Summary of
https://agenda.infn.it/event/18113/contributions/85363/attachments/63303/76199/ReactionID.pdf

Variable Classes

Observables
- Invariant masses
- Production/Decay angles
- CoM kinematics

Discriminatory
- Exclusive Process
  - Missing Mass
  - Missing Energy
  - Missing Momentum
- Decay Processes
  + Invariant masses

Particle Identification
- Momentum, Position
- Time of Flight
- Delta Energy
- Cherenkov ...

Used to extract physics

Don't want to mess with these

Used to remove backgrounds

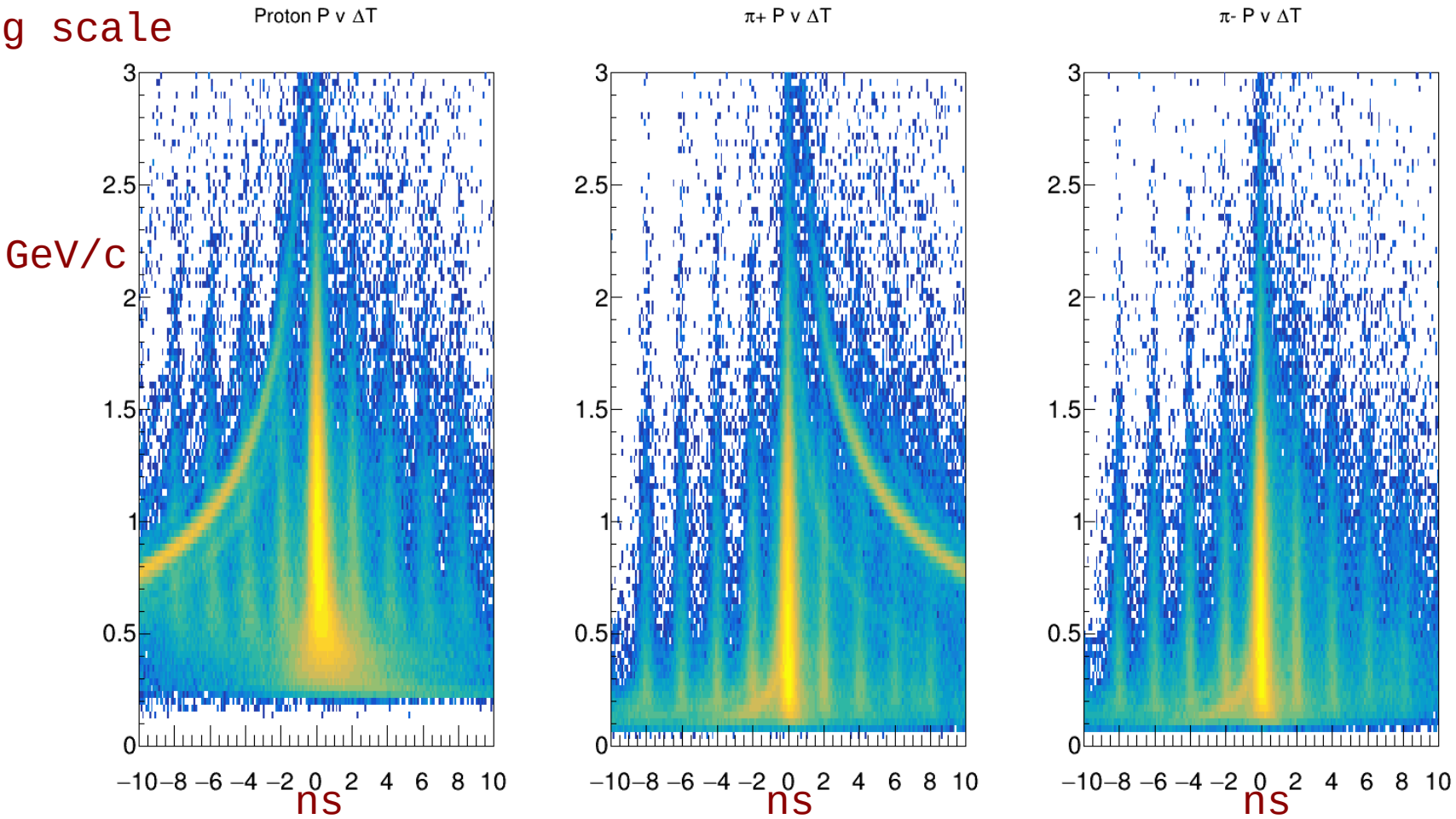
They still need to be useable after any ML (or other) cuts

Used to minimise backgrounds

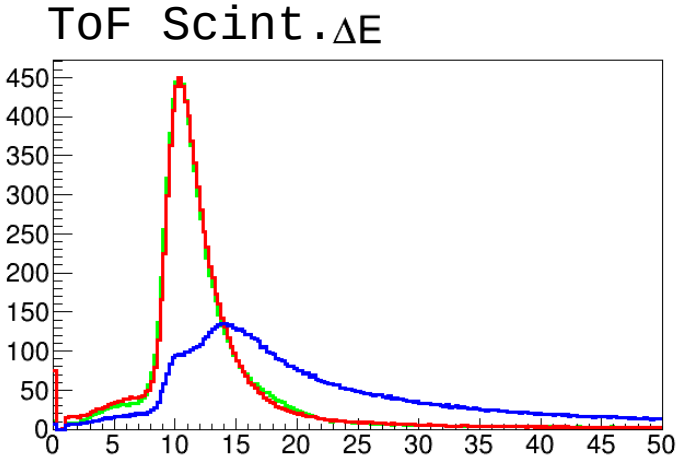
Best candidate for ML algorithms??

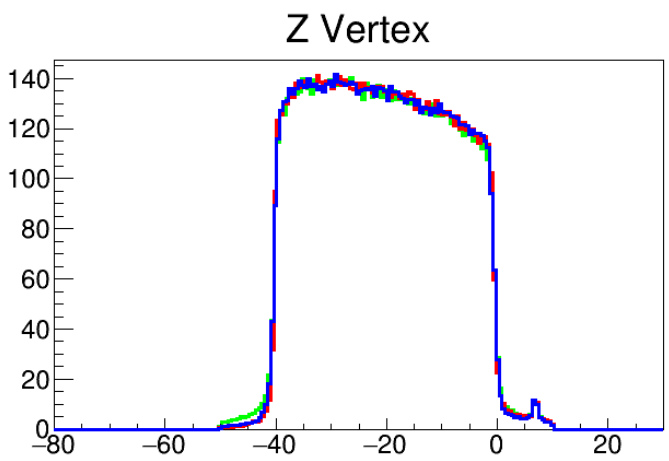
Example from CLAS ΔTime PID

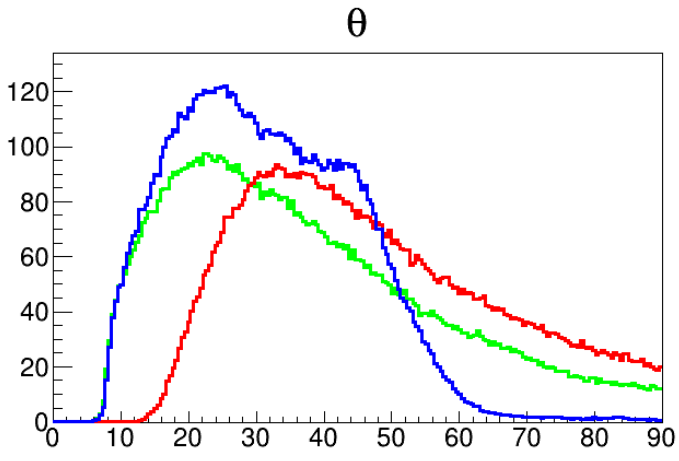
*Log scale

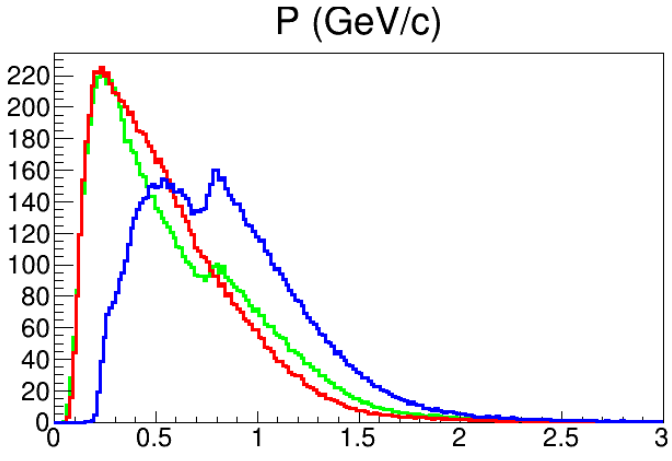


Other Input Variables







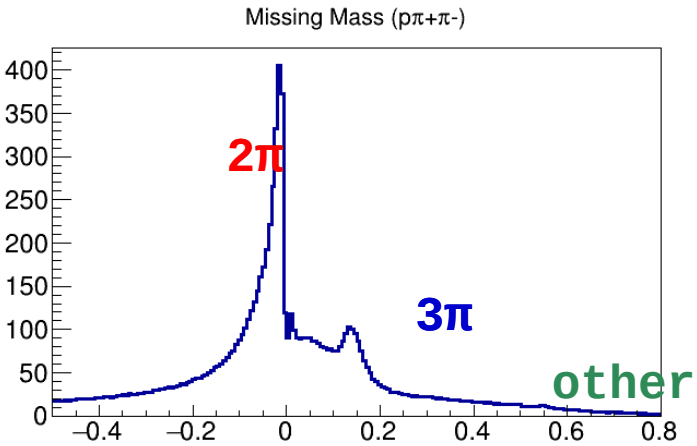


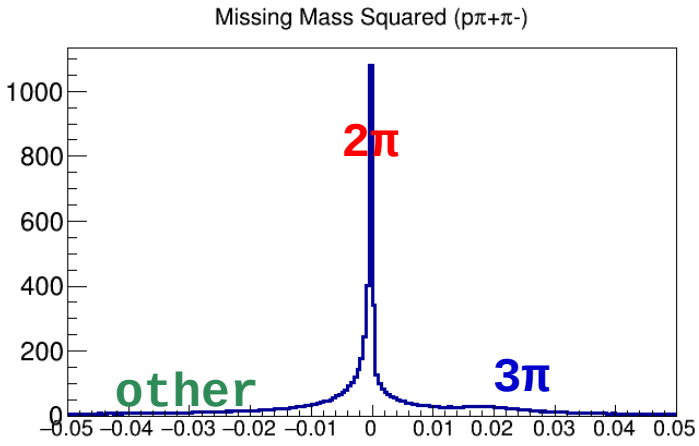
Proton

π^+

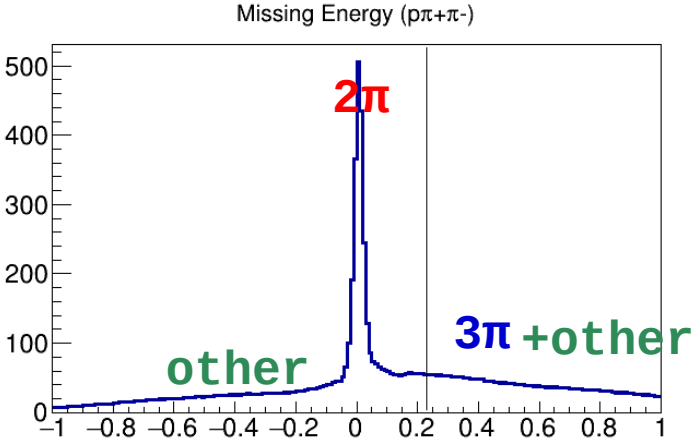
π^-

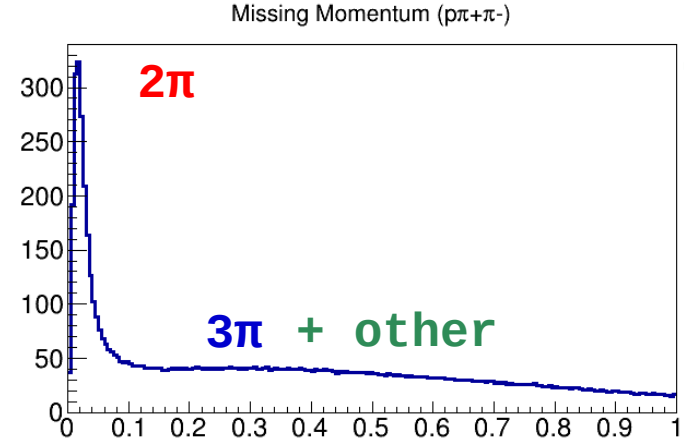
Exclusivity variables





NOT FOR TRAINING WITH





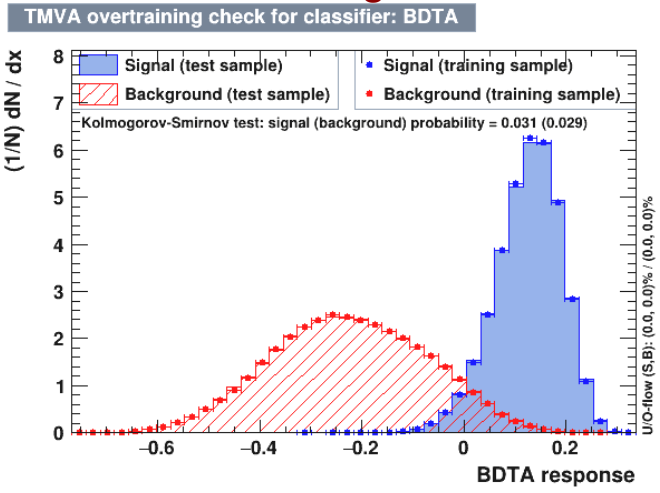
GOAL :
Using PID variables
Retain all 3π
Remove backgrounds

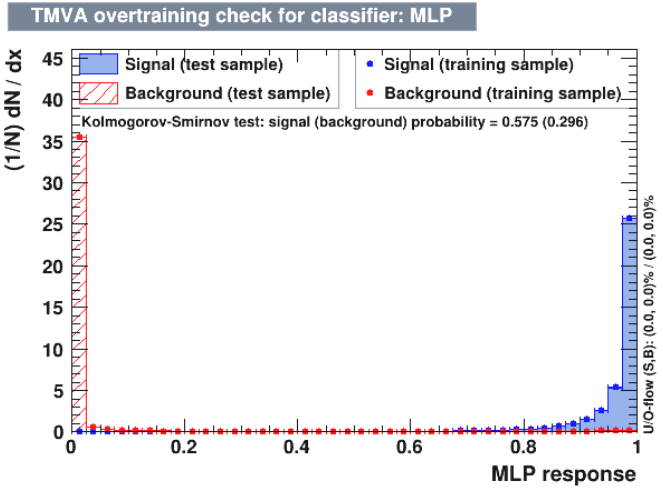
Train with Simulation

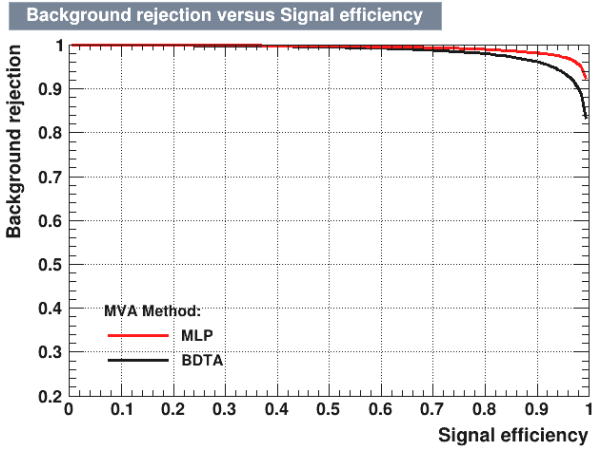
Signal : **Simulated** 3π (Truth Matched)
Background : All Experimental Data Events +(Simulated Non Truth)

Training variables + E_γ
Particle(ΔTime, ΔEnergy, P, Vz, θ, φ)

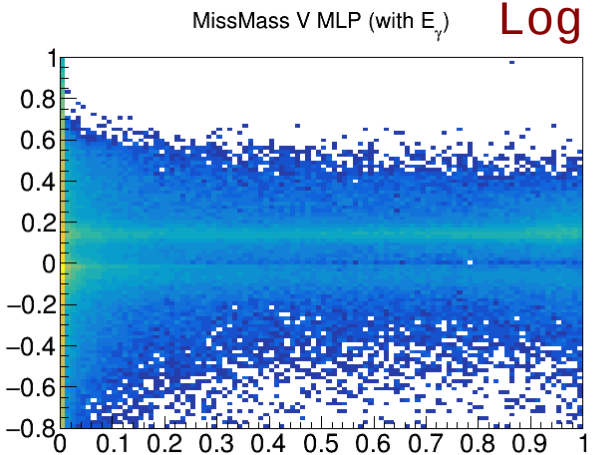
200k Training and Test data



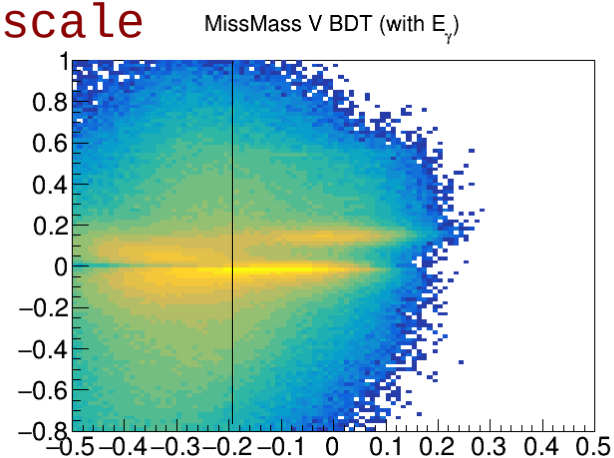




Apply Simulated training



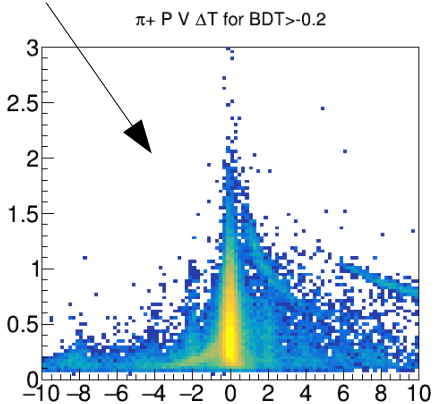
Log z scale

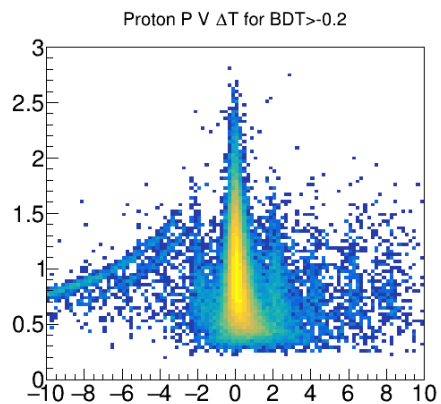


Real Experiment distributions shifted to background values compared to Simulation

Simulation must be very Realistic!

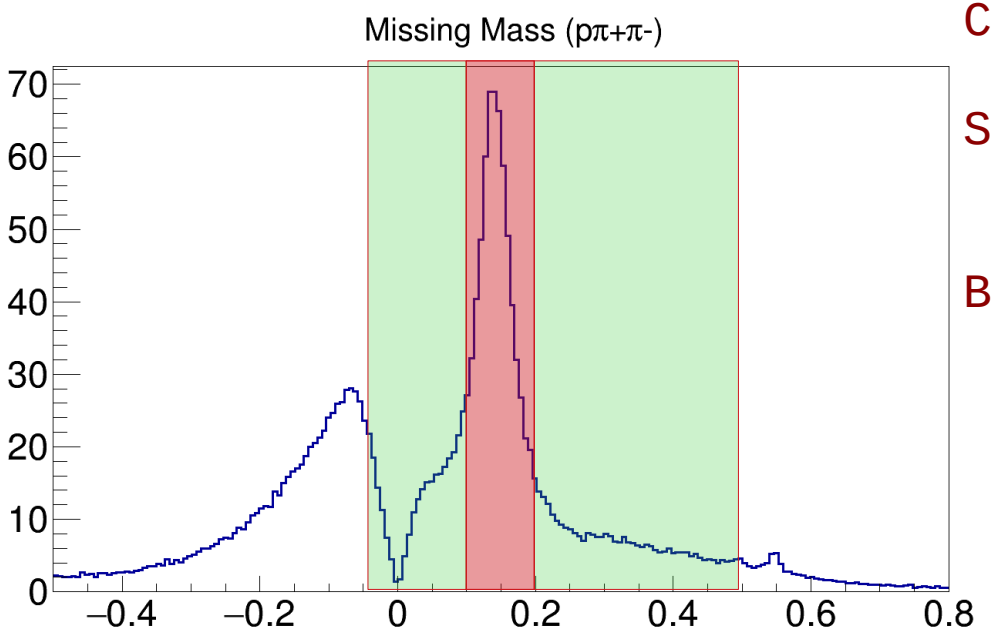
Small seperation between 2π and 3π
- OK PID variables similar





Cuts still clean up data, but must use low BDT value

Mixed events training



Cut on region with high signal density

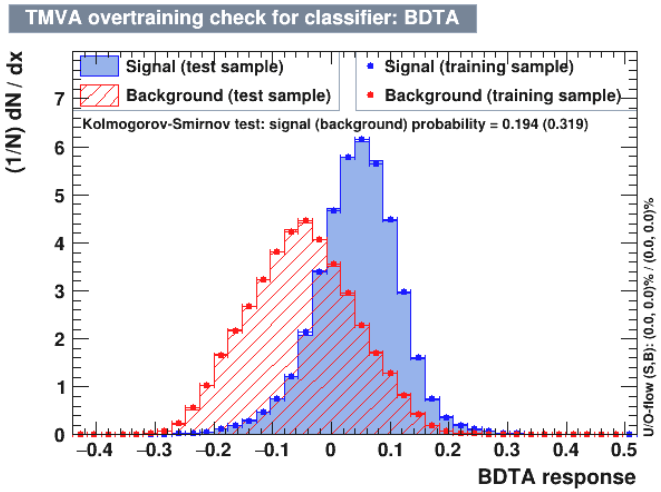
Signal = (Missing mass peak ~ π^0)

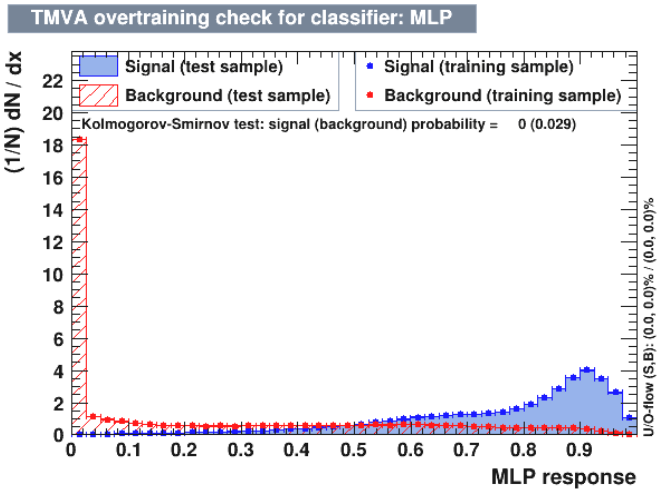
Background = sidebands inc 2π

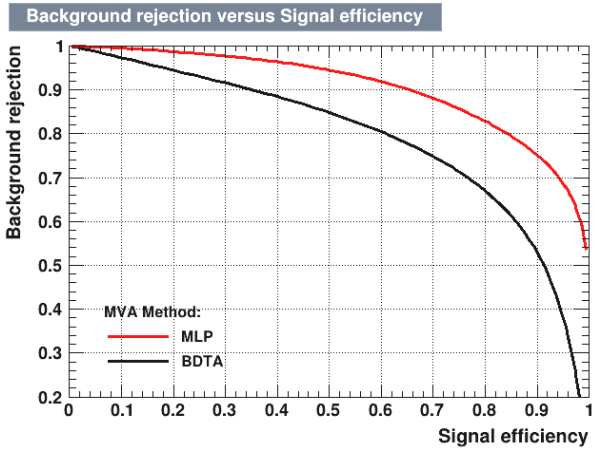
Train with Mixed Events (III)

Signal : 0.1<MissMass<0.2
Background : -0.04<MissMass<0.1 and 0.2<MissMass<0.5

Variables + E_y



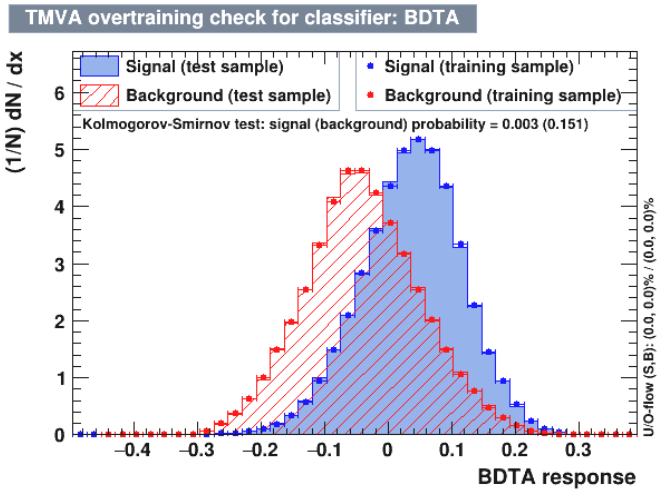


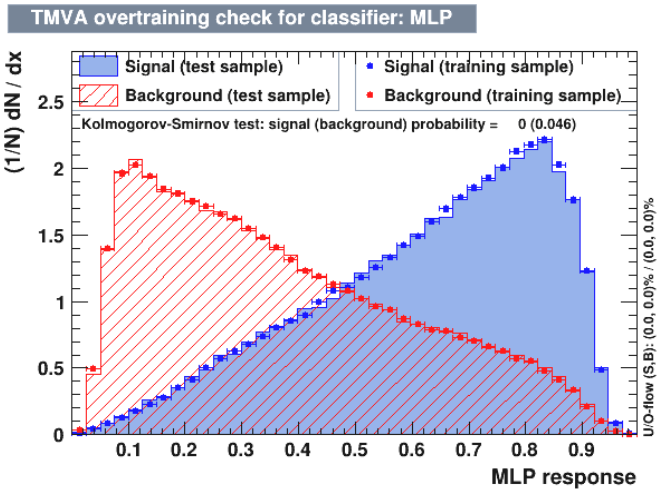


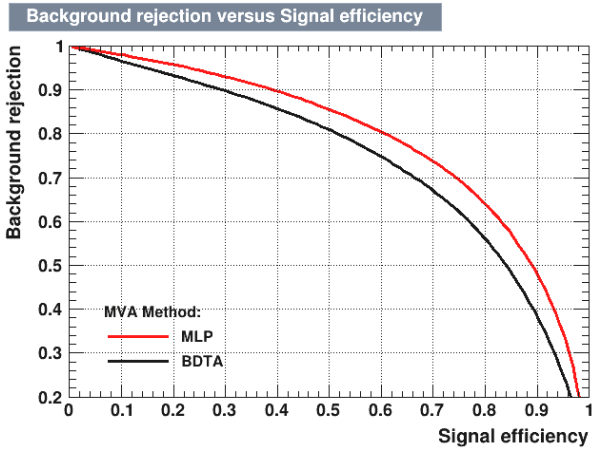
Train with Mixed Events (IV)

Signal : 0.1<MissMass<0.2
Background : -0.04<MissMass<0.1 and 0.2<MissMass<0.5

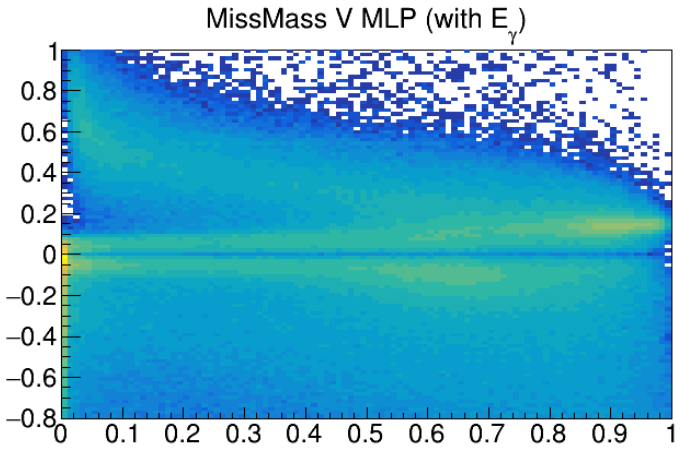
Variables - E_y

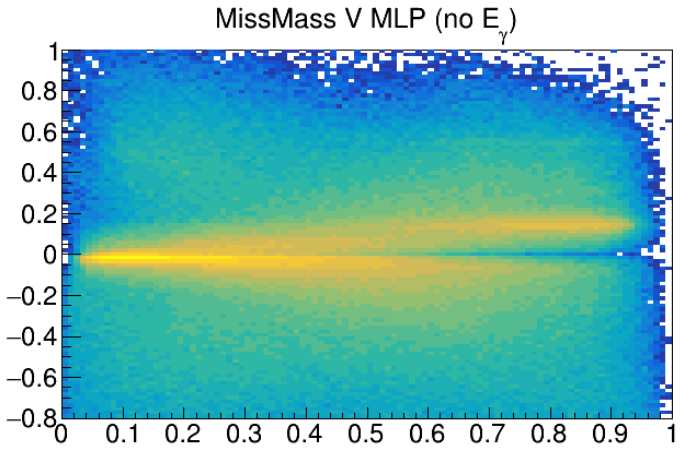


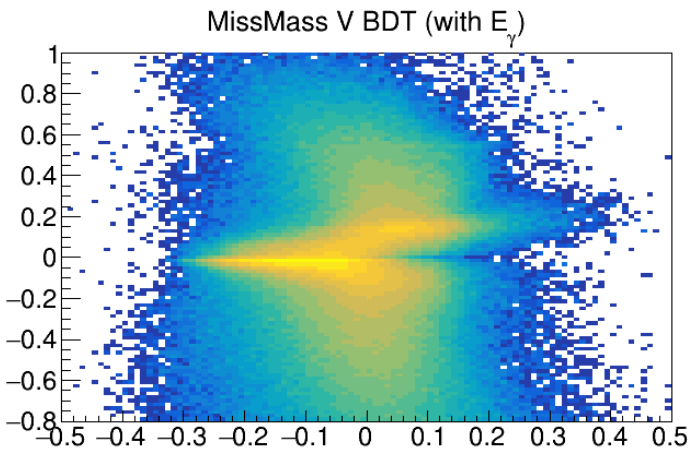


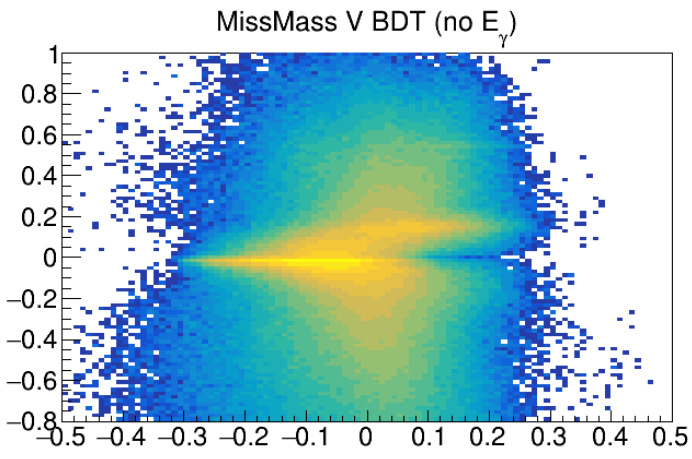


Miss Mass V Responses

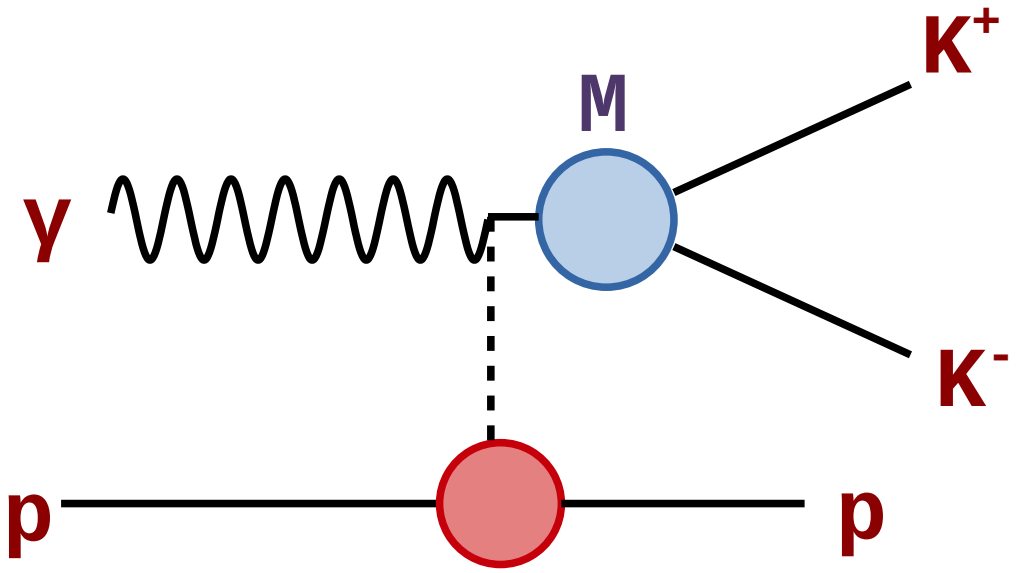








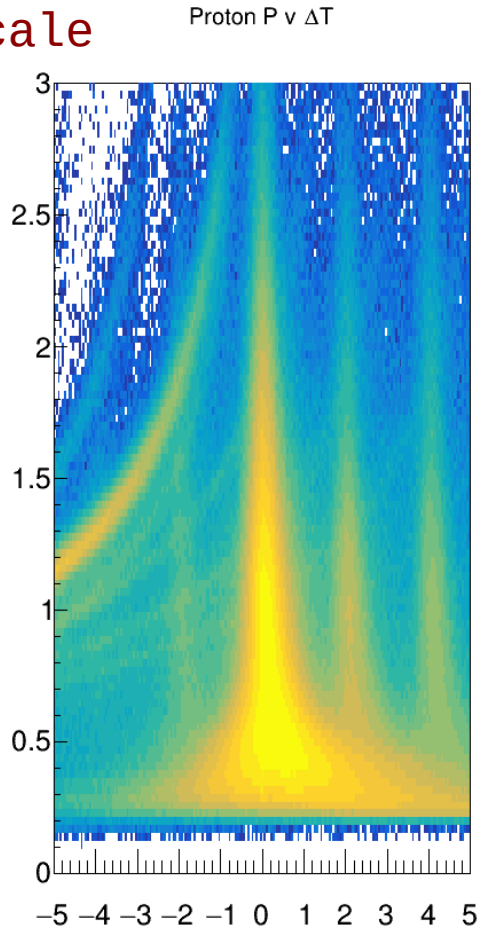
Example Reaction

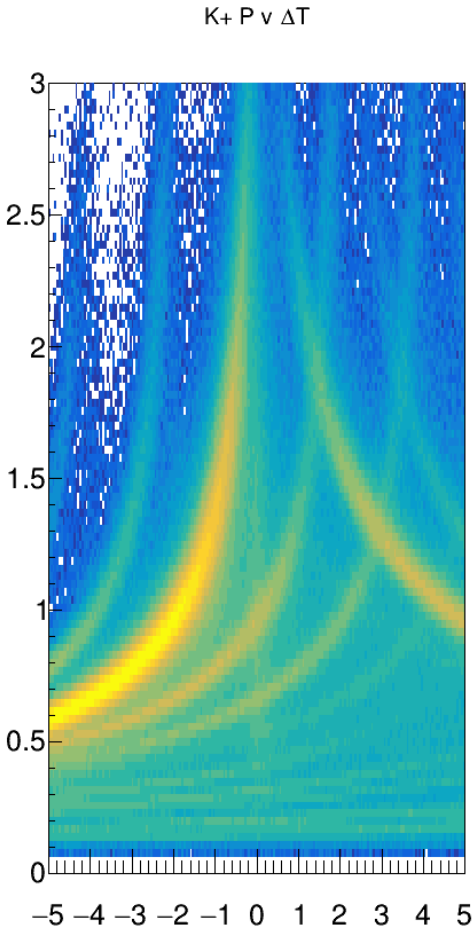


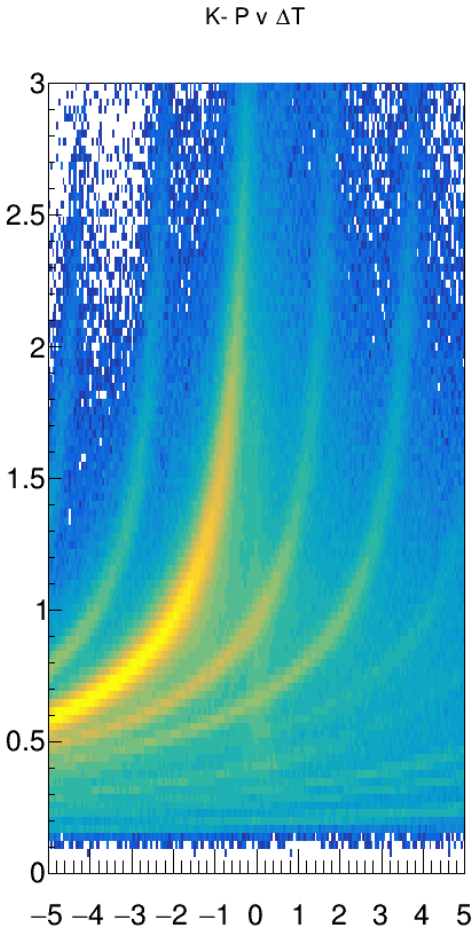
Large background from
Pions IDed as Kaons

ΔTime PID

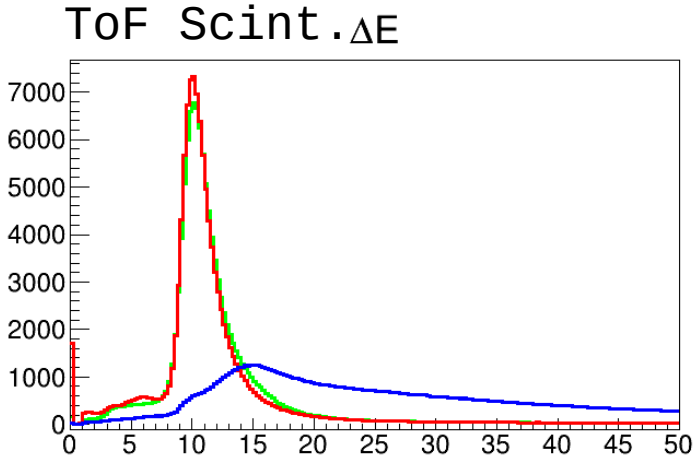
*Log scale

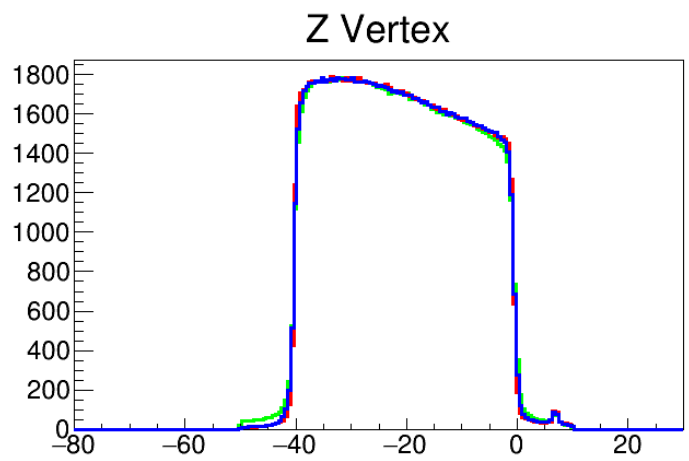






Other Input Variables

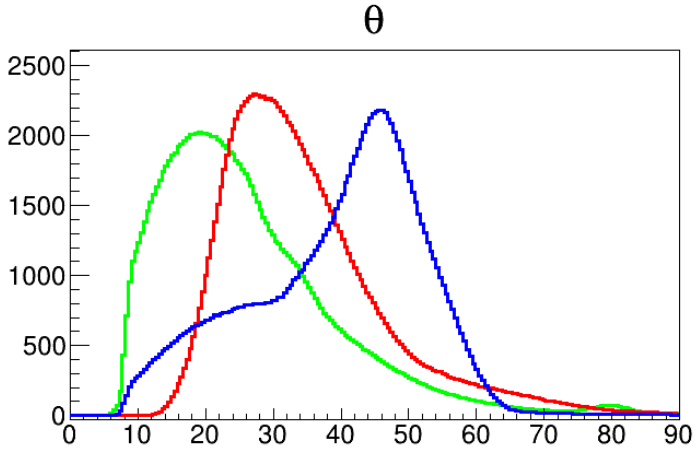


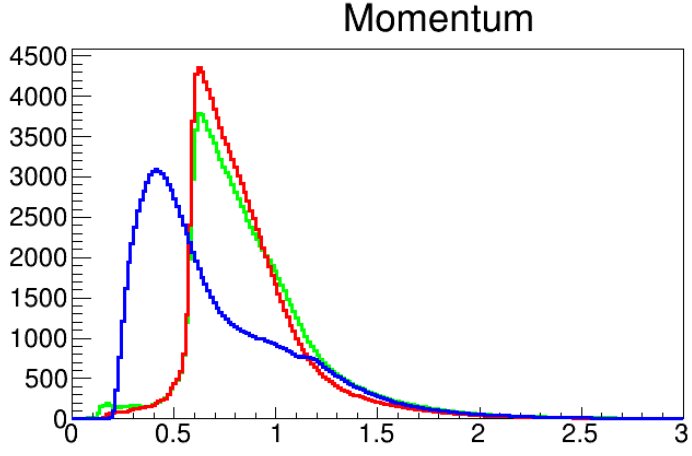


Proton

K^+

K^-





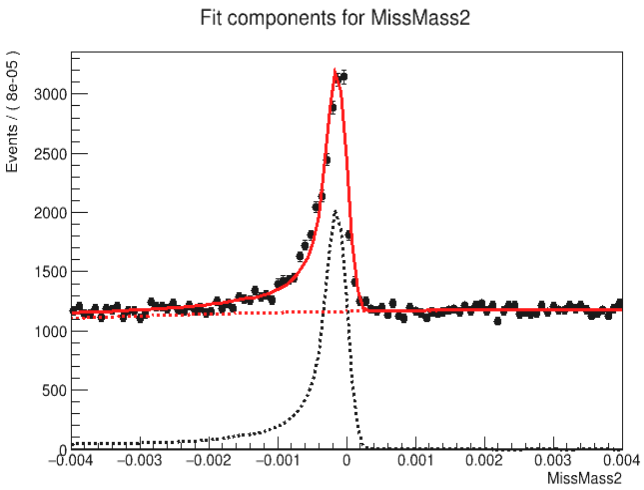
Train BDT with sWeights

SPlot – technique for disentangling different event species using a discriminatory variable (generalised side-band subtraction)

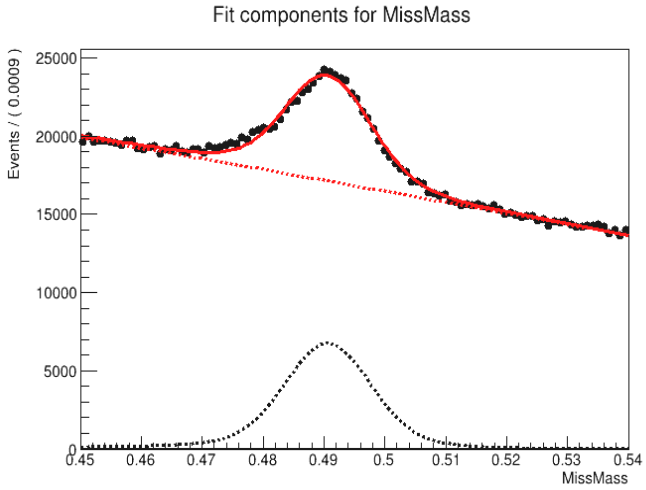
M. Pivk,F.R. Le Diberder,Nucl.Inst.Meth.A 555, 356-369, 2005

Used RooStats implementation

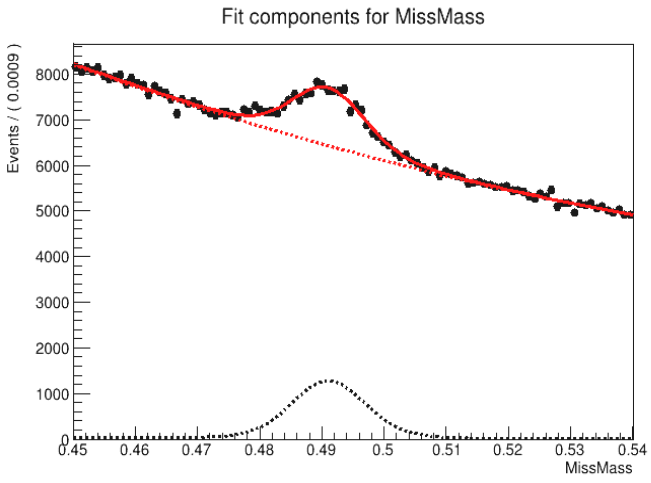
Exclusive



Missing K-



Missing K+



In TMVA BDT accept negative event weights

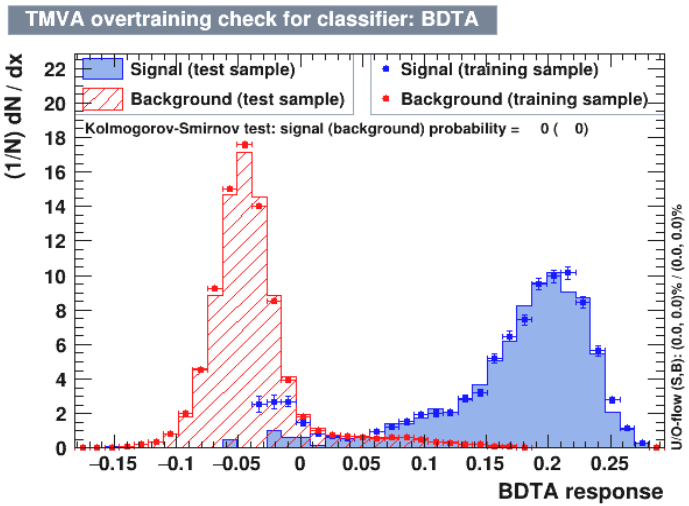
MLP does not

$$ {}_s\mathcal{P}_{\mathrm{n}}(y_e) = \frac{\sum_{j=1}^{\mathrm{Ns}} \mathbf{V}_{\mathrm{n}j} \mathrm{f}_j(y_e)}{\sum_{k=1}^{\mathrm{Ns}} N_k \mathrm{f}_k(y_e)} $$

Train BDT with sWeights

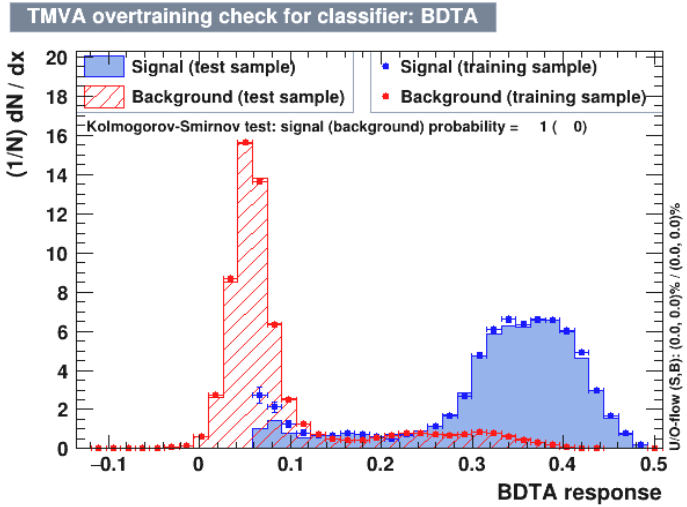
80k test and 20k train

Exclusive



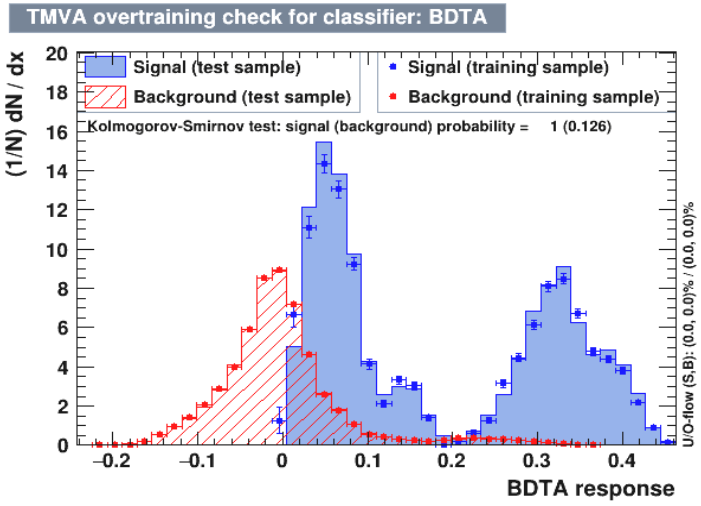
300k test and train

Missing K-



300k test and train

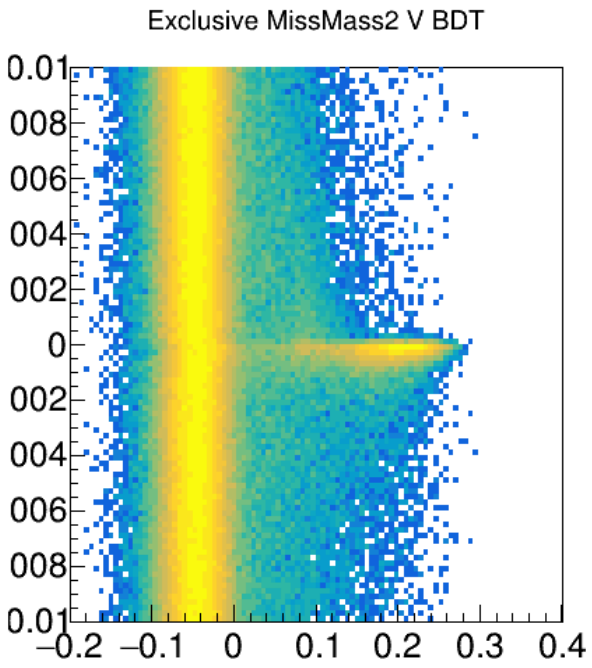
Missing K+

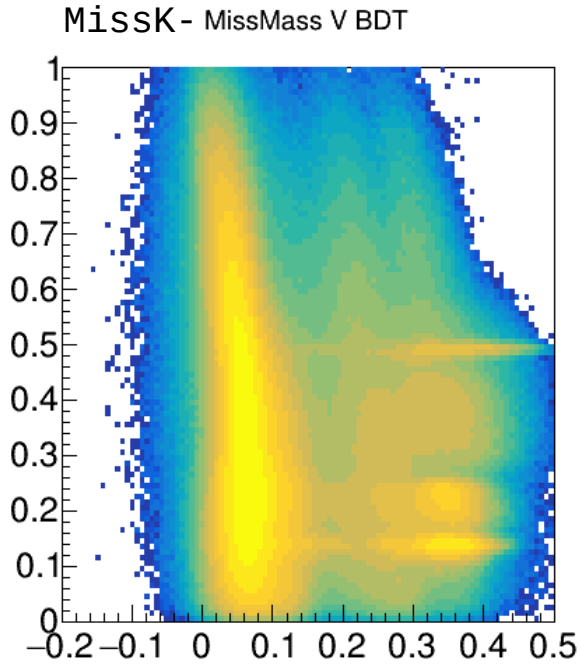


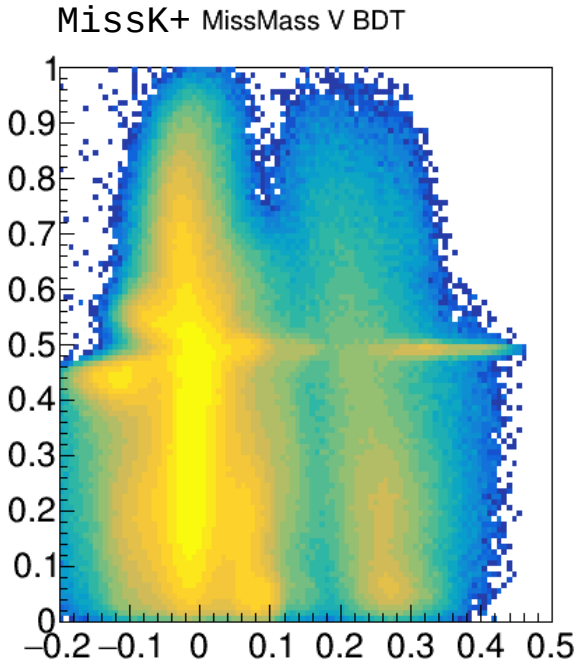
2 types of events in Missing K+ signal weights

Seperated by training

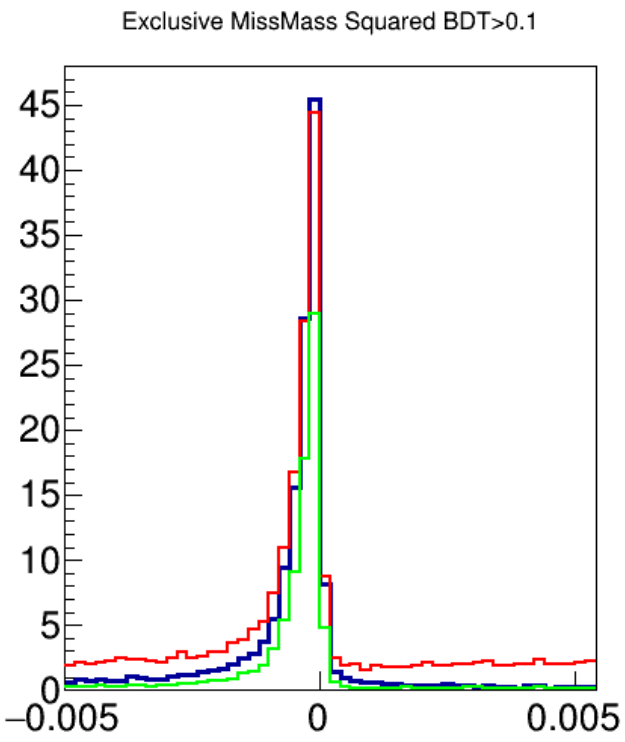
Exclusivity variables V BDT

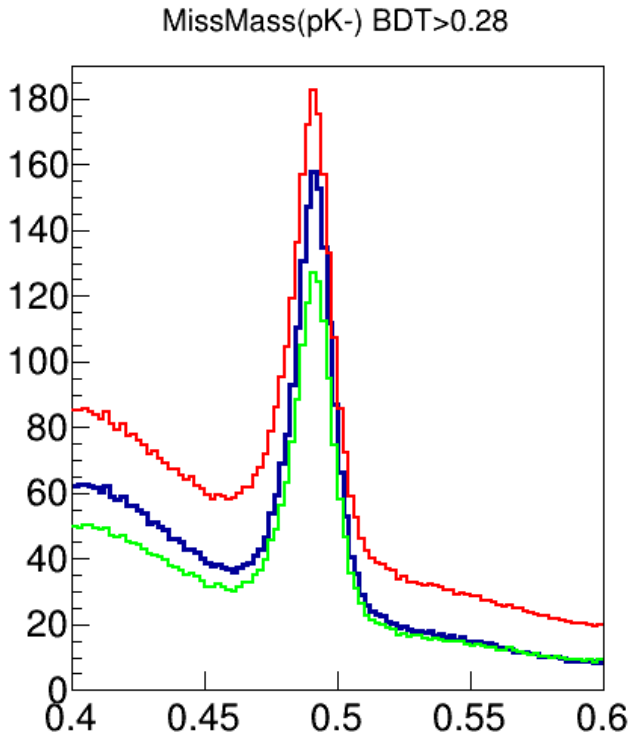


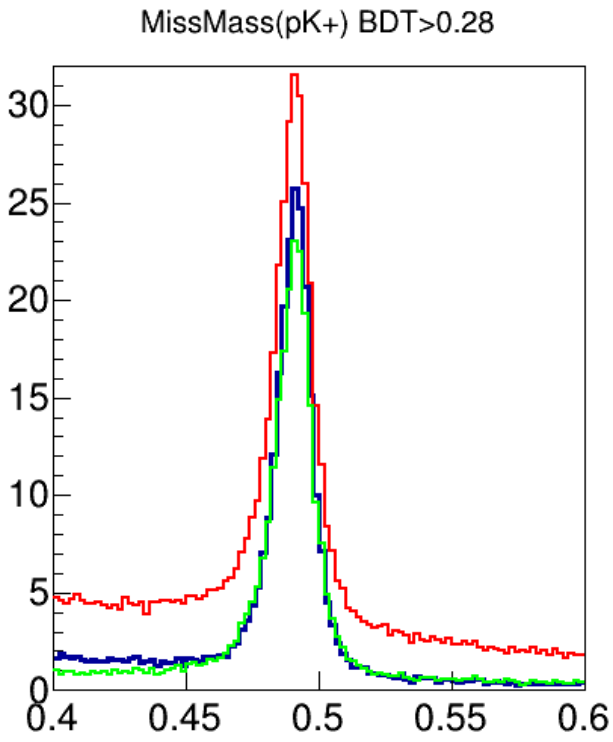




Event exclusivity with BDT cut



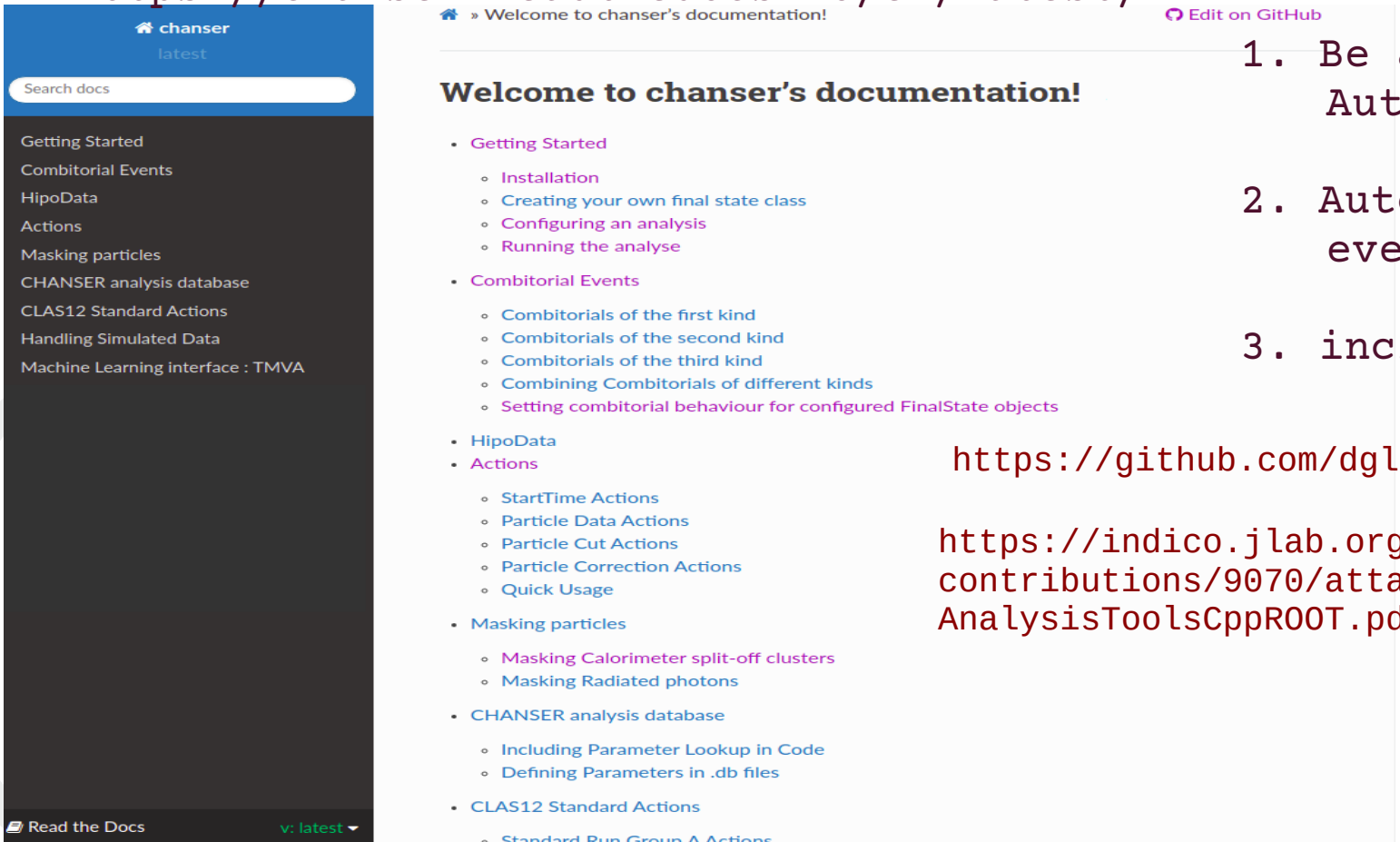




Loose Cuts Based(1ns)
Tight Cuts Based(0.5ns)
BDT Cut

CHANSER
CLAS12 HIPO ANALYSER

https://chanser.readthedocs.io/en/latest/



1. Be able to Automate everything
2. Automate everything
3. including ML

https://github.com/dglazier/chanser/wiki/TMVA

https://indico.jlab.org/event/472/contributions/9070/attachments/7345/10144/AnalysisToolsCppROOT.pdf