Preliminary study of a Boosted Decision Tree (BDT) for particle identification in KLOE

Niels Vestergaard - Muon general meeting 10/2 2025

UNIVERSITY OF LIVERPOOL

LEVERHULME TRUST

KLOE processes

In KLOE, we are measuring the total cross section:

$$\sigma^0_{\pi\pi(\gamma)} = \frac{4\pi\alpha^2}{3s'}(1 + 2m_\mu^2/s')\beta_\mu \frac{d\sigma_{\pi\pi\gamma}/ds'}{d\sigma_{\mu\mu\gamma}/ds_\mu(1 - \eta_\mu(s_\mu))}$$

As such, the two main processes are: $e^+e^- \to \pi^+\pi^-\gamma$ and $e^+e^- \to \mu^+\mu^-\gamma$

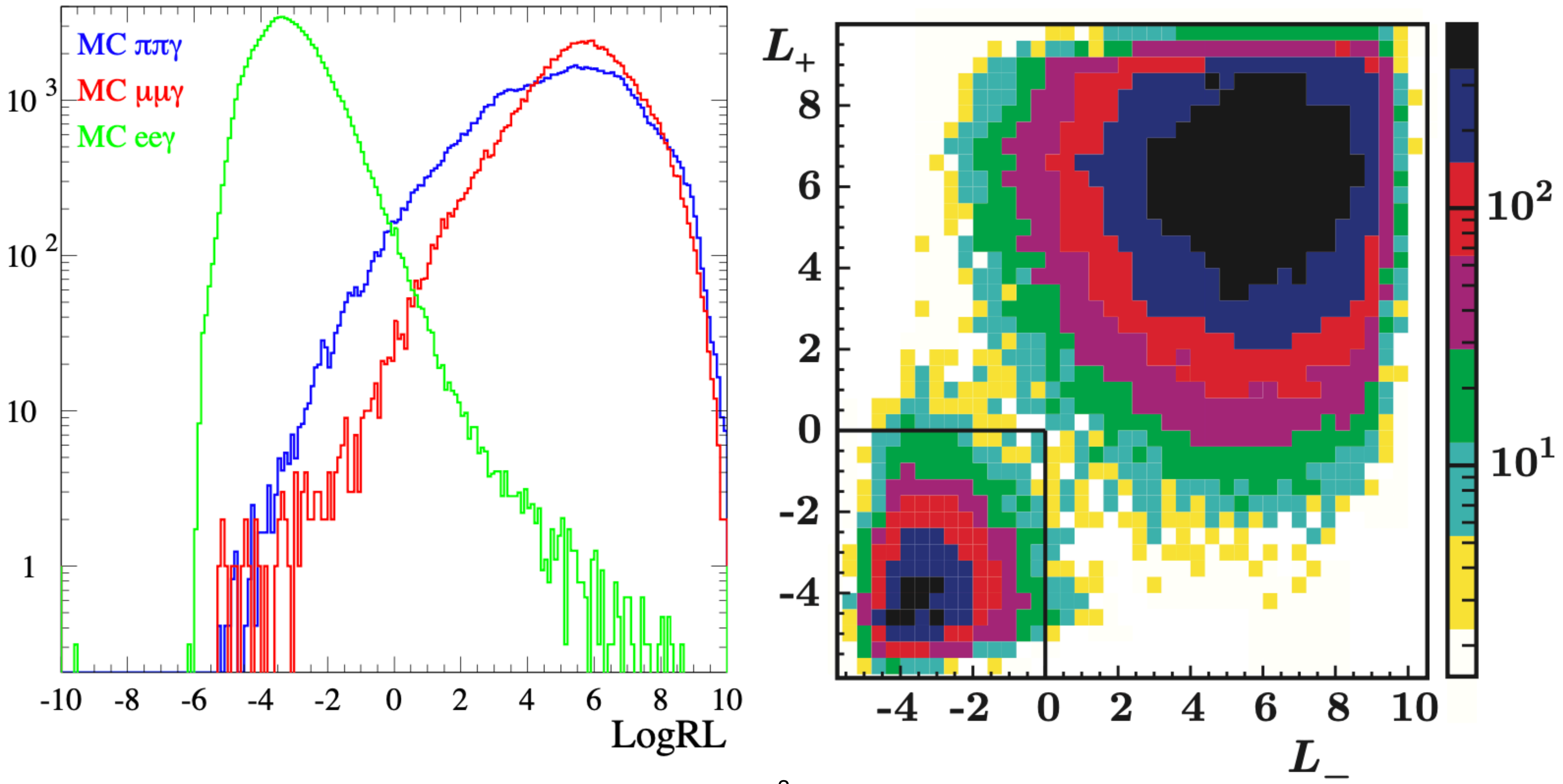
By normalising to the cross section of $e^+e^- \to \mu^+\mu^-\gamma$, many quantities and their associated error cancel. This however, comes at the cost of having to perform an additional analysis on muon events, which necessitates good separation of pions and muons

Current KLOE PID

Current PID works on individual tracks based on log-likelihood function using calorimeter information andTime of Flight:

Excellent rejection of $e^+e^-\gamma$ events with $\epsilon > 99.9\,\%$,

however muon and pion events cannot be distinguished



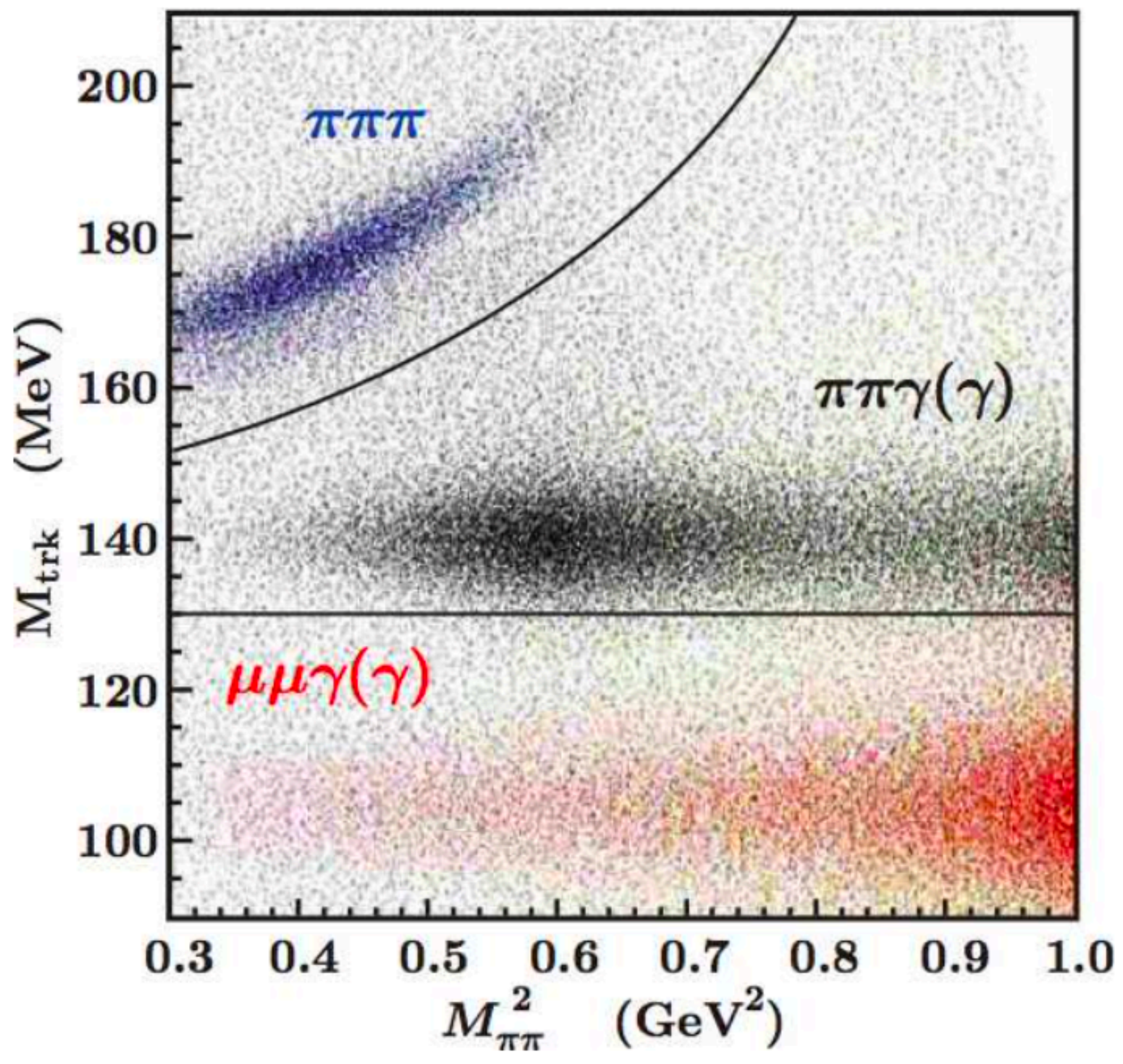
Current π/μ separation: trackmass cut

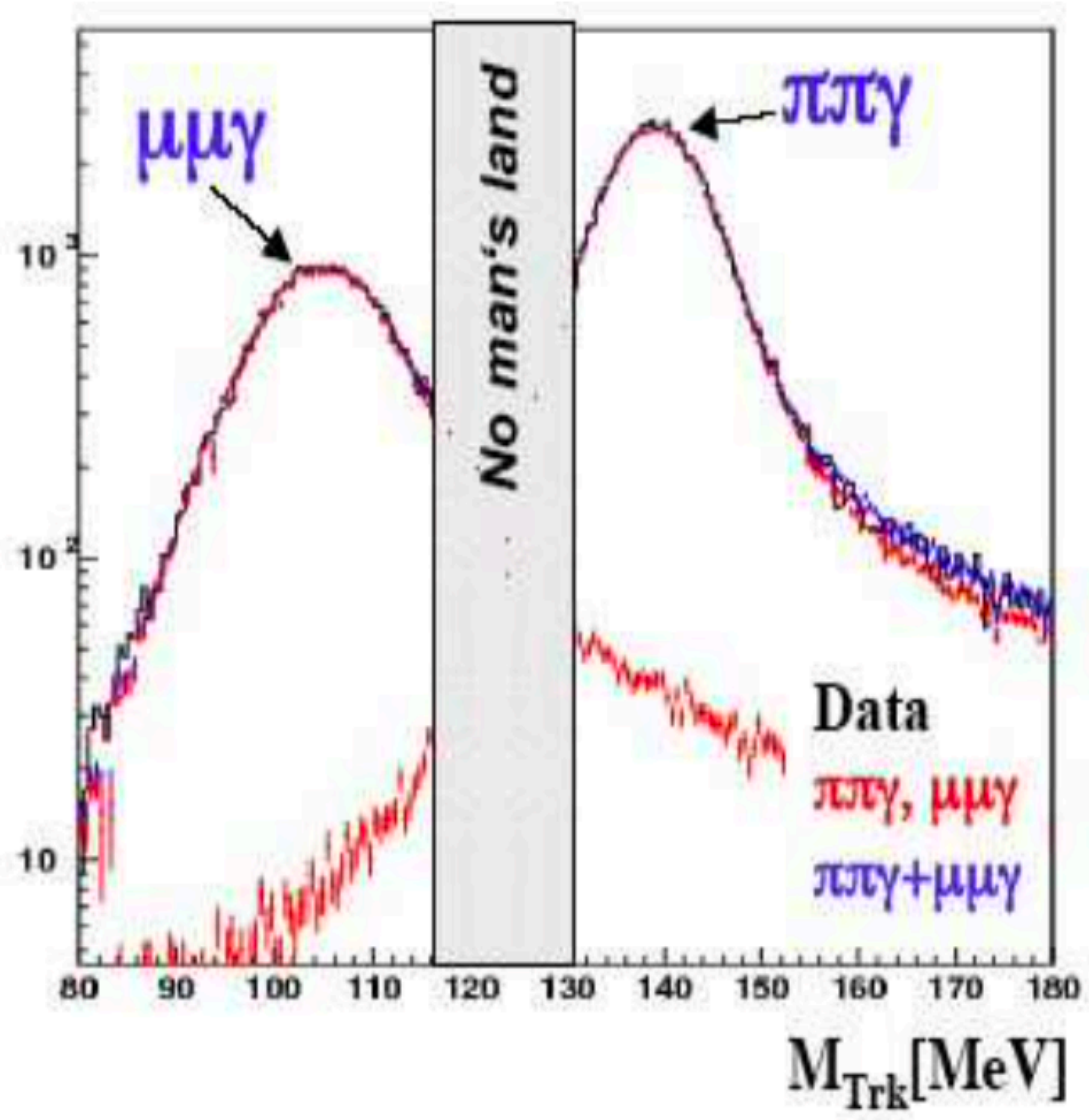
$$\left(\sqrt{s} - \sqrt{|\mathbf{p}_+|^2 + M_{\text{trk}}^2} - \sqrt{|\mathbf{p}_-|^2 + M_{\text{trk}}^2}\right)^2 - (\mathbf{p}_+ + \mathbf{p}_-)^2 = 0$$

Principle variable in KLOE, based only on information from the drift chamber

Calculated using energy and momentum conservation under the assumption of an unobserved photon, and two particles of the same mass

Current π/μ separation: trackmass cut





Current π/μ separation: trackmass cut

2002 $\pi^+\pi^-\gamma$ and $\mu^+\mu^-\gamma$ MC sample STENTU (two tracks already selected)

Based on sample of ~8,000,000 $\pi^+\pi^-\gamma$ events and ~3,500,000 $\mu^+\mu^-\gamma$ events

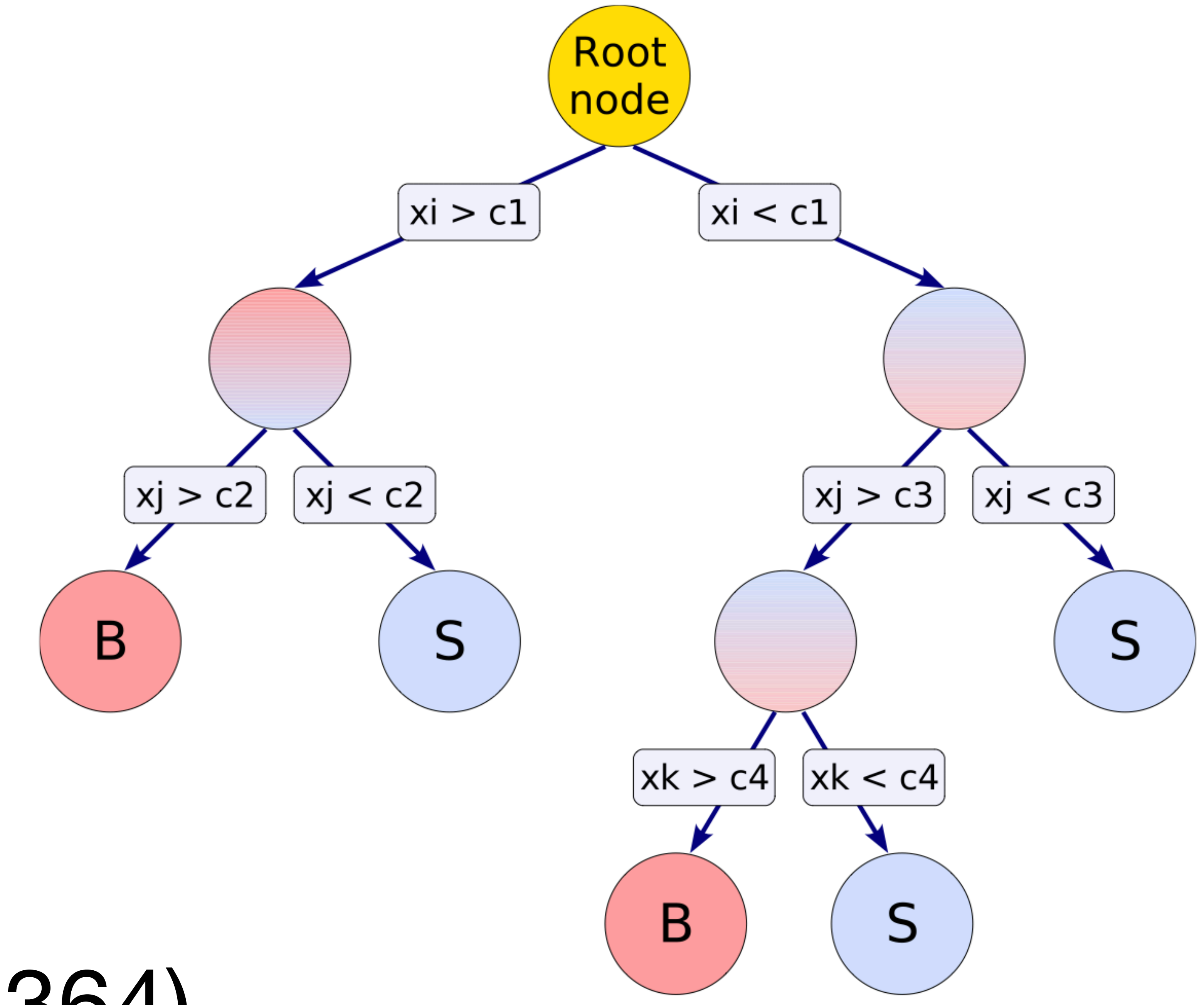
- **Pions** (selected by $m_{\text{trk}} > 130$ MeV)
  Efficiency: 96.93 %
  Purity*: 96.77 %

- **Muons** (selected by 80 MeV $< m_{\text{trk}} <$ 115 MeV)
  Efficiency: 80.49 %
  Purity*: 98.15 %

*purity only considers mutual contamination of $\pi^+\pi^-\gamma$ and $\mu^+\mu^-\gamma$

BDT

- Very popular ML method in particle physics for classification of particles
- Easy to use (available in the TMVA package in ROOT)
- Separates signal from background events based on input variables until some stop condition that then labels event as either signal or background
- “Boosting” means that wrongly classified events are given a higher weight
- MLP used in past KLOE analyses (Internal Kloe Memo 364) for tracking efficiency studies



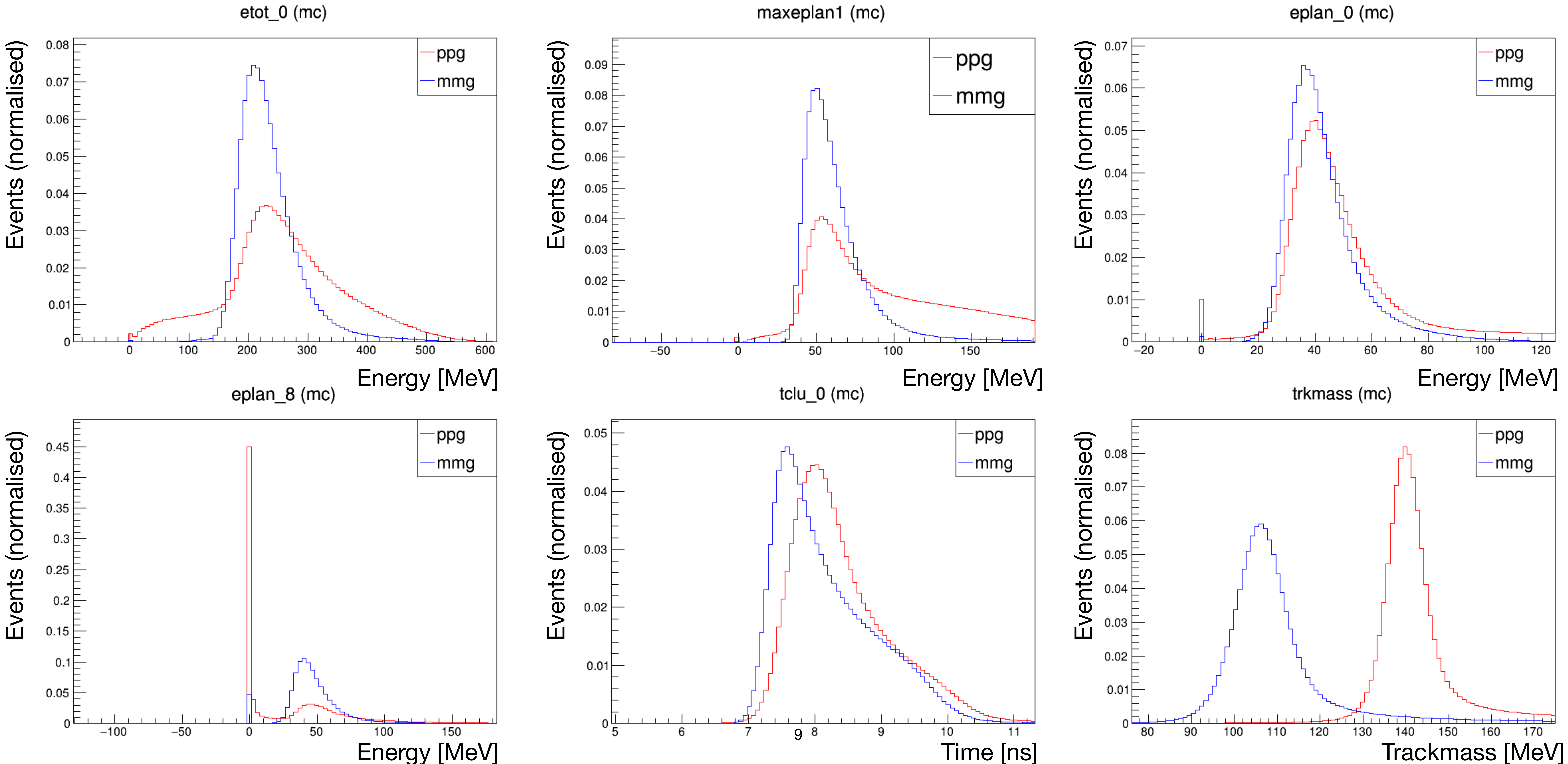
Input variables used in training

Inspiration from the π/e likelihood estimator (Internal Kloe Memo 295) as well as BESIII*

- Total energy of each cluster [etot]
- Plane of the calorimeter with highest energy deposit [eplan]
- Energy deposit in the first and last plane of the calorimeter [eplan]
- Timing of each cluster (Time of Flight preferred but not available in STENTU) [Tclu]
- Trkmass [trkmass]

*https://www.epj-conferences.org/articles/epjconf/pdf/2024/05/epjconf_chep2024_09027.pdf

Input variables used in training

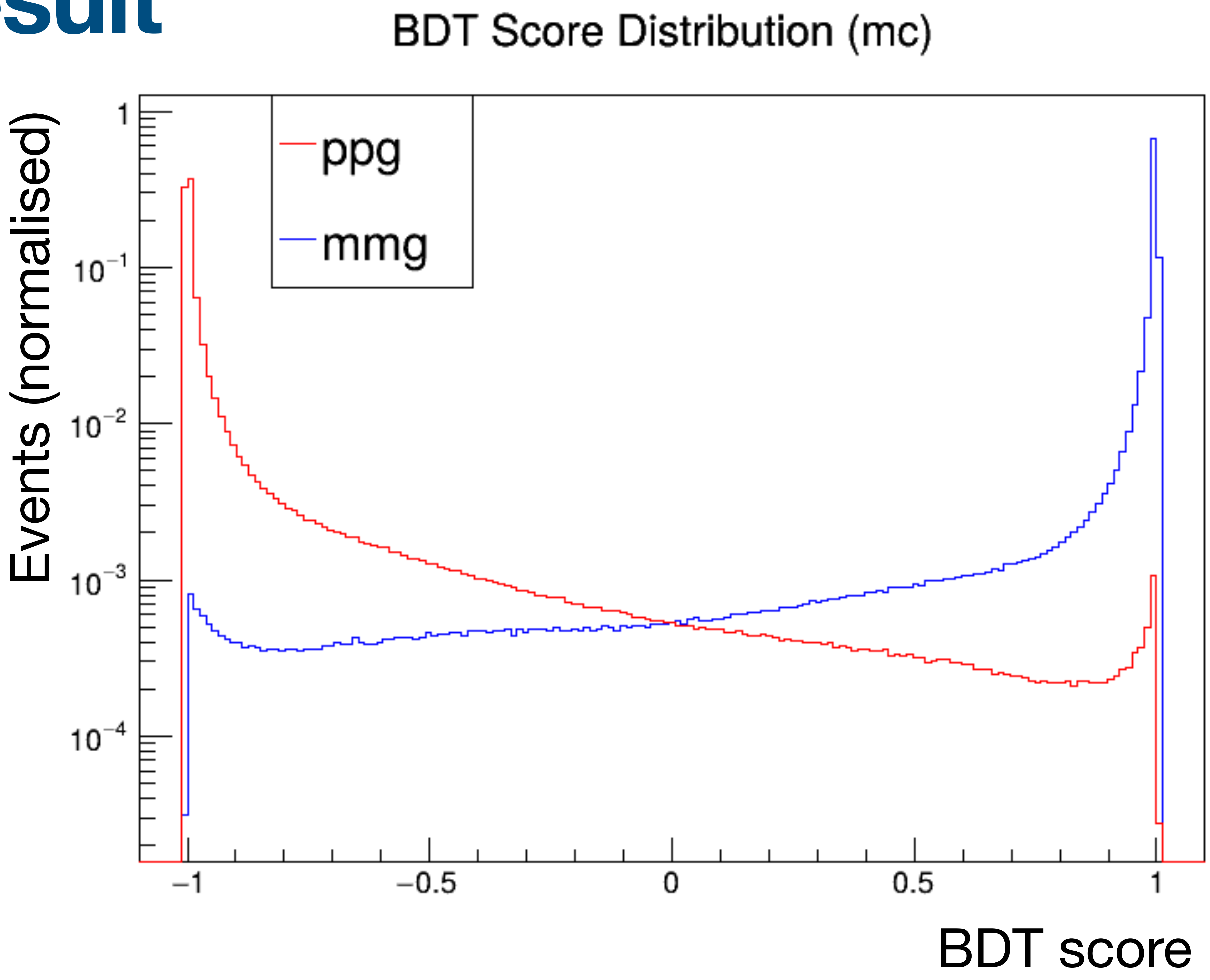


BDT training and testing

2002 $\pi^+\pi^-\gamma$ and $\mu^+\mu^-\gamma$ MC sample STENTU (two tracks already selected)

- Trained on ~10 million events each of $\pi^+\pi^-\gamma$ and $\mu^+\mu^-\gamma$
- Pre-selection criteria of KLOE12, eg angle, trackmass, first hit, pt etc already applied
- Evaluated and tested on a different subset of sample of ~8,000,000 $\pi^+\pi^-\gamma$ events and ~3,500,000 $\mu^+\mu^-\gamma$ events using the output weight file from TMVA
- $\mu^+\mu^-\gamma$ are given a score of 1 and $\pi^+\pi^-\gamma$ are given a score of -1

BDT result



BDT Score Distribution (mc)
ppg
mmg
Events (normalised)
1
10^{-1}
10^{-2}
10^{-3}
10^{-4}
−1
−0.5
0
0.5
1
BDT score

BDT result

Selection to optimise for purity*efficiency:

- $\pi^+\pi^-\gamma$ (selected with BDT < 0.366)
  Efficiency: 98.59 %
  Purity*: 97.64 %

- $\mu^+\mu^-\gamma$ (selected with BDT > 0.366)
  Efficiency: 92.88%
  Purity*: 97.77%

Very similar purity as KLOE12
but >10% increase in efficiency for muon events

Important considerations for current and future work

Currently working on:

- Getting realistic estimate of systematics
- Data/MC comparisons in order to make sure selection works on data
- Studying the amount of background events that remain ($e^+e^-\gamma$ and $\pi^+\pi^-\pi^0$)

Future work (once data becomes available):

- Find a more optimal set of input variables
- Re-train using ROOT-samples with lower level variables once files are available
- Modify to work on track-level rather than event-level, similar to what was done in past KLOE analyses
- Best stage of analysis to implement the BDT (selection, background subtraction, tracking efficiency)