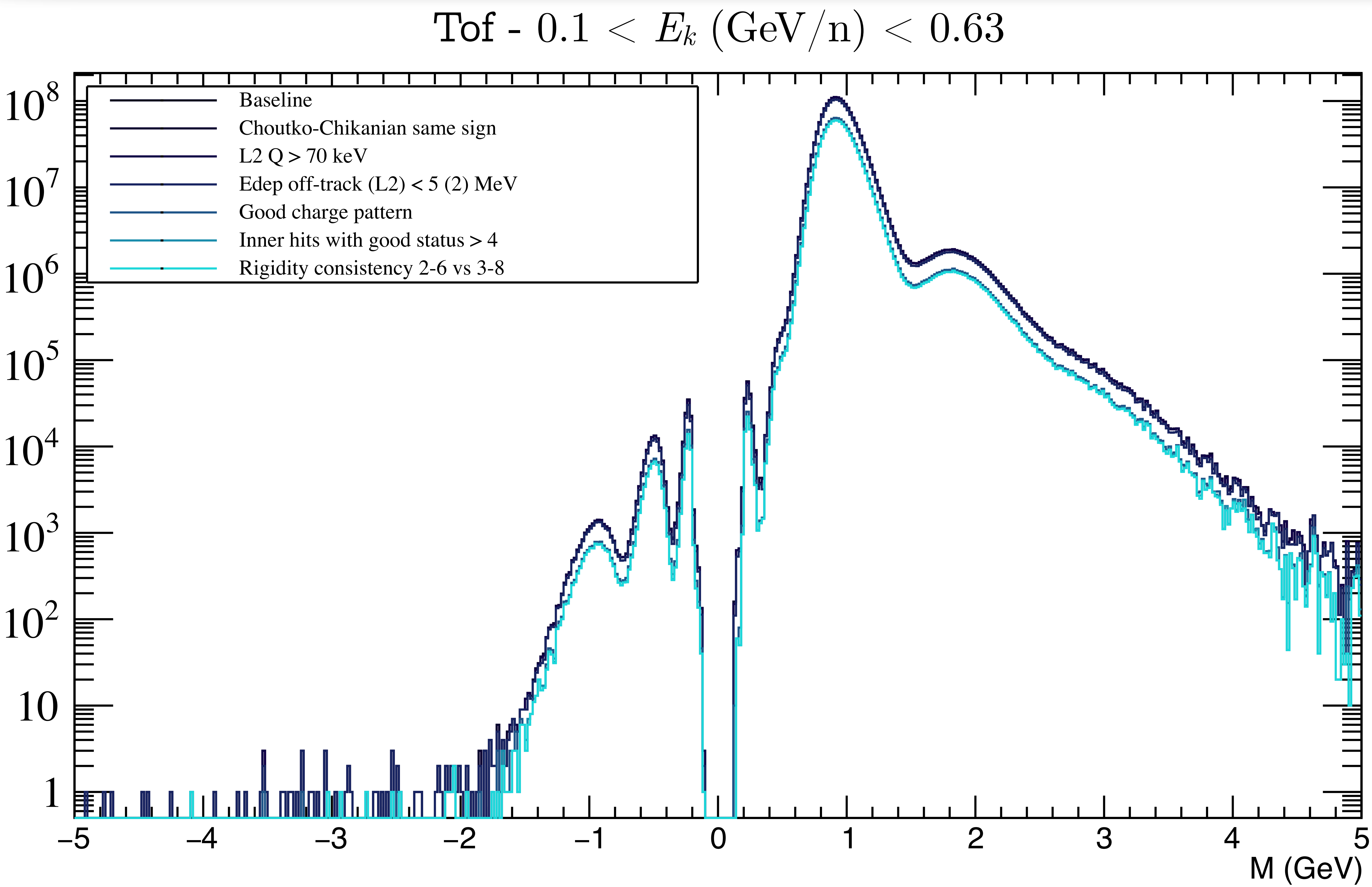
Tracker feet?



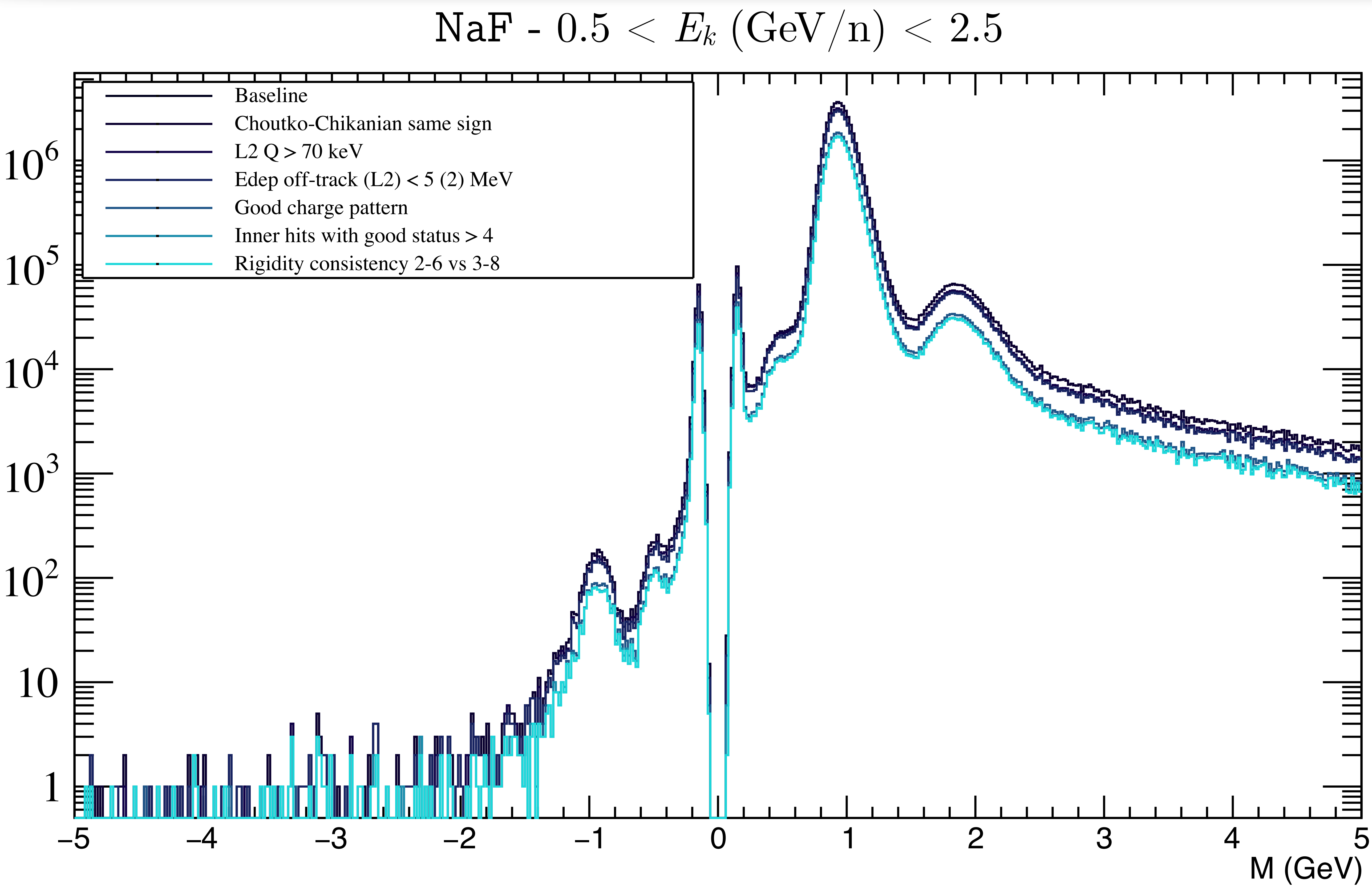
Tof mass likelihood



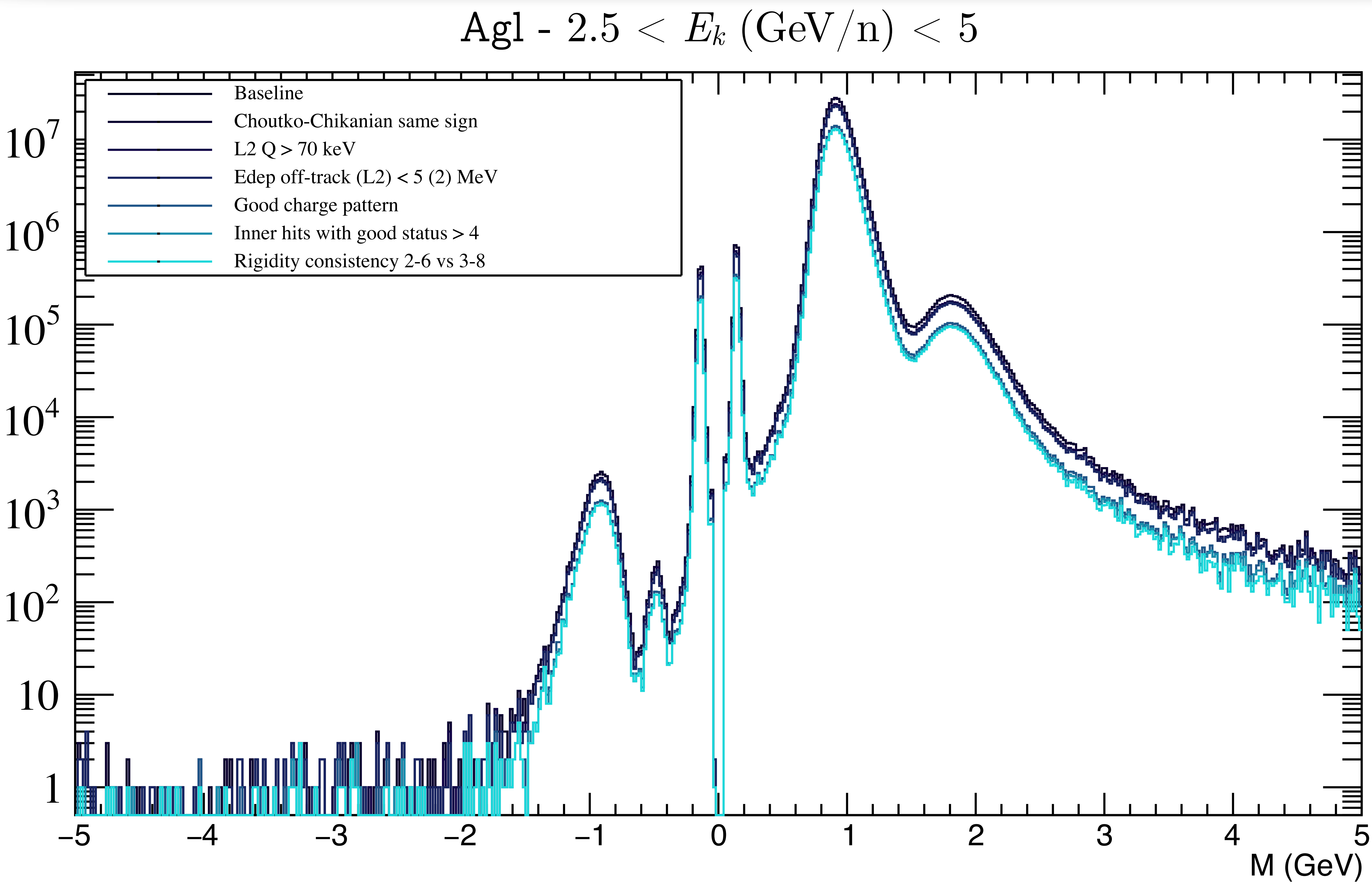
Tof mass distribution



NaF mass distribution



Agl mass distribution



Next

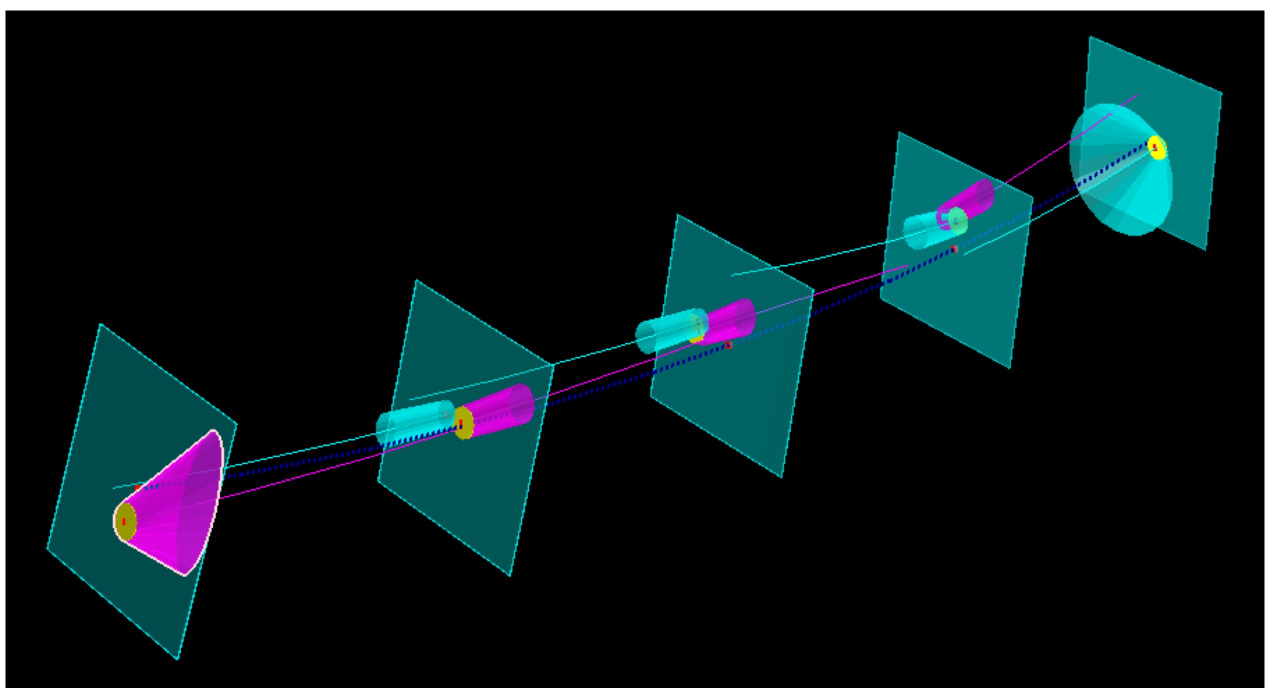
Where to go from here:

- Selection tuning
  - Multiple-scattering dependent variables? - Ciemat/PG
- Multivariate analysis:
  - Beta estimator (BDT regression) - MIT
  - RICH Mass quality (BDT) - Trento
  - Mass classification/estimator (BDT/DNN) - PG
- Mass templates
  - From MC
    - Requires good weights for MC - Ciemat
    - Requires validation of elastic/inelastic xsec
    - Requires actual production
    - Requires (possibly) variables tuning
  - From data
    - Good luck

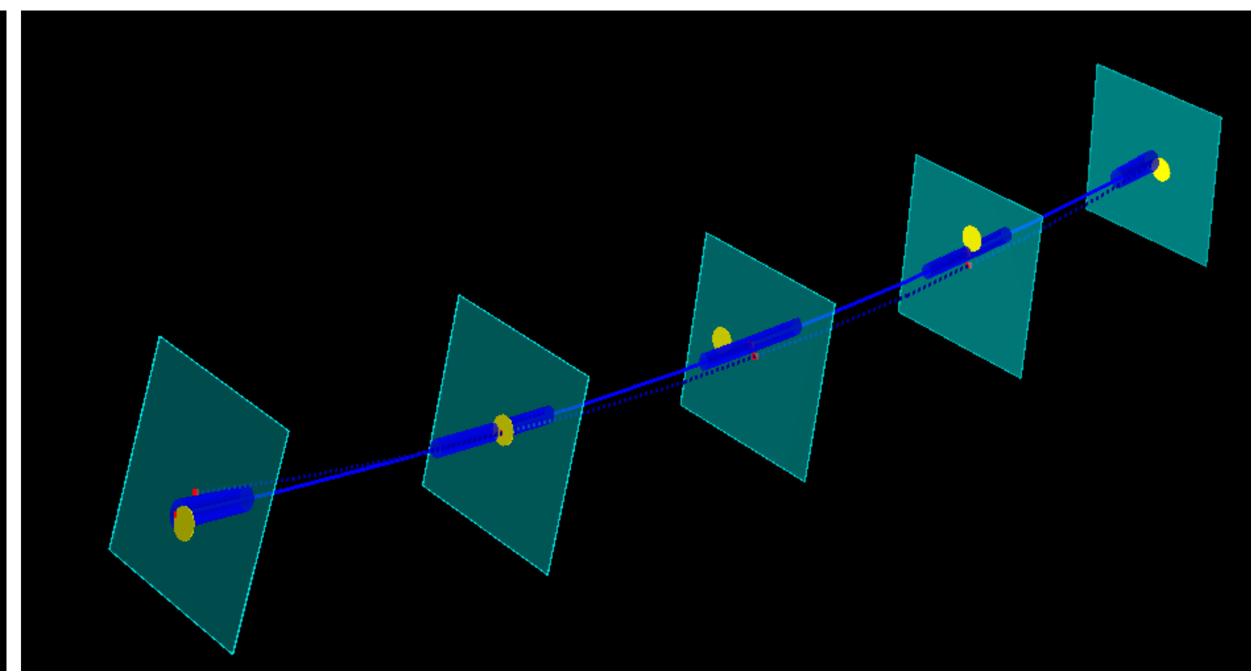
Currently...

At the moment we are updating the dbar DST to v4, which means:

- Bugfixes in the RICH variables
- RICH hit-by-hit beta information
- Multiple scattering naive estimator
- New track fitting by QY (GENFIT+Kalman)



forward and backward fitting



average smoothing

Höppner, C. and Neubert, S. and Ketzer, B. and Paul, S. , *A novel generic framework for track fitting in complex detector systems*, *Nuclear Inst. and Methods in Physics Research,* A **620.2**, (2010) 518-525

Tools

We are currently setting up a real "Analysis Train":

- We will run the analysis code every night on the pass6 and MC
  - …loading each user's code (provided it compiles) and running it
  - …sharing resources on lxplus/CNAF (using pybatch)
  - …keeping the output objects separate (single ROOTfile with one directory for each plugin)

This is nothing new (but it shows that such an idea actually makes sense…)

The ALICE analysis train system

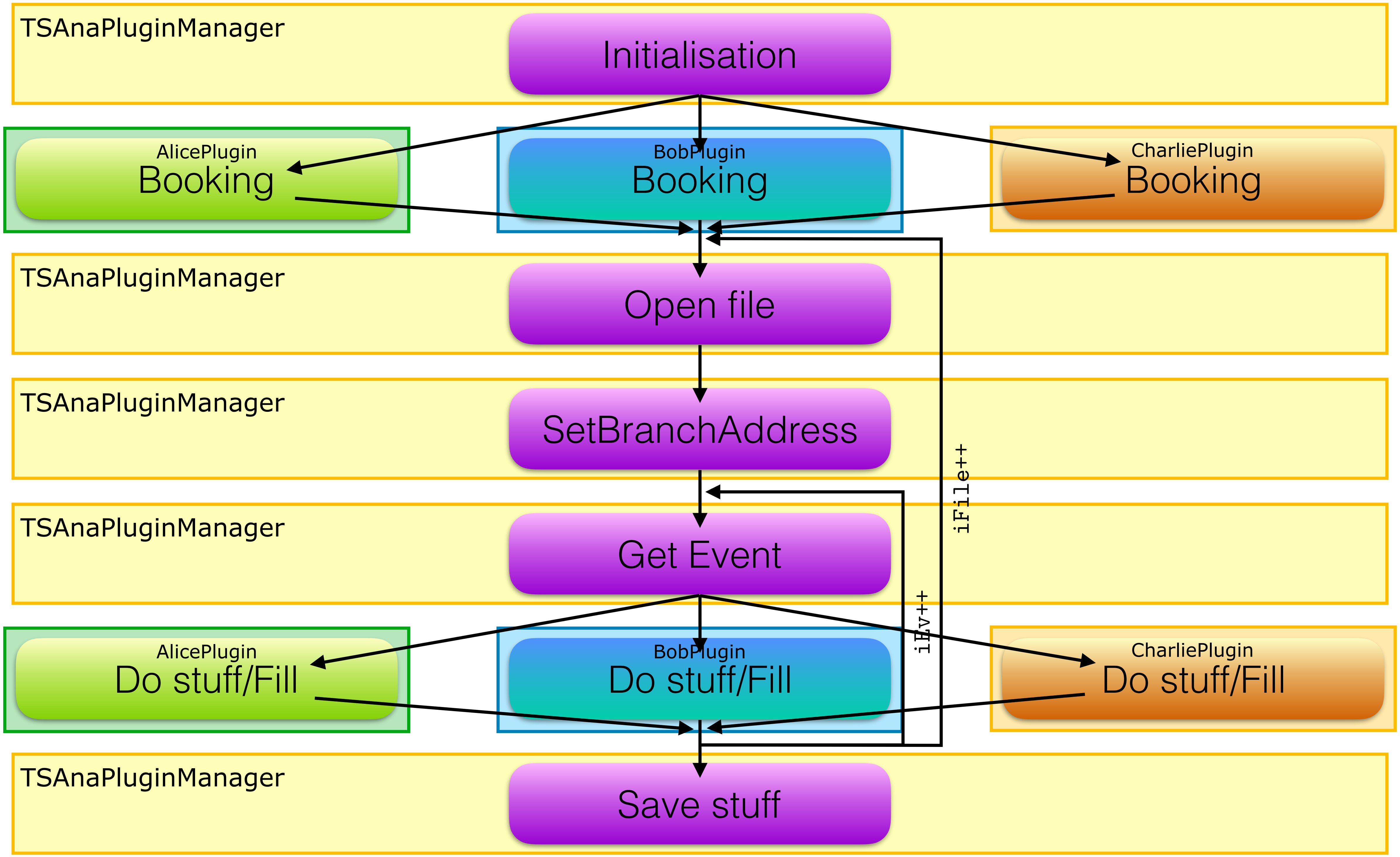
Markus Zimmermann for the ALICE collaboration
CERN, 1211 Geneva 23, Switzerland
E-mail: `m.zimmermann@cern.ch`

Abstract. In the ALICE experiment hundreds of users are analyzing big datasets on a Grid system. High throughput and short turn-around times are achieved by a centralized system called the LEGO trains. This system combines analysis from different users in so-called analysis trains which are then executed within the same Grid jobs thereby reducing the number of times the data needs to be read from the storage systems. The centralized trains improve the performance, the usability for users and the bookkeeping in comparison to single user analysis. The train system builds upon the already existing ALICE tools, i.e. the analysis framework as well as the Grid submission and monitoring infrastructure. The entry point to the train system is a web interface which is used to configure the analysis and the desired datasets as well as to test and submit the train. Several measures have been implemented to reduce the time a train needs to finish and to increase the CPU efficiency.

https://arxiv.org/pdf/1502.06381.pdf

Tools

Alice & Bob & Charlie



Rediscover physics