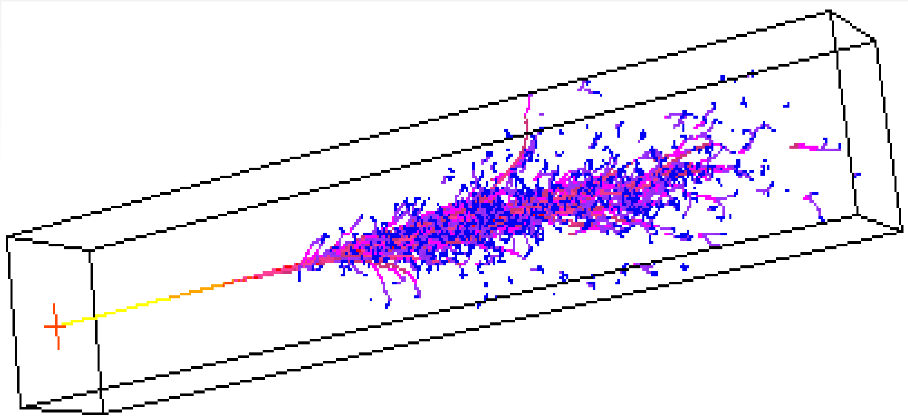
Charged and neutral particle ID with the EMC



Ronald Kunne
IPN Orsay, France
CNRS/IN2P3 – Université Paris Sud

- What are Bayesian classifiers ?

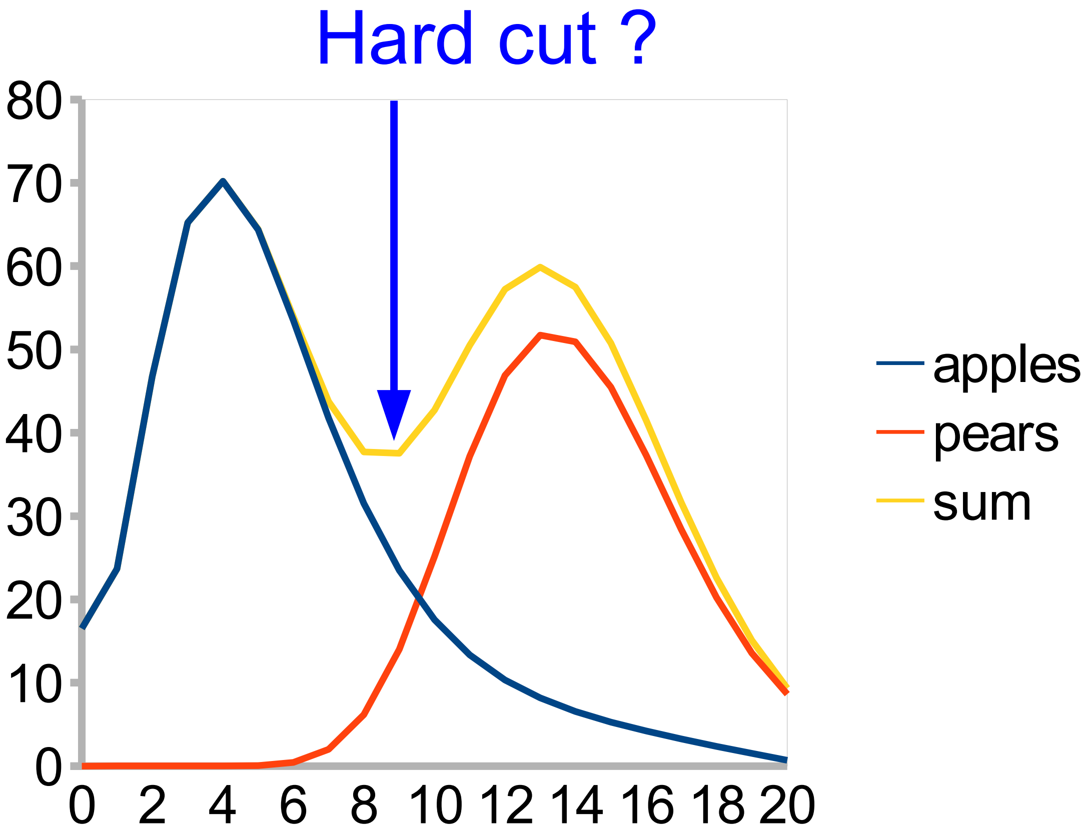
- Charged PID

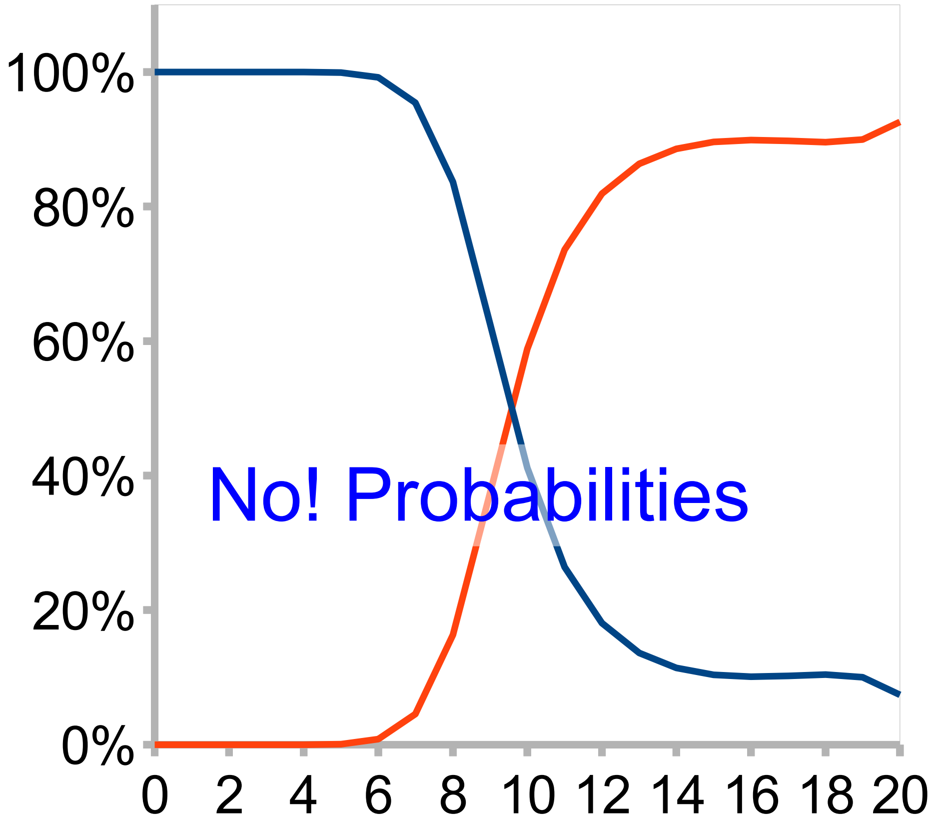
- Neutral PID

- PandaRoot Version : 12490 (26/6/11)
- externals: may 2011

What are Bayesian classifiers ?

Hard cuts are not optimal





Enter Bayes' Theorem

$$P(e \mid \text{data}) = \frac{P(\text{data} \mid e) * P(e)}{P(\text{data} \mid e) * P(e) + P(\text{data} \mid \pi) * P(\pi) + \cdots}$$

Enter Bayes' Theorem

$$P(e \mid data) = \frac{P(data \mid e) * P(e)}{P(data \mid e) * P(e) + P(data \mid \pi) * P(\pi) + \cdots}$$

$$= \frac{P(data \mid e)}{P(data \mid all)} \; \frac{P(e)}{P(all)}$$

Enter Bayes' Theorem

$$P(e \mid data) = \frac{P(data \mid e) * P(e)}{P(data \mid e) * P(e) + P(data \mid \pi) * P(\pi) + \cdots}$$

$$= \frac{P(data \mid e)}{P(data \mid all)} \quad \frac{P(e)}{P(all)}$$

$P(all)$ = 1

When the data are uncorrelated we can write :

$$P(data \mid all) = P(d_1 \mid all) * P(d_2 \mid all) * \ldots..$$

Bayesian classifiers

P(e | data) =

$$\frac{P(d_1 \mid e)}{P(d_1 \mid all)} * \frac{P(d_2 \mid e)}{P(d_2 \mid all)} * \frac{P(d_3 \mid e)}{P(d_3 \mid all)} * P(e)$$

Bayes' factor (from MC)

From flux, other data

Very practical !

P(e | data) =

$$\frac{P(d_1 \mid e)}{P(d_1|all)} * \frac{P(d_2 \mid e)}{P(d_2|all)} * \frac{P(d_3 \mid e)}{P(d_3|all)} * P(e)$$

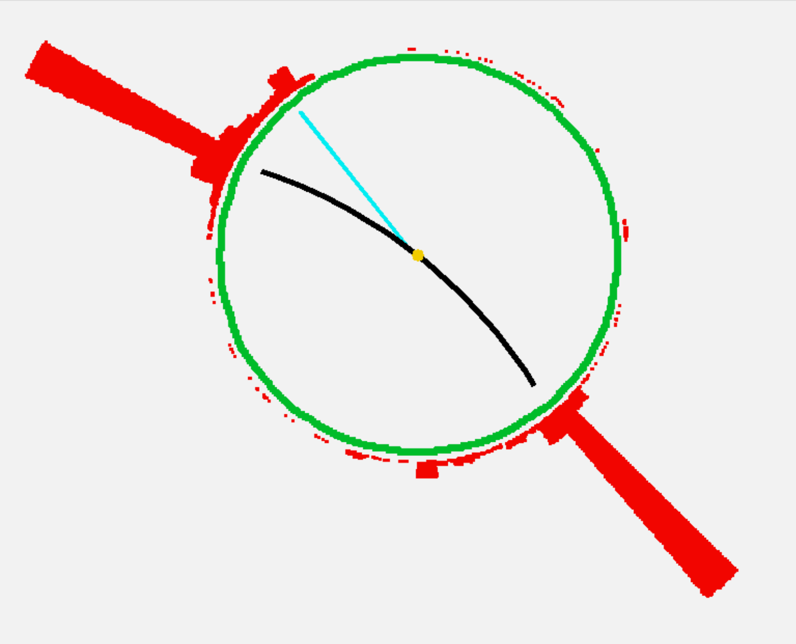
Any number of variables

Classifiers work with correlated data too

Have to take out E **=>** *always use (E,z) plane*

!

Charged particle ID



- Problem :

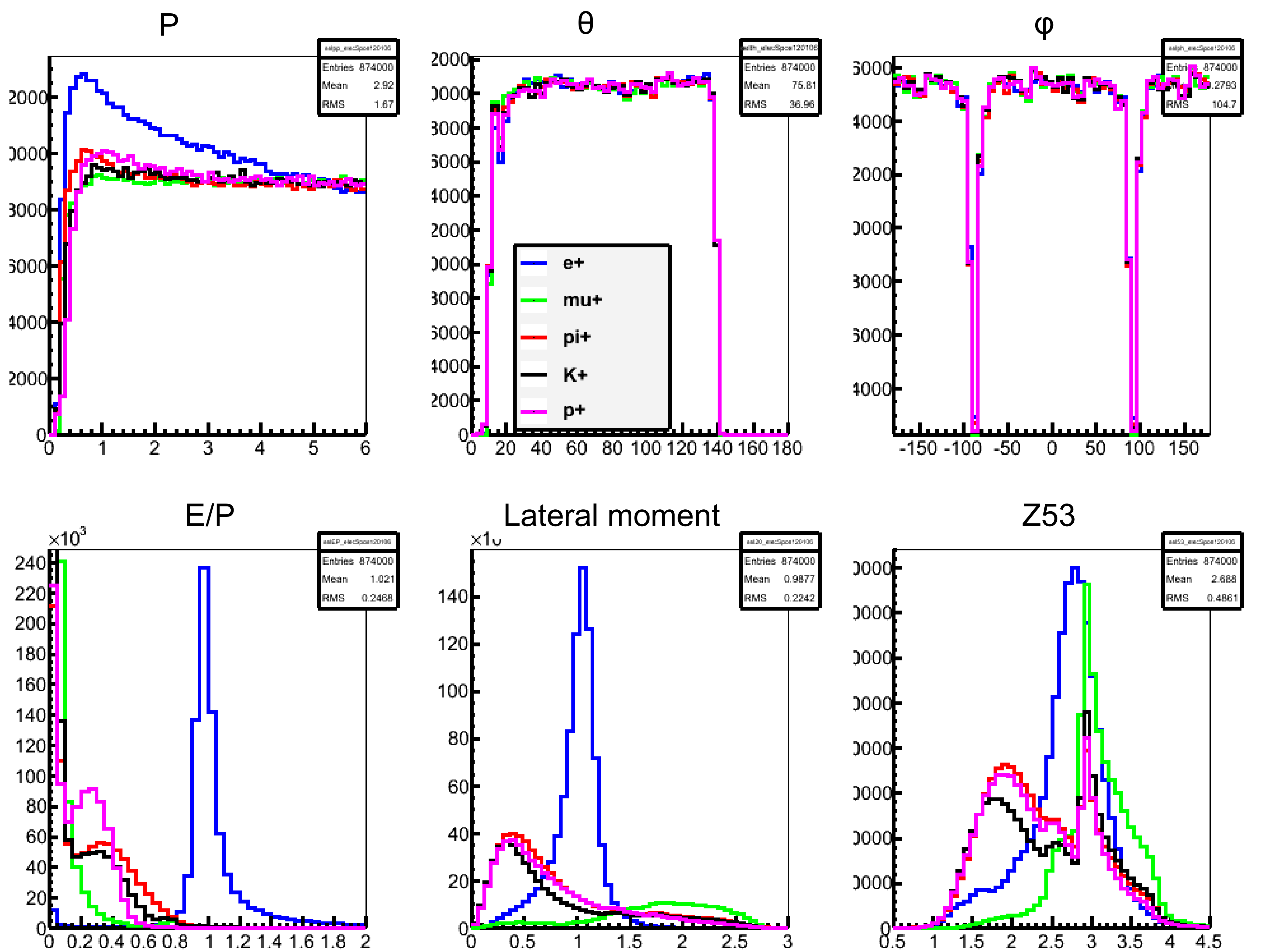
  assign PID probabilities to tracks using EMC variables like E/P, Zernike moments

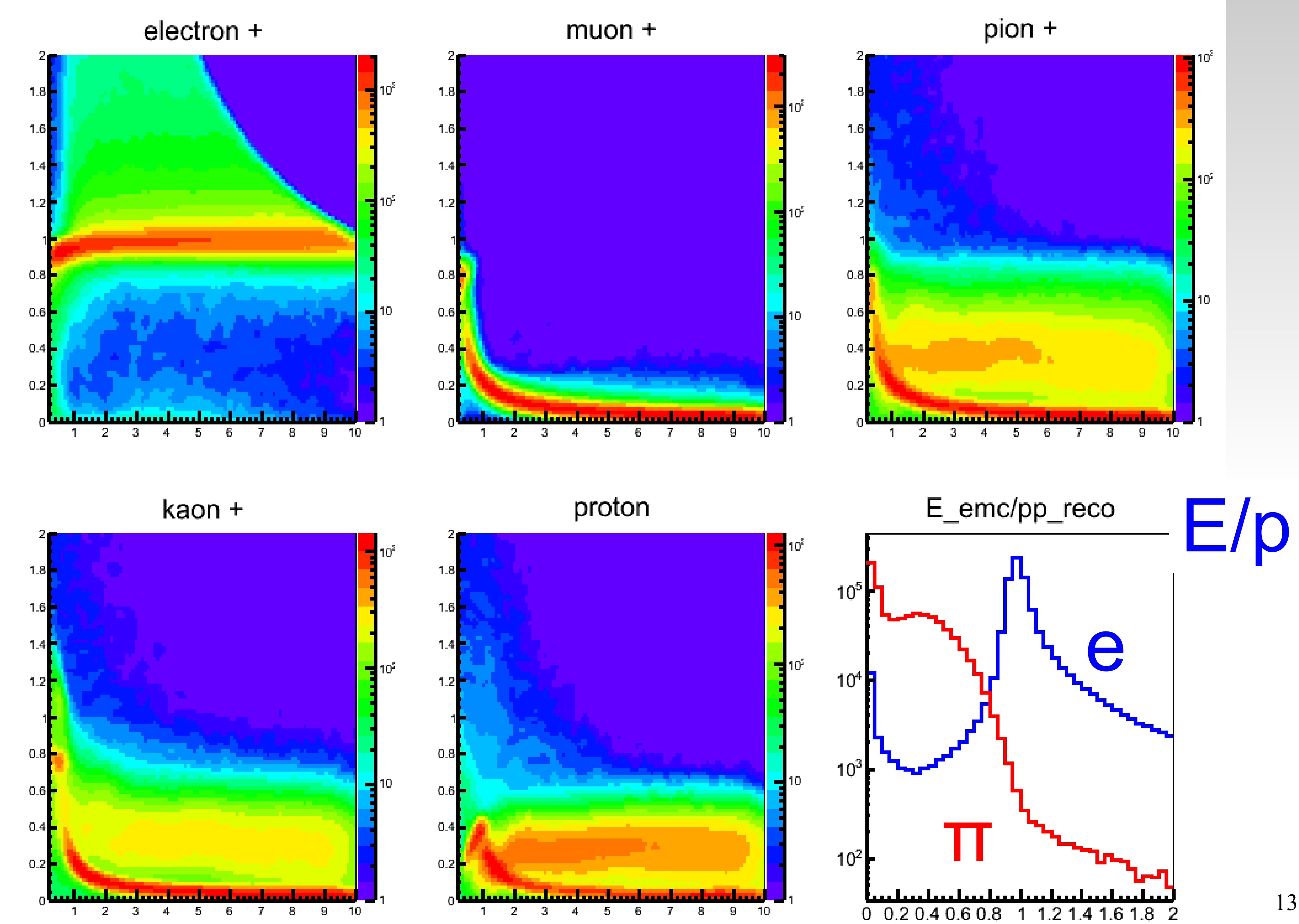
Procedure

- Generate $e^{\pm}$, $\mu^{\pm}$, $\pi^{\pm}$, $K^{\pm}$, $p^{\pm}$ tracks, with

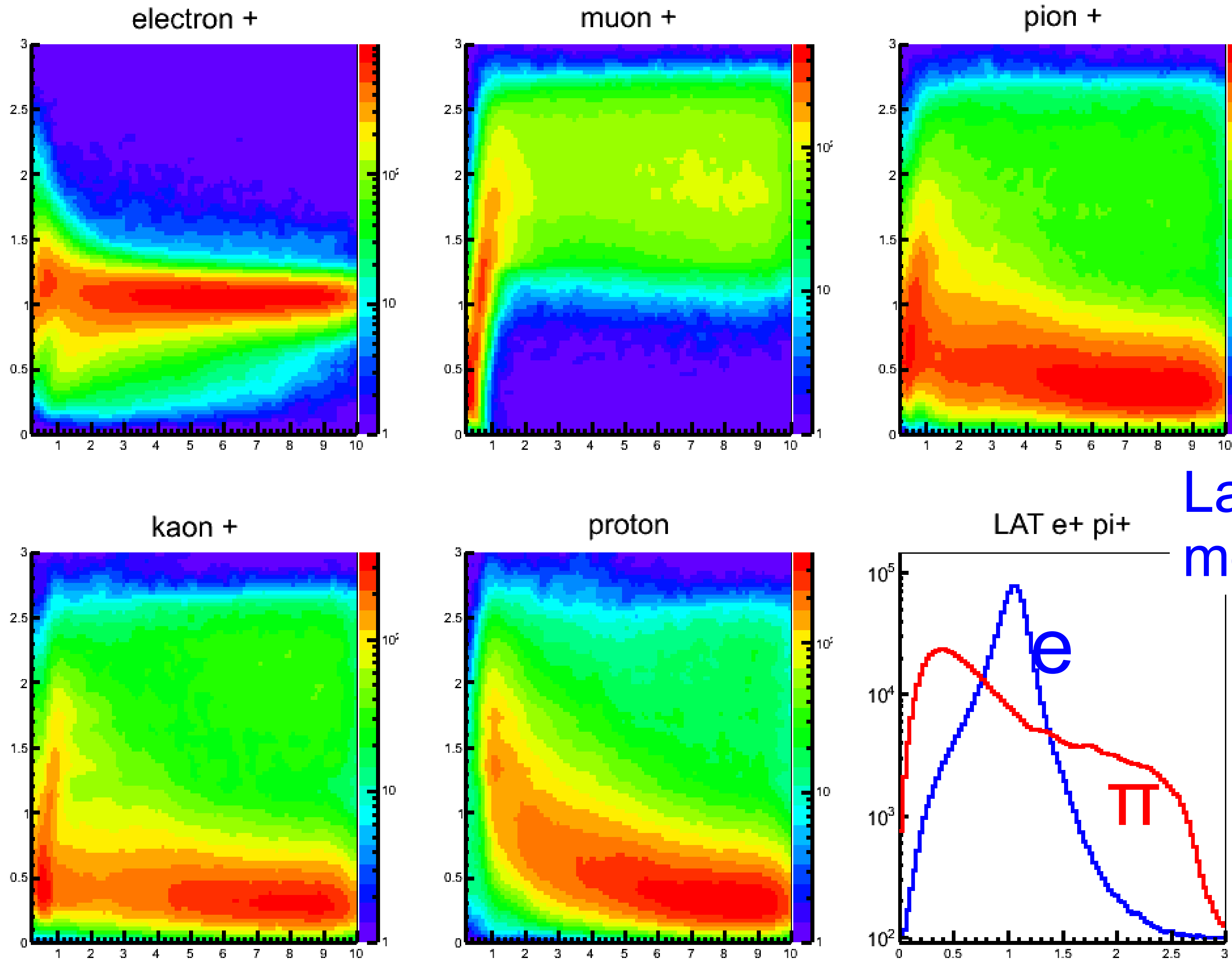
  $0.2 < p < 10$ GeV/c, $0° < \varphi < 360°$, $5° < \Theta < 140°$

- Run reconstruction macros μ hypothesis
- Save p_reco, φ, Θ, E_emc, Lateral moment, Z53, ...
- Calculate probabilities
- Assign all events using probabilities and check result.

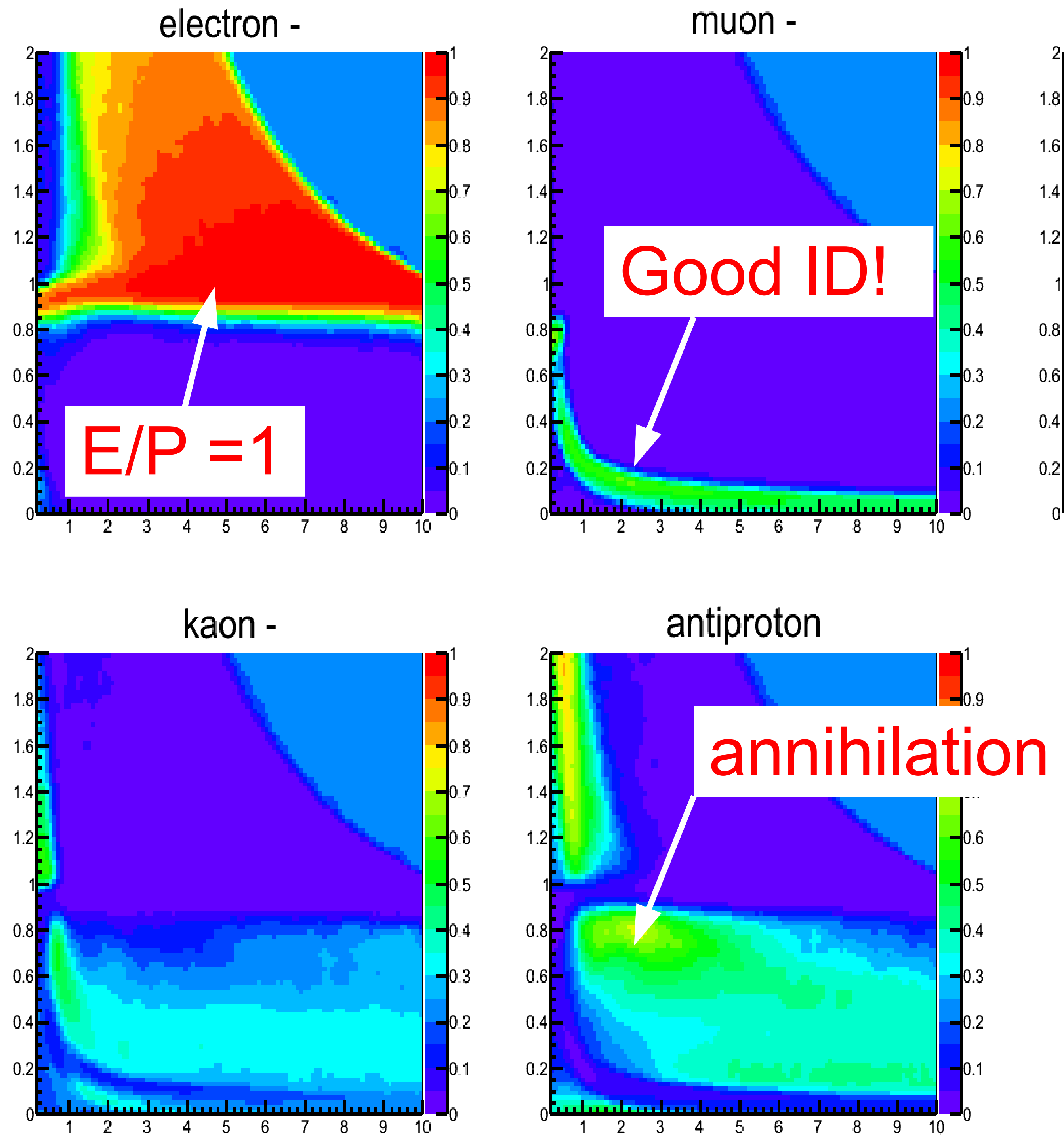

P
Entries 874000
Mean 2.92
RMS 1.67
θ
Entries 874000
Mean 75.81
RMS 36.96
e+
mu+
pi+
K+
p+
φ
RMS 104.7
E/P
Entries 874000
Mean 1.021
RMS 0.2468
Lateral moment
Entries 874000
Mean 0.9877
RMS 0.2242
Z53
Entries 874000
Mean 2.688
RMS 0.4861



electron +
muon +
pion +
kaon +
proton
E_emc/pp_reco
E/p
e
π



Lateral moment



Probabilities

Overall efficiencies

MC Input

Reco& PID

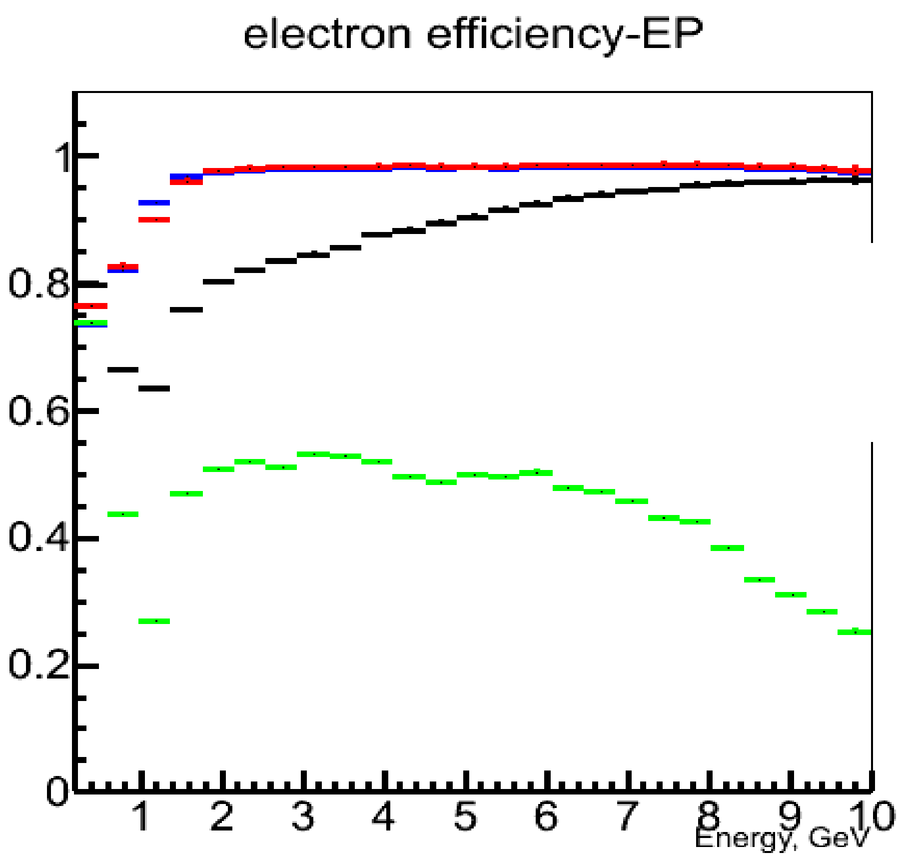
	e	μ	π	K	p
e	95	0	2	2	2
μ	0	97	36	41	24
π	2	1	23	15	11
K	1	1	3	5	3
p	2	1	36	37	60

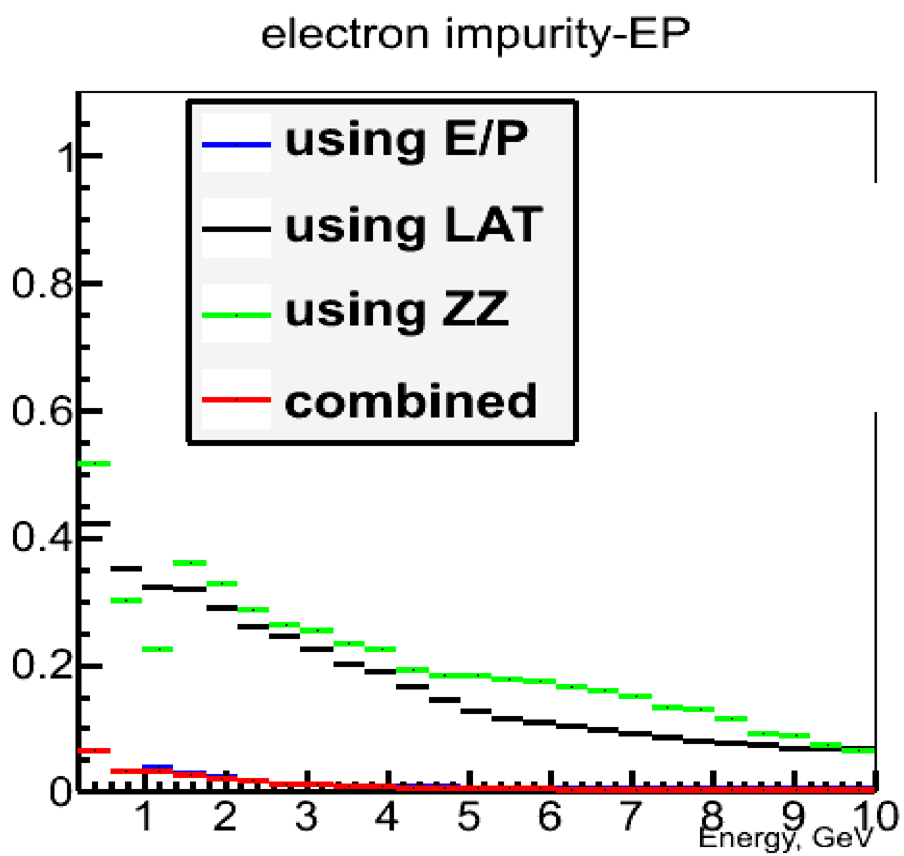
e and μ OK

μ impure

Confusing the hadrons

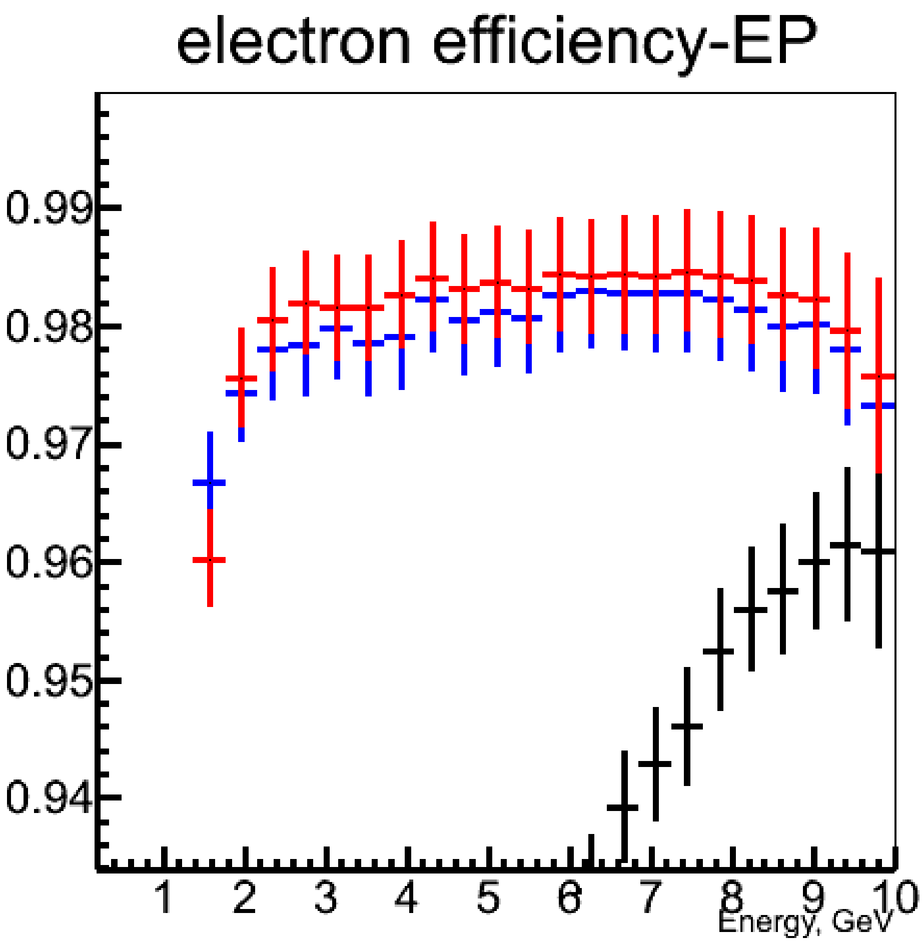
Efficiency et impurity

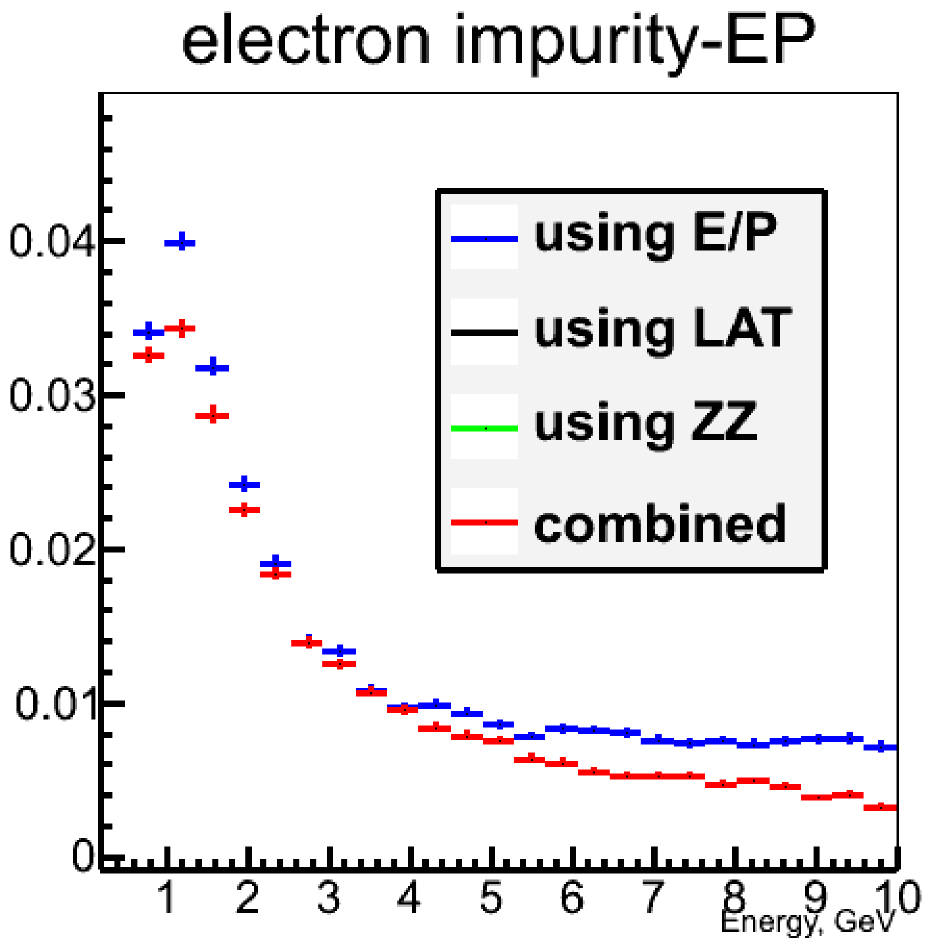




- E/P does most of the job
- Low E problematic

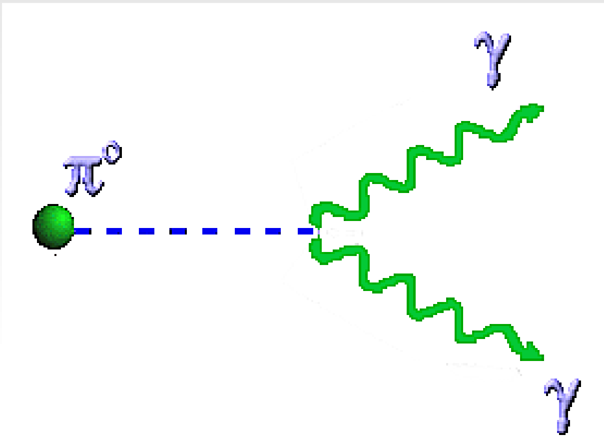
Efficiency et impurity





- 98% efficiency above 2 GeV
- 1% impurity per track

Neutral particle ID

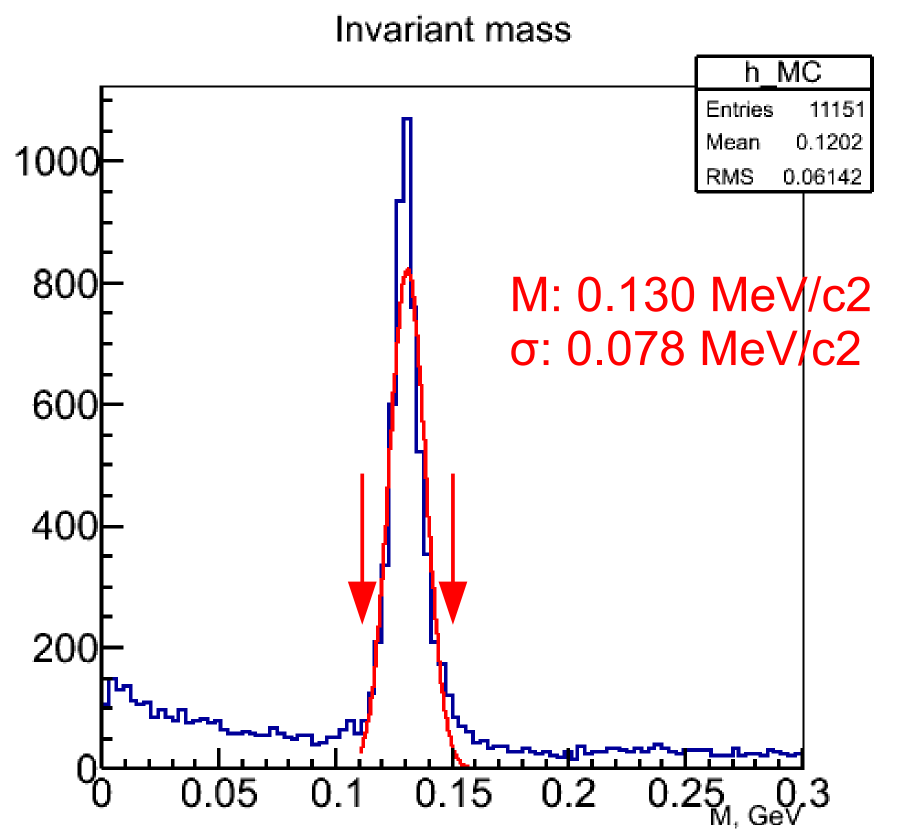


•Problem : π°'s with small opening angles do not resolve and can not be distinguished from γ's

Step1 : reconstruction from two γ

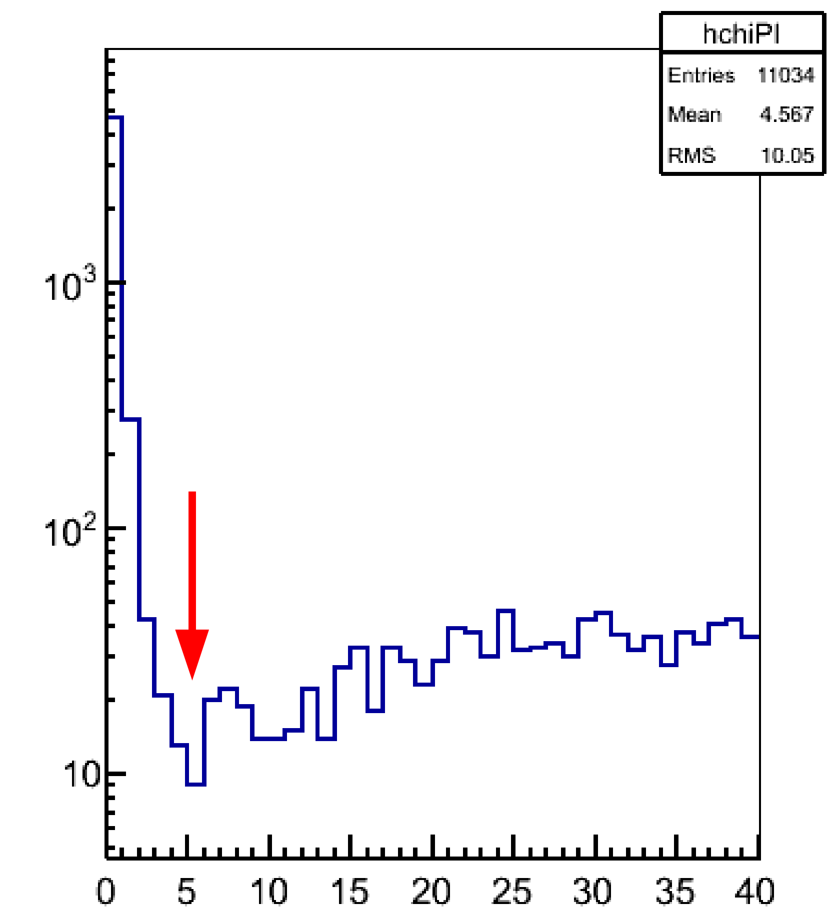
- Reconstruct π°'s from two bumps using the information available in PndEmcBump

- Cut on invariant mass $0.11 < M_{\pi^{\circ}} < 0.15$ GeV



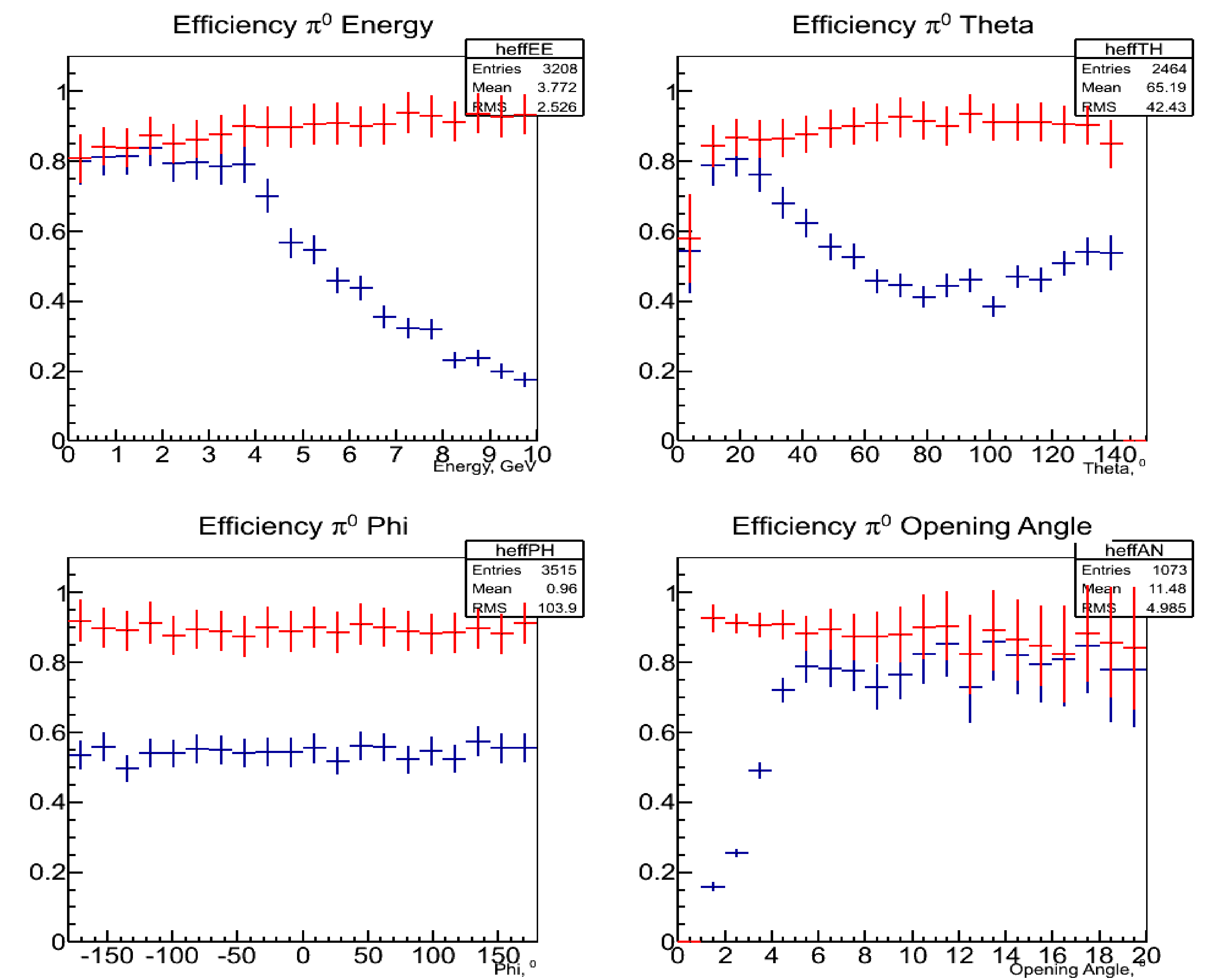
Correlation MC π°- reconstructed π°

- Succesfull reconstruction if $\chi^2 = d\theta^2 + d\varphi^2 < 5\ deg^2$ when compared to MC values



Overall efficiency : About 50%

Reconstruction efficiency

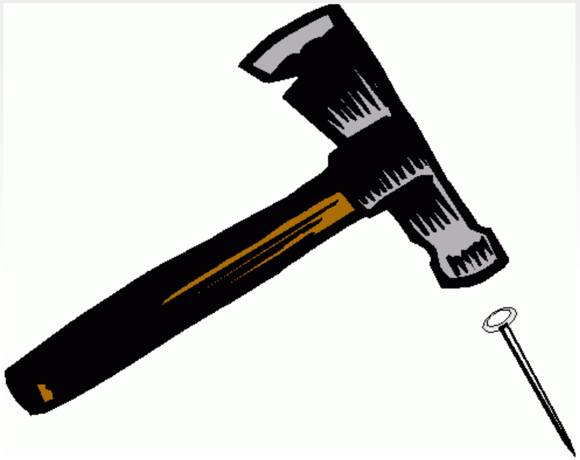


Theoretical efficiency when all unused clusters are taken as π°:

About 85%

Can we recuperate these, without admitting too much gammas?

Step 2 : recuperate unsplit clusters



- Construct Bayesian classifiers using the Zernike moments

Cluster Moments

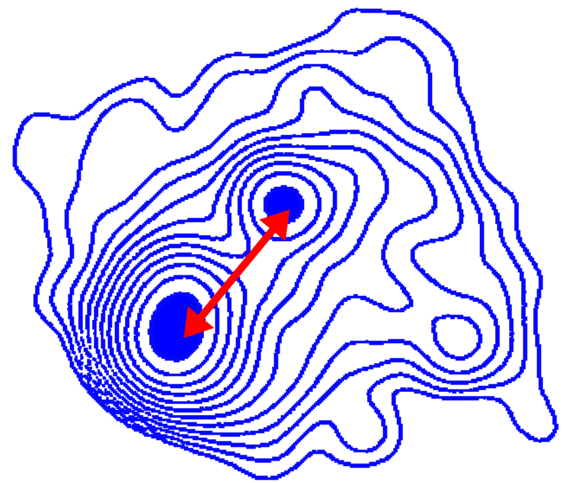
- Two extra "moments" are defined :

  - Shower mass : $(\Sigma E_i)^2 - (\Sigma p_i)^2$

  - Correlated maximum separation :

  $\max [E_i E_j (d\theta_{ij}^2 + d\varphi_{ij}^2)]$

  (where the indices run over the cluster hits)



Sixteen moments tested

Z00	Z20	Z40	Z11
Z31	Z51	Z22	Z42
Z62	Z33	Z53	Z44
Z64	LatMom	ShowerMass	CorMax

- Not all moments are efficient classifiers
- Many are correlated between them

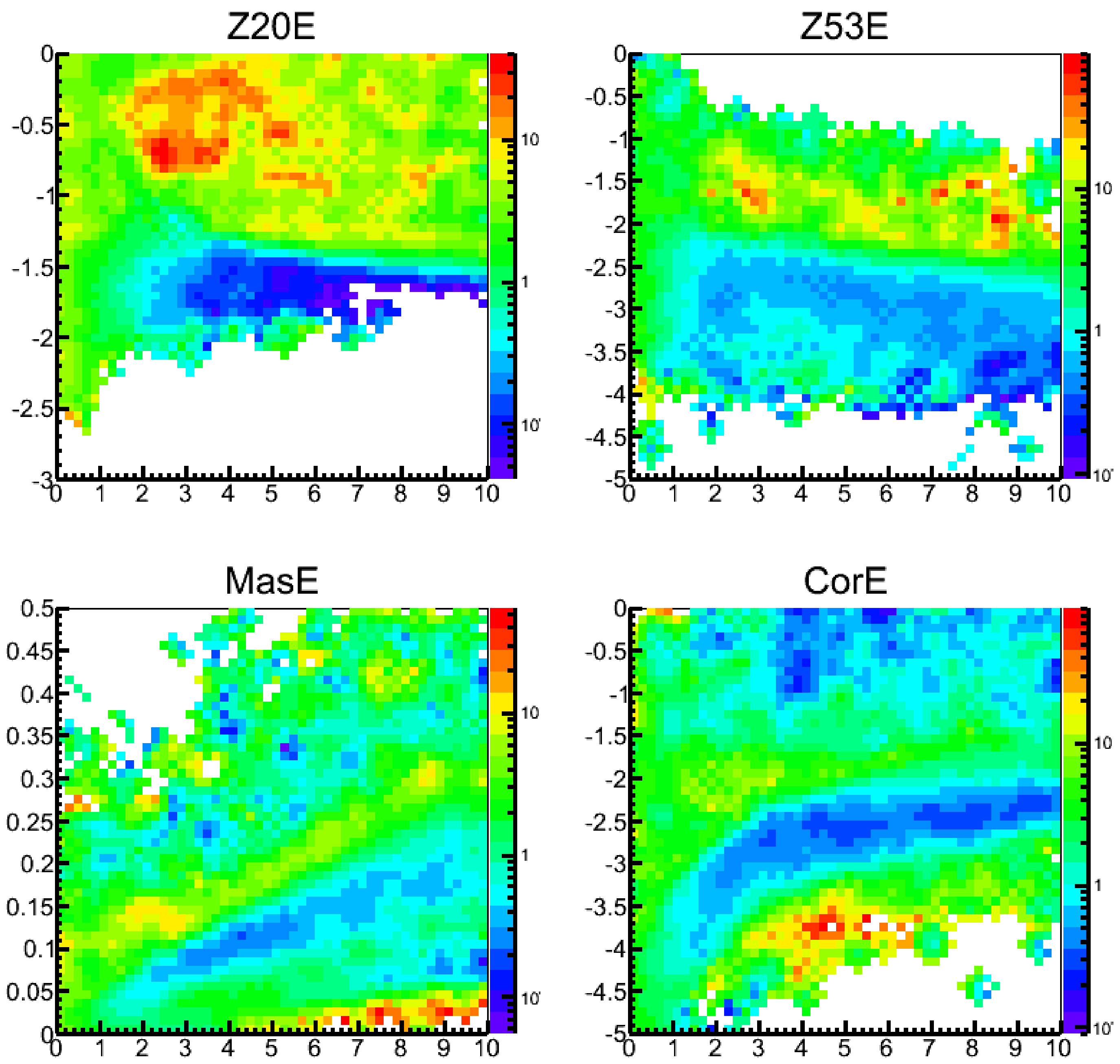
Construction of a "Naive" Bayesian classifier

- Procedure :

  - Create 2d-histos with ClusterEnergy E versus moment z for γ and $\pi°$

  - Factor(E,z) = N (E,z |$\pi°$) / N(E,z |γ)

  - classifier = log(factor1) + log(factor2) + …

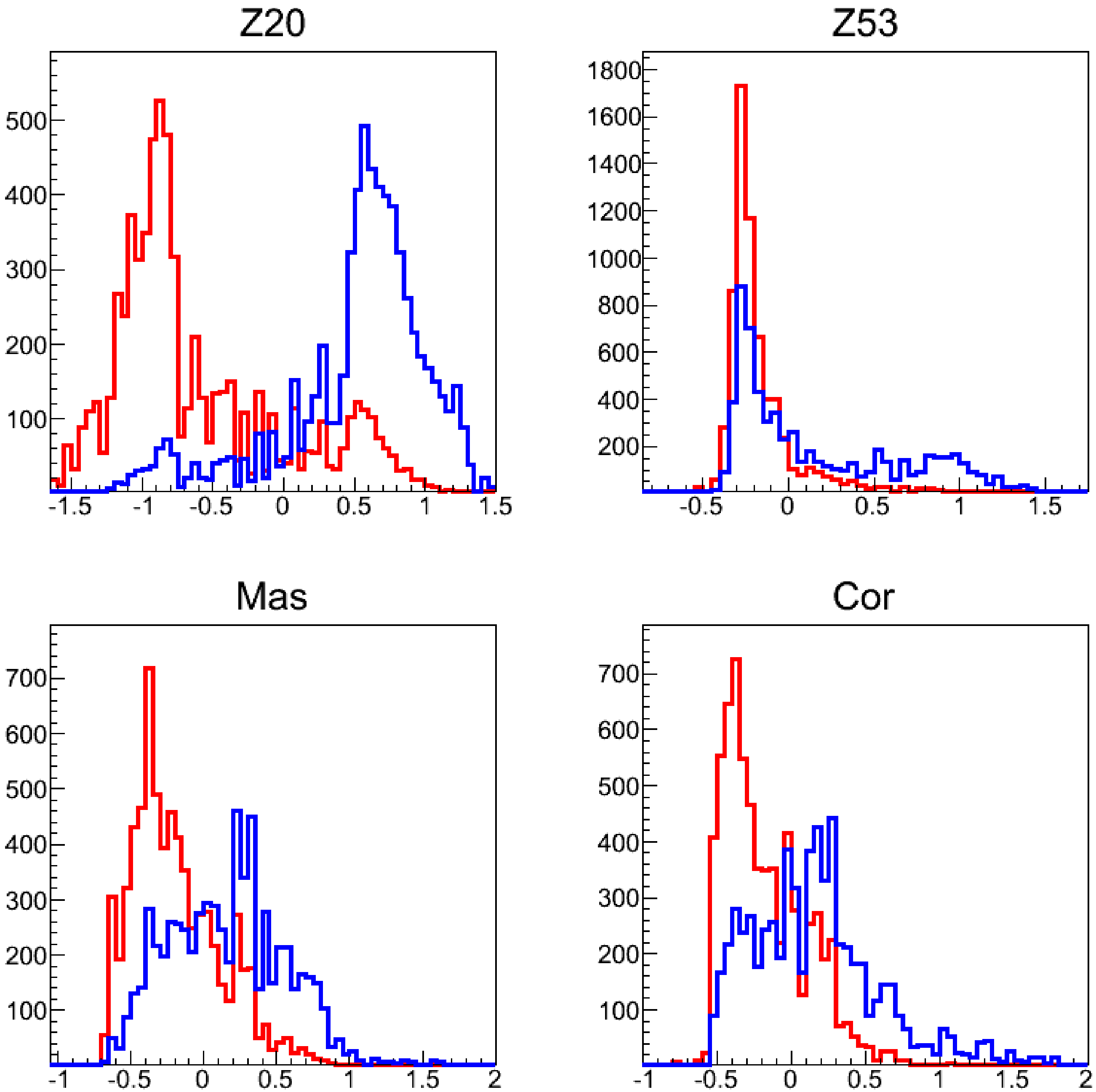


Factor (E,z)

Blue : more γ than π°

Red : more π° than γ

Green : about equal

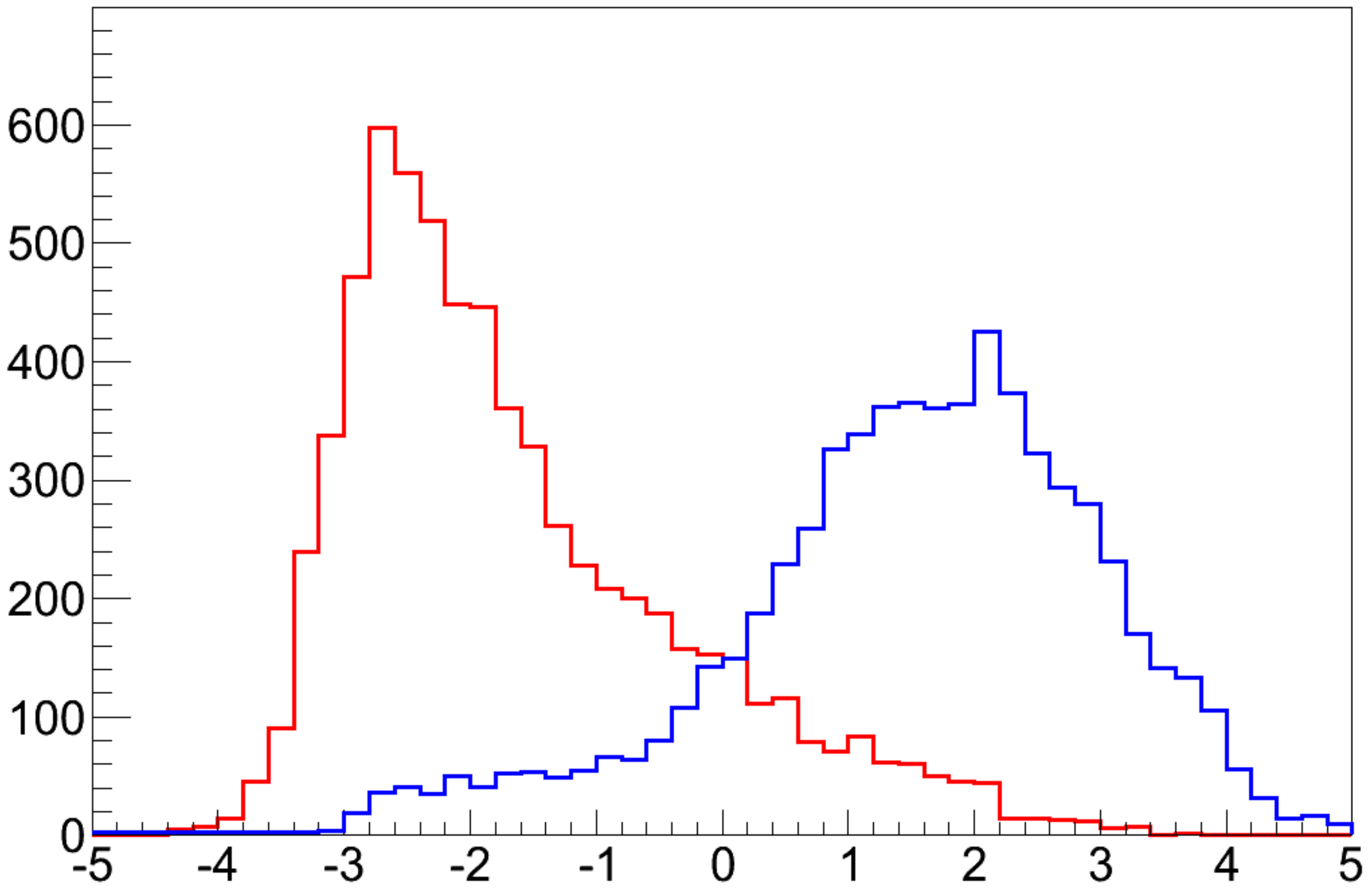


events as function of log(factor)

Blue : π°

Red : γ

Bayesian classifier = Σ log(factor)



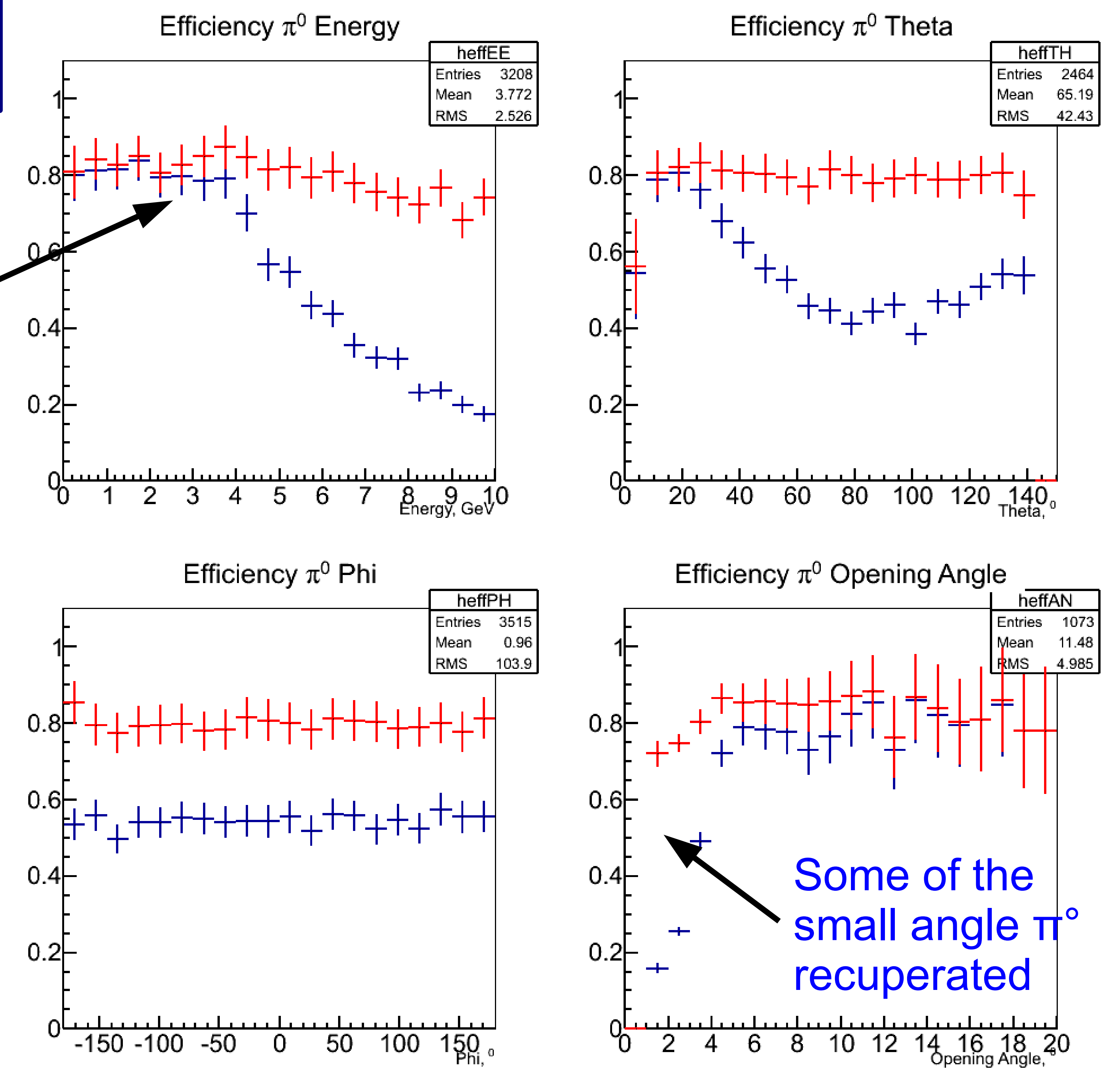
events
as function of
Bayesian
classifier

Blue : π°
Red : γ

Efficiency π° : five moments

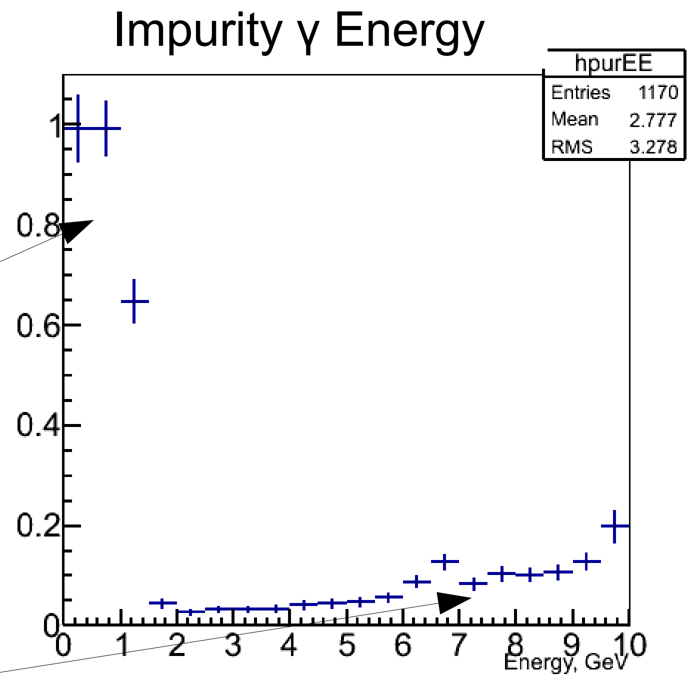
Not needed below 4GeV/c

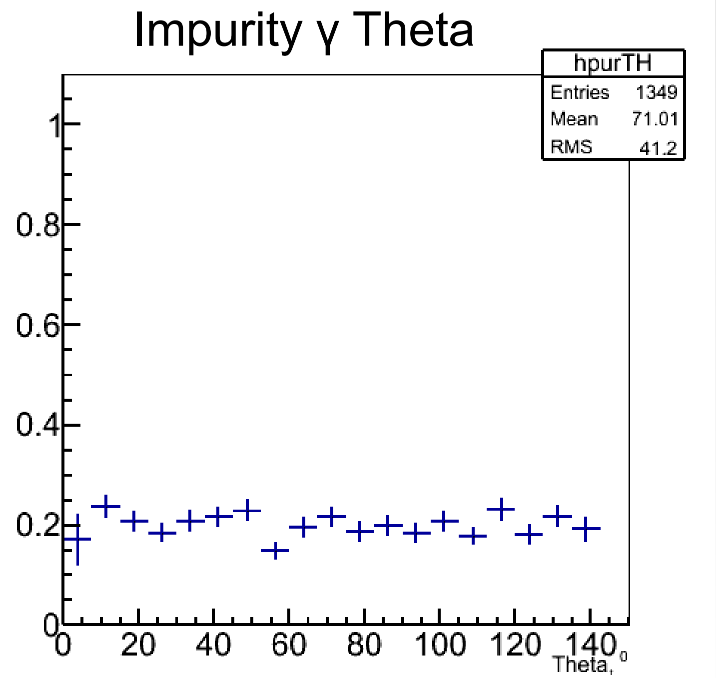
80% total efficiency



Some of the small angle π° recuperated

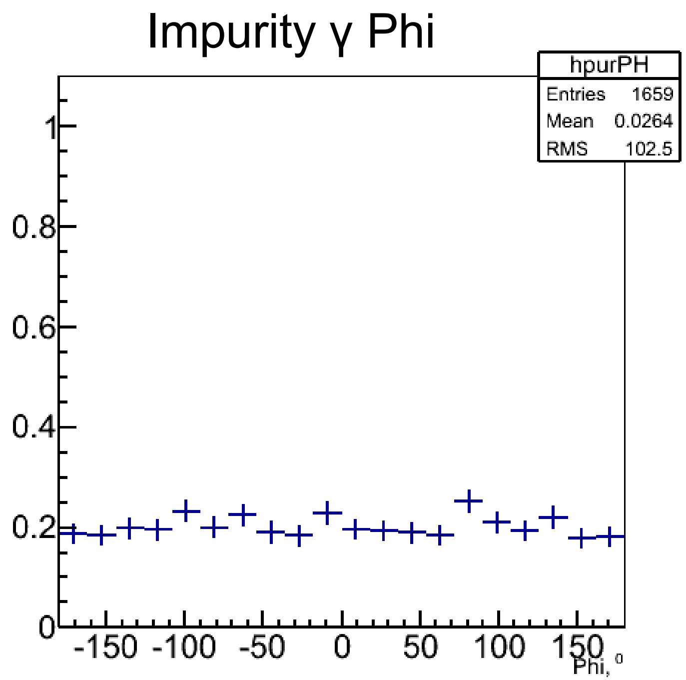
Impurity from gammas





Harmful below 2GeV/c

5-20% above 2 GeV/c



Fraction of γ accepted as π°

Conclusions

- Charged particle ID:

  - Bayesian classifiers combine E/P, LatMom, Z53

  - 98% efficiency for e and μ, above 2 GeV/c

  - 1% impurity for Π, above 3 GeV/c

- Π° reconstruction using bumps :

  - loss due to unresolved clusters, can be recuperated with cluster moments

  - price to pay : 5-20% of the single γ pass the test

Thank you for your attention !

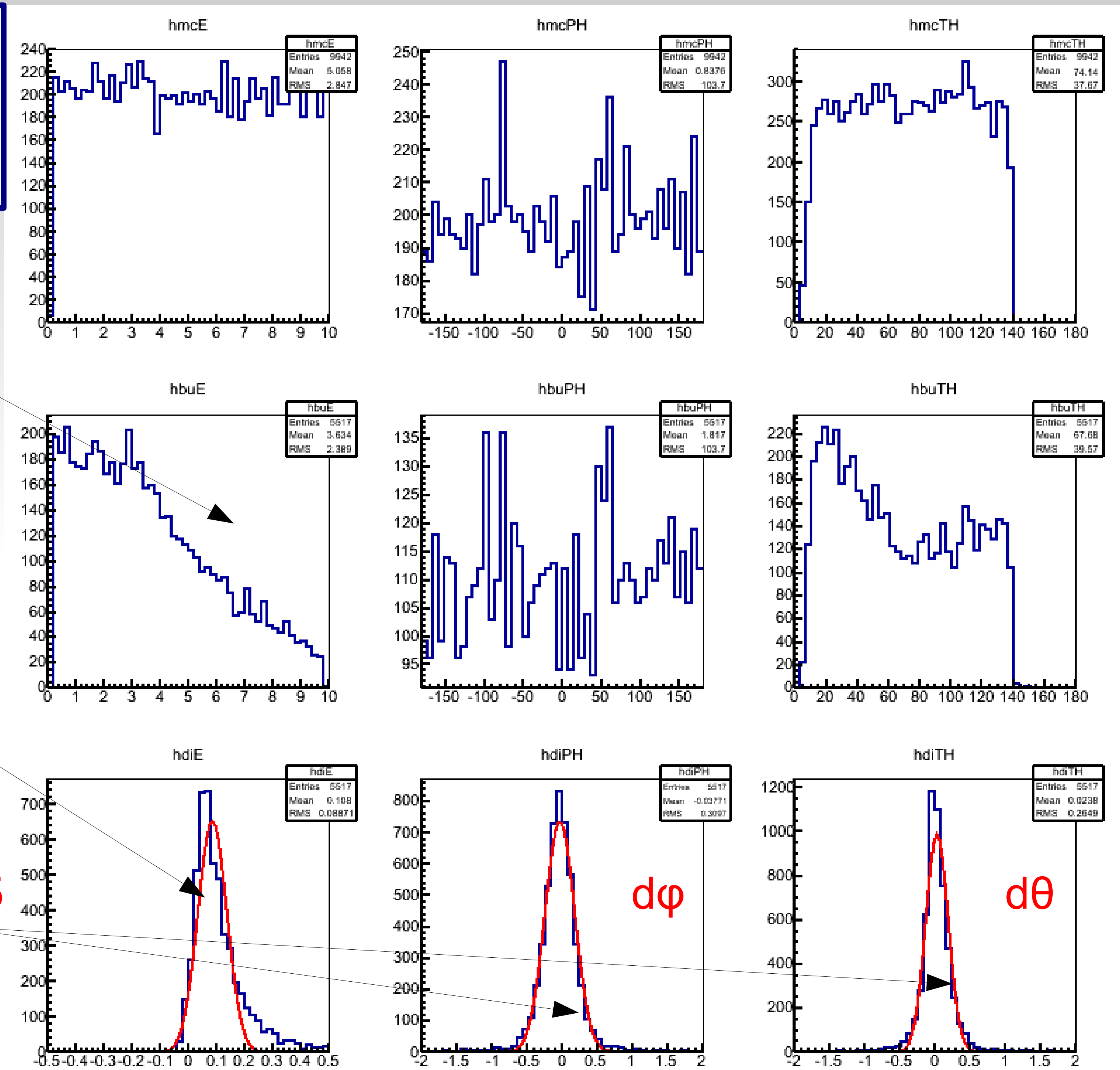
Correlation
MC π°-
reconstructed
π°

Losses !

10K
events

Using
Bump
energies

$\chi^2 = d\theta^2 + d\varphi^2 < 5$



MC

Reco

Residu