Improving Reaction Identification With ROOT TMVA

Derek Glazier
University of Glasgow

Overview

Exclusive reactions

JLAB + CLAS + CLAS12

Discriminatory variables

Training variables

TMVA

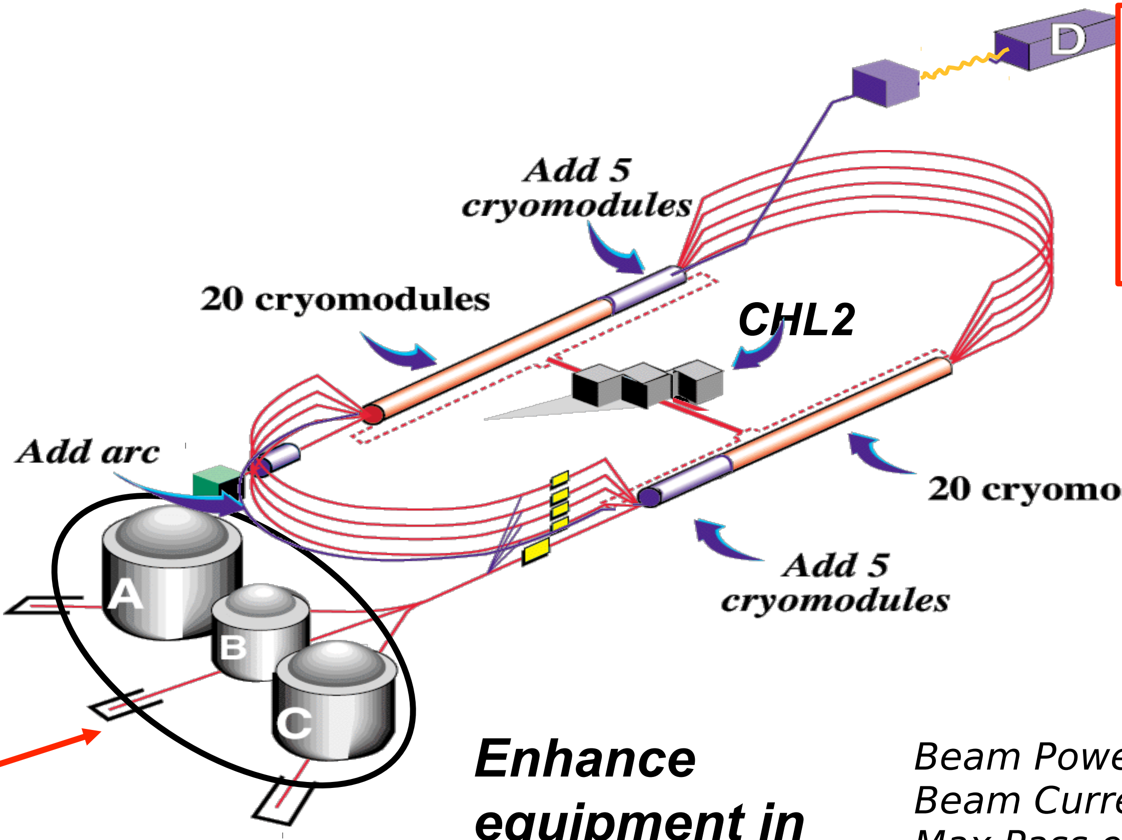
Simulation training

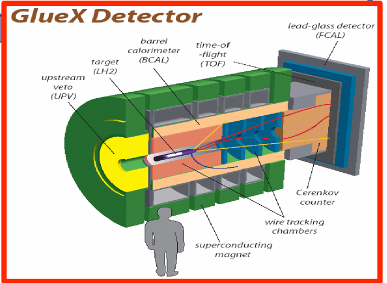
Mixed events training

sPlot training

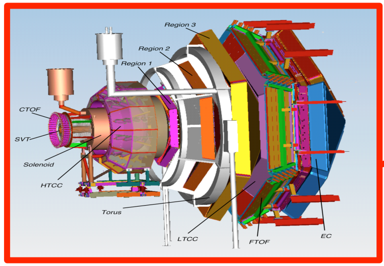
Accelerate BDT->MLP

JLAB at 12GeV





GLUEX



CLAS12

Enhance equipment in existing halls

Beam Power: 1MW
Beam Current: 90 μA
Max Pass energy: 2.2 GeV
Max Enery Hall A-C: 10.9 GeV
Max Energy Hall D: 12 GeV

Variable classes

Observables

Invariant masses
Production/Decay angles
CoM kinematics

Used to extract physics

Discriminatory

Exclusive Process
- Missing Mass
- Missing Energy
- Missing Momentum

Decay Processes
+ Invariant masses

Used to remove backgrounds

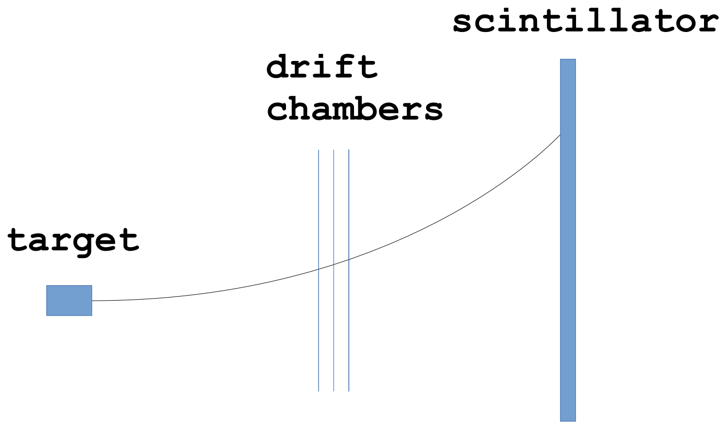
Particle Identification

Momentum, Position
Time of Flight
Delta Energy
Cherenkov ...

Used to minimise backgrounds

Focus of this talk

“Standard” CLAS track PID



Measure momentum, charge C, vertex V, from tracking in Toroidal field

Measure Time of Flight T, to scintillator from reaction vertex

“Swim” track in field to reconstuct path length D.

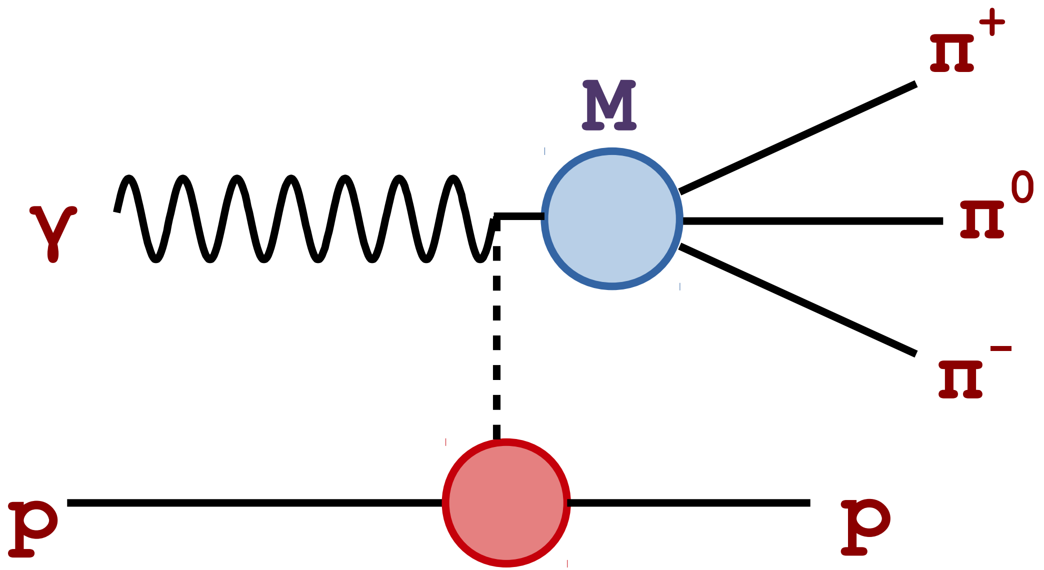
Determine Time of flight Ti,from momentum and path for species hypothesis i e.g. i = proton, K+ , π+,...

$$T_i = \frac{D}{\beta_i \cdot c} \, with \, \beta_i = \frac{P}{\sqrt{P^2 + M_i^2}}$$

then $\Delta T = T - T_i$ ~0 for correct hypothesis

Also measure mass, Δβ, … but just the same information

Example Reaction

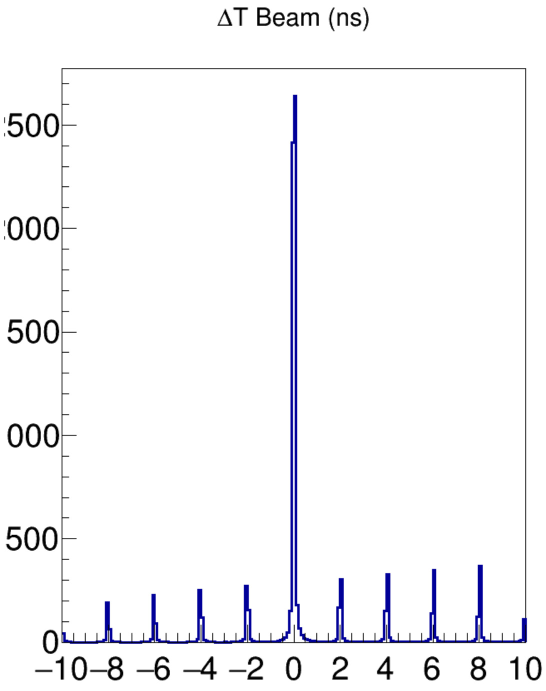


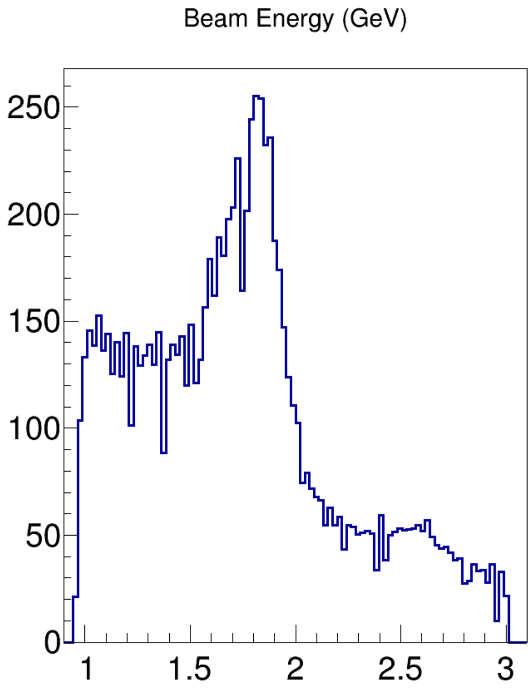
Focus on CLAS charge particle detection
– miss π^0

Filter events with 1 -ve , 2 +ve and N tagged photons

assume -ve = π^-
create 2 events for each positive as proton and π^+

Tagged Photon





CEBAF e- beam ~2ns bunches

Photon beam produced from e- bremsstrahlung

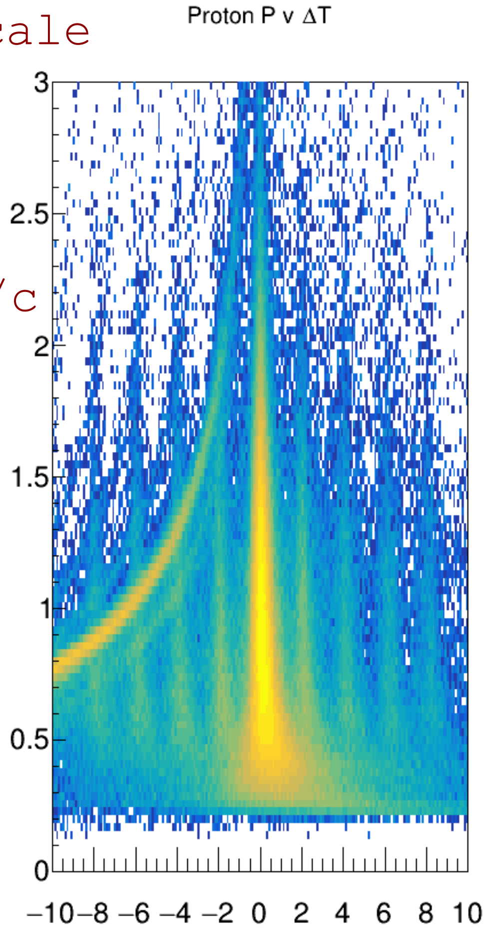
Use of diamond radiator gives Coherent peak (polarised)

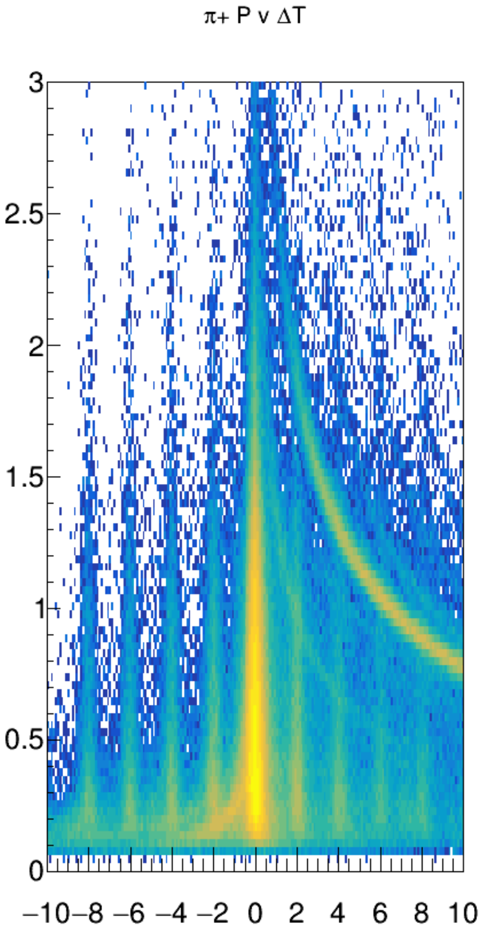
Recoil electron detected in Magnetic spectrometer "tagger" => the photon energy

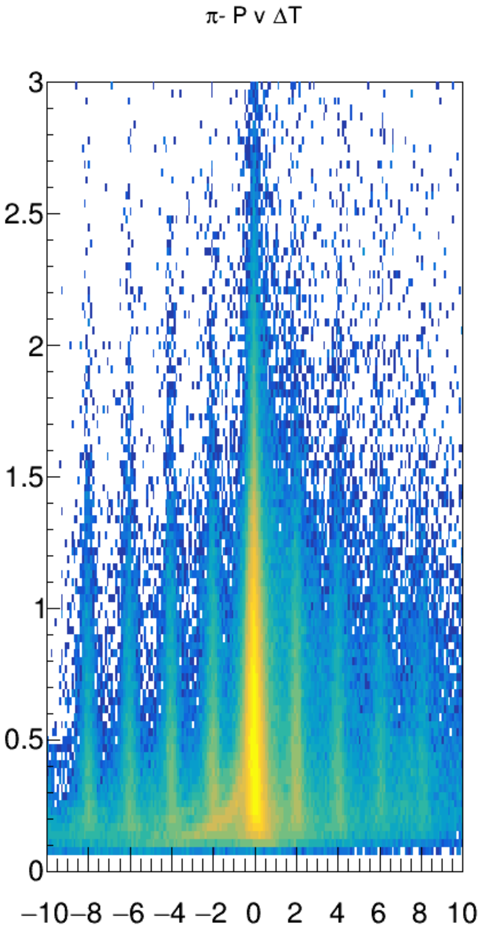
ΔT Beam gives coincidence of tagger with CLAS and reaction vertex time

ΔTime PID

*Log scale



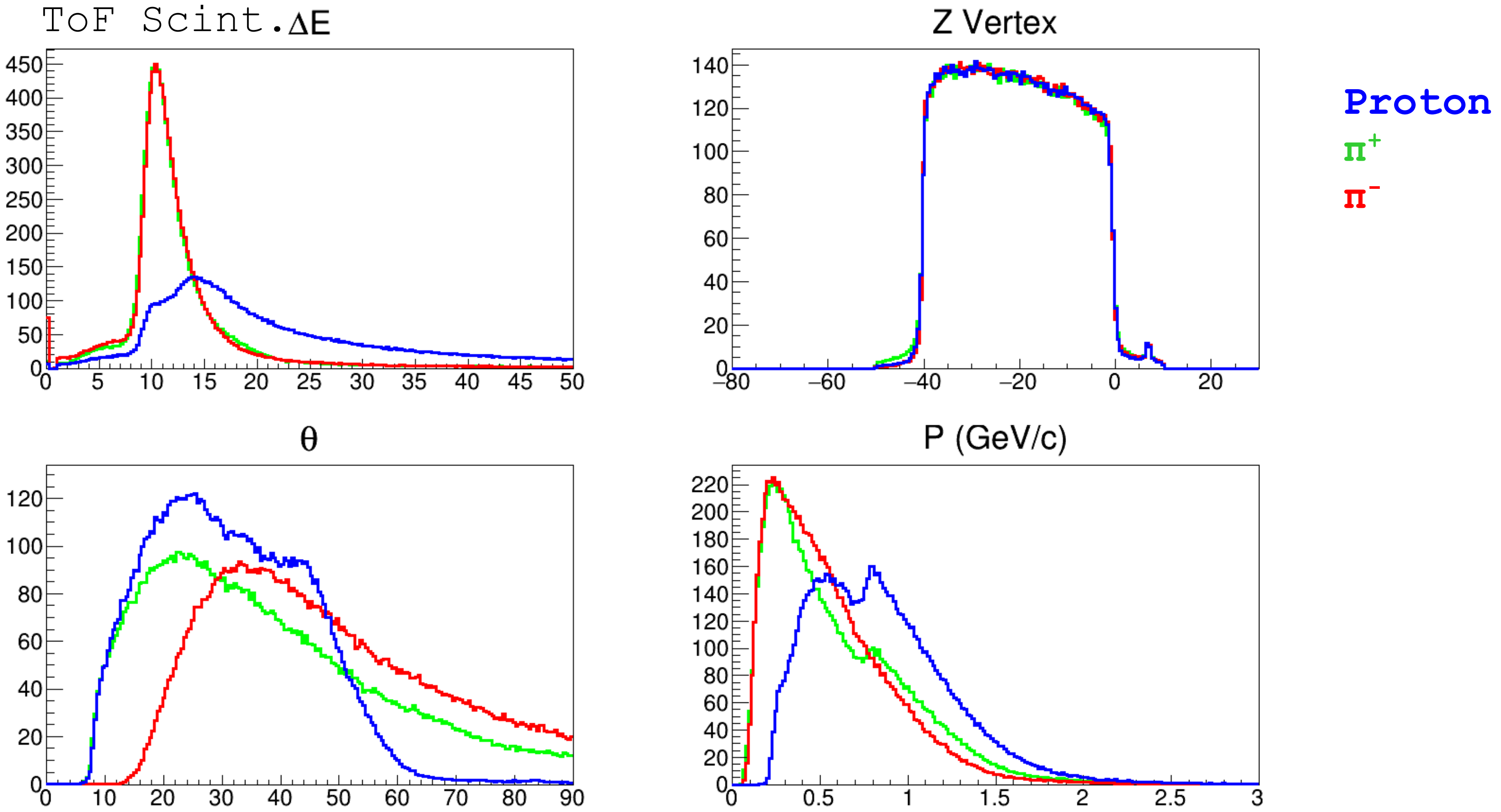




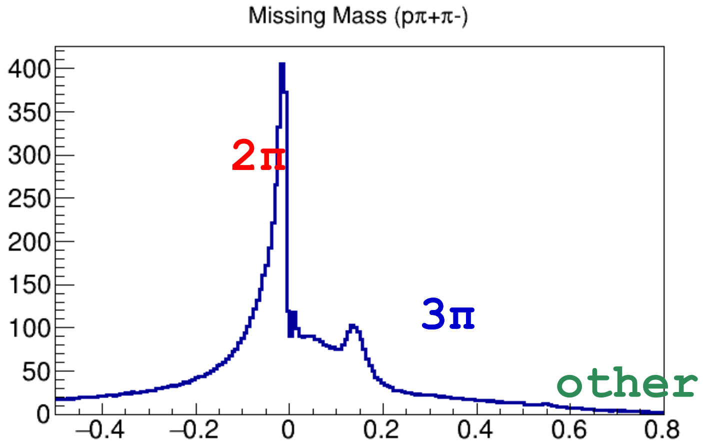
GeV/c

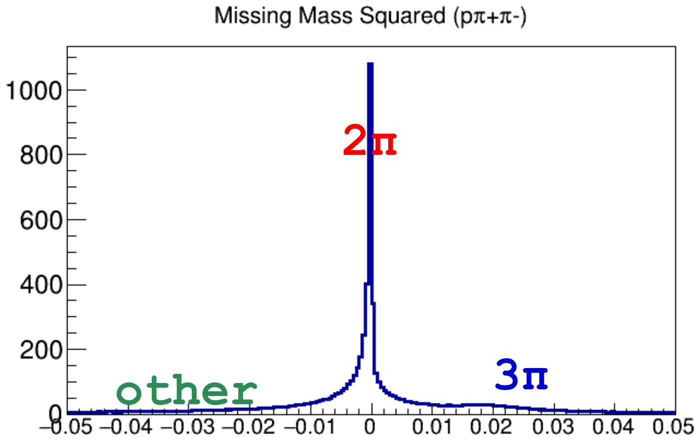
ns ns ns

Other Input Variables

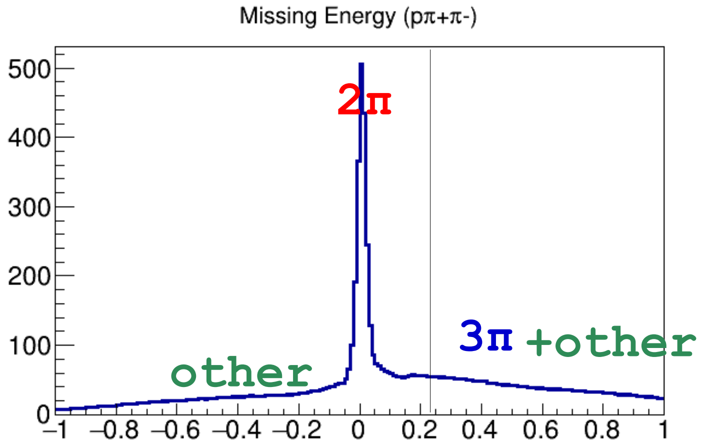


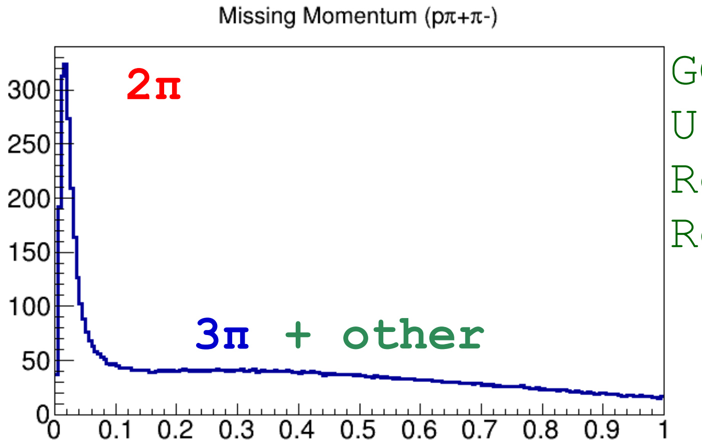
Exclusive variables





NOT FOR TRAINING WITH





GOAL :
Using PID variables
Retain all 3π
Remove backgrounds

ROOT-TMVA

Range of classifiers

- Rectangular cut optimization
- Projective likelihood estimation (PDE approach)
- Multidimensional probability density estimation (PDE - range-search approach)
- Multidimensional k-nearest neighbor classifier
- Linear discriminant analysis (H-Matrix and Fisher discriminants)
- Function discriminant analysis (FDA)
- Predictive learning via rule ensembles (RuleFit)
- Support Vector Machine (SVM)
- Artificial neural networks (MLP)
- Boosted/Bagged decision trees (BDT)
- (Newer) Deep NN, Convolutional NN,…
- Interface for R, Keras,…

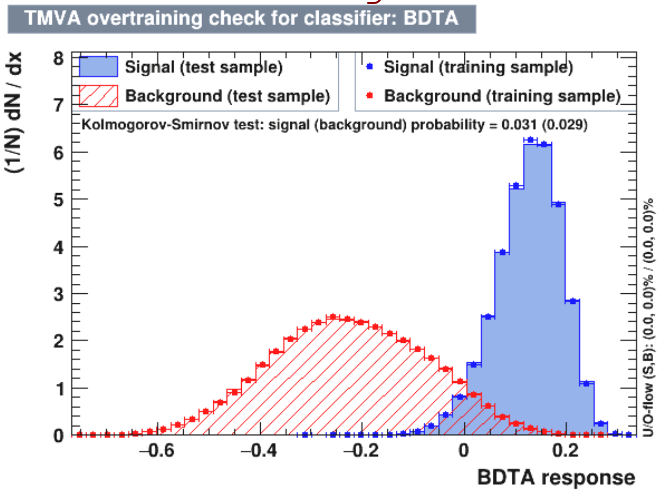
+ Jupyter ROOTbooks or PyROOT

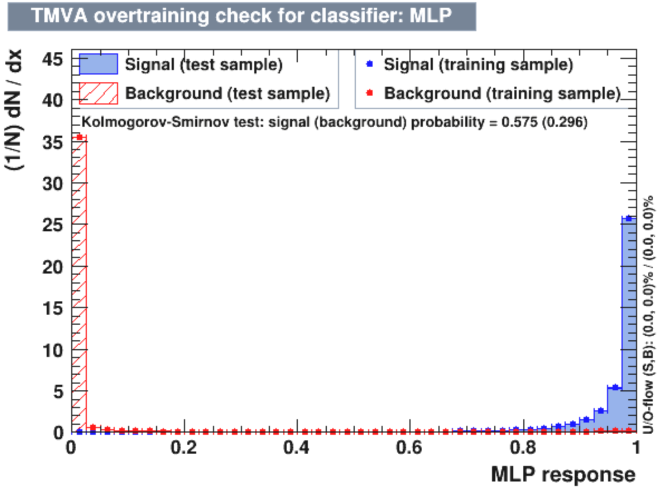
Train with Simulation

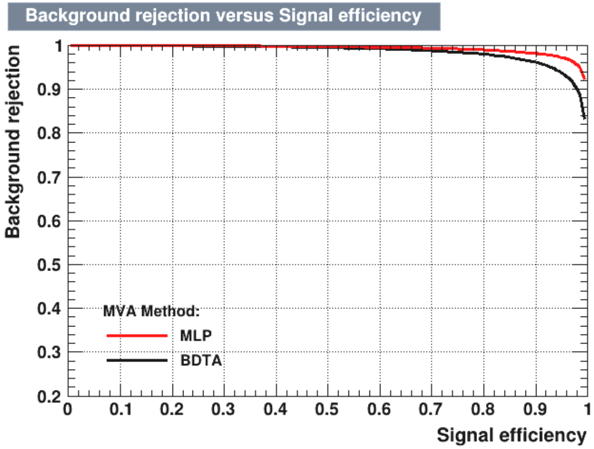
Signal : **Simulated** 3π (Correct Combitorial)
Background : All Experimental Data Events +(Simulated Wrong Combitorial)

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

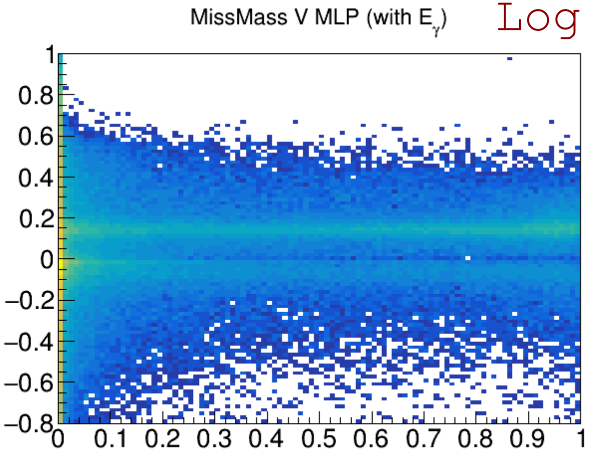
200k Training and Test data



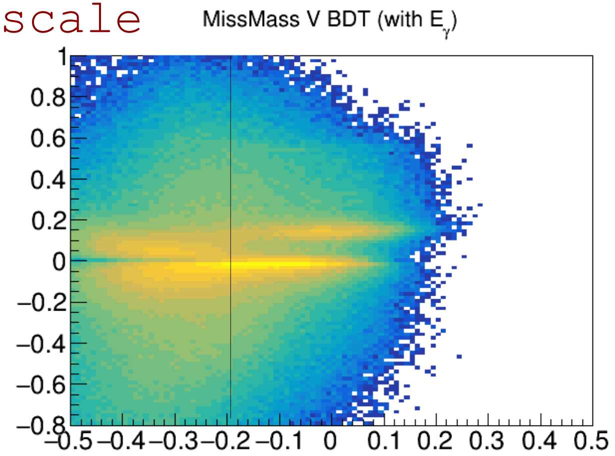




Apply Simulated training



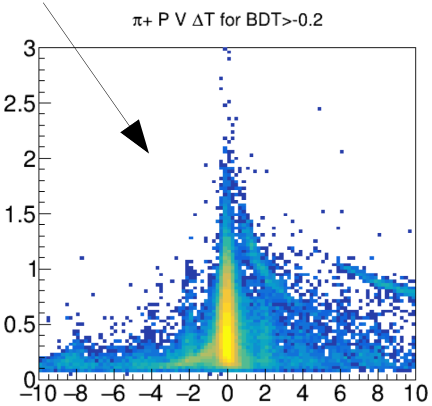
Log z scale

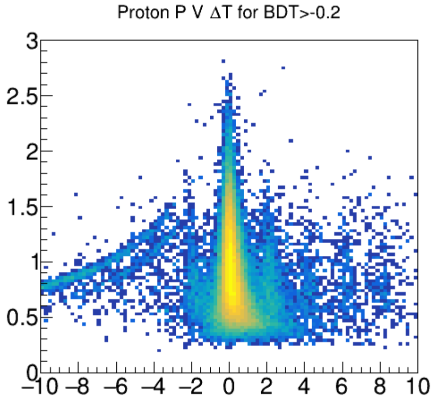


Real Experiment distributions shifted to background values compared to Simulation

Simulation must be very Realistic!

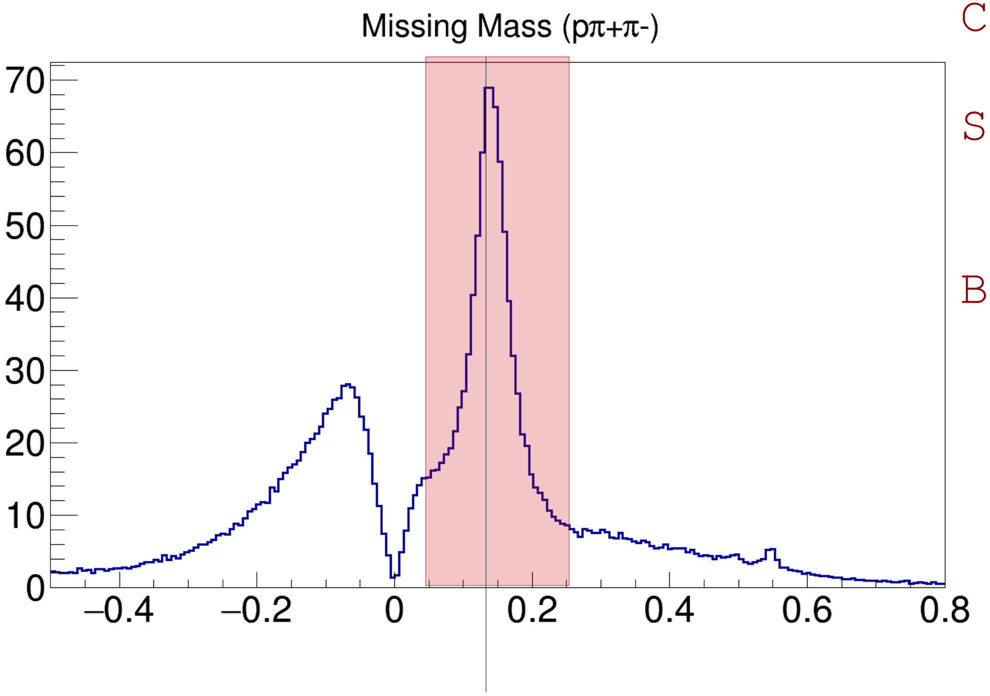
Small seperation between 2π and 3π
- OK PID variables similar





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

Mixed events training



M(π^0)

Cut on region with high signal density

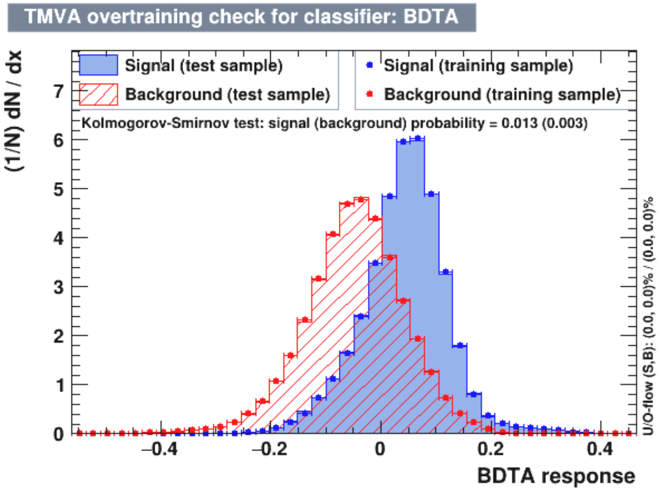
Signal = (Missing mass peak ~ π^0)

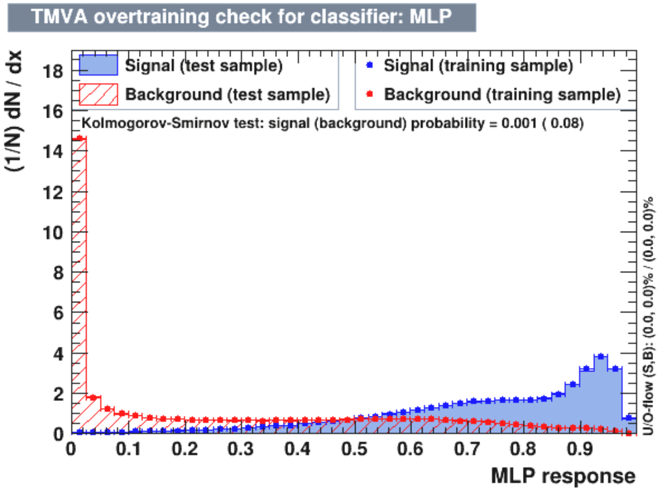
Background = Everything else

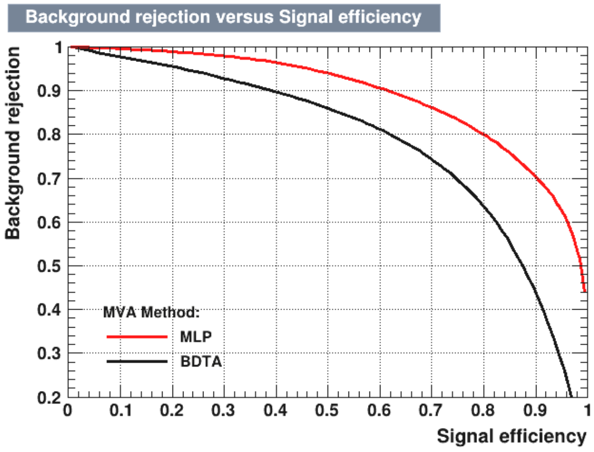
Train with Mixed Events (I)

Signal : 0.1<MissMass<0.2
Background : !Signal

Variables + E_γ



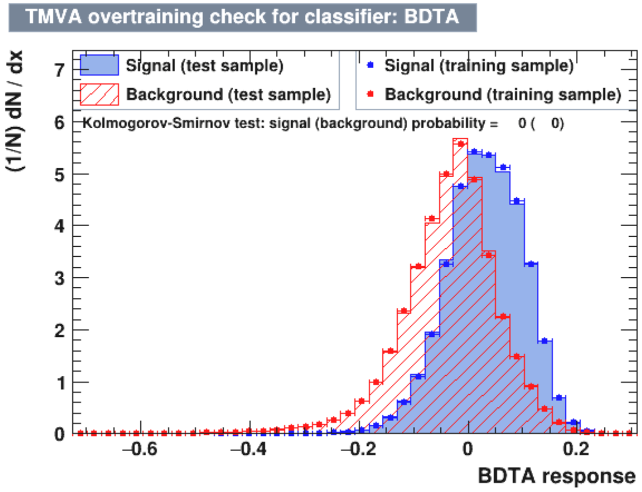


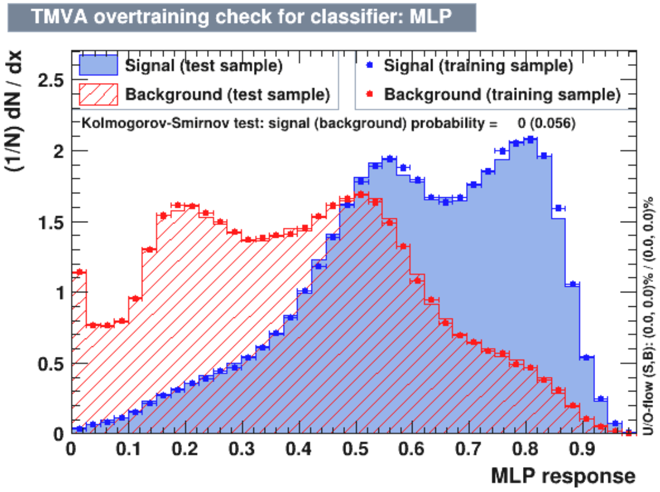


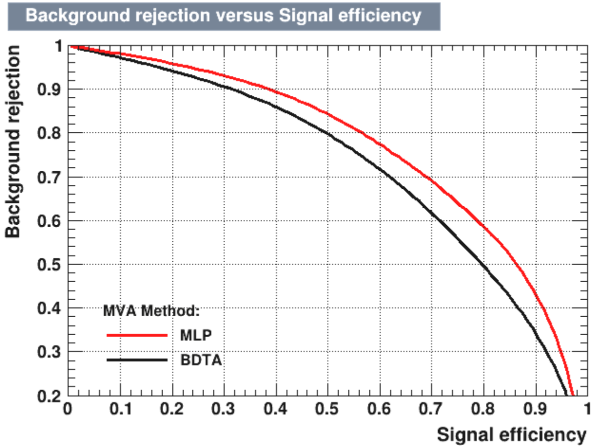
Train with Mixed Events (II)

Signal : 0.1<MissMass<0.2
Background : !Signal

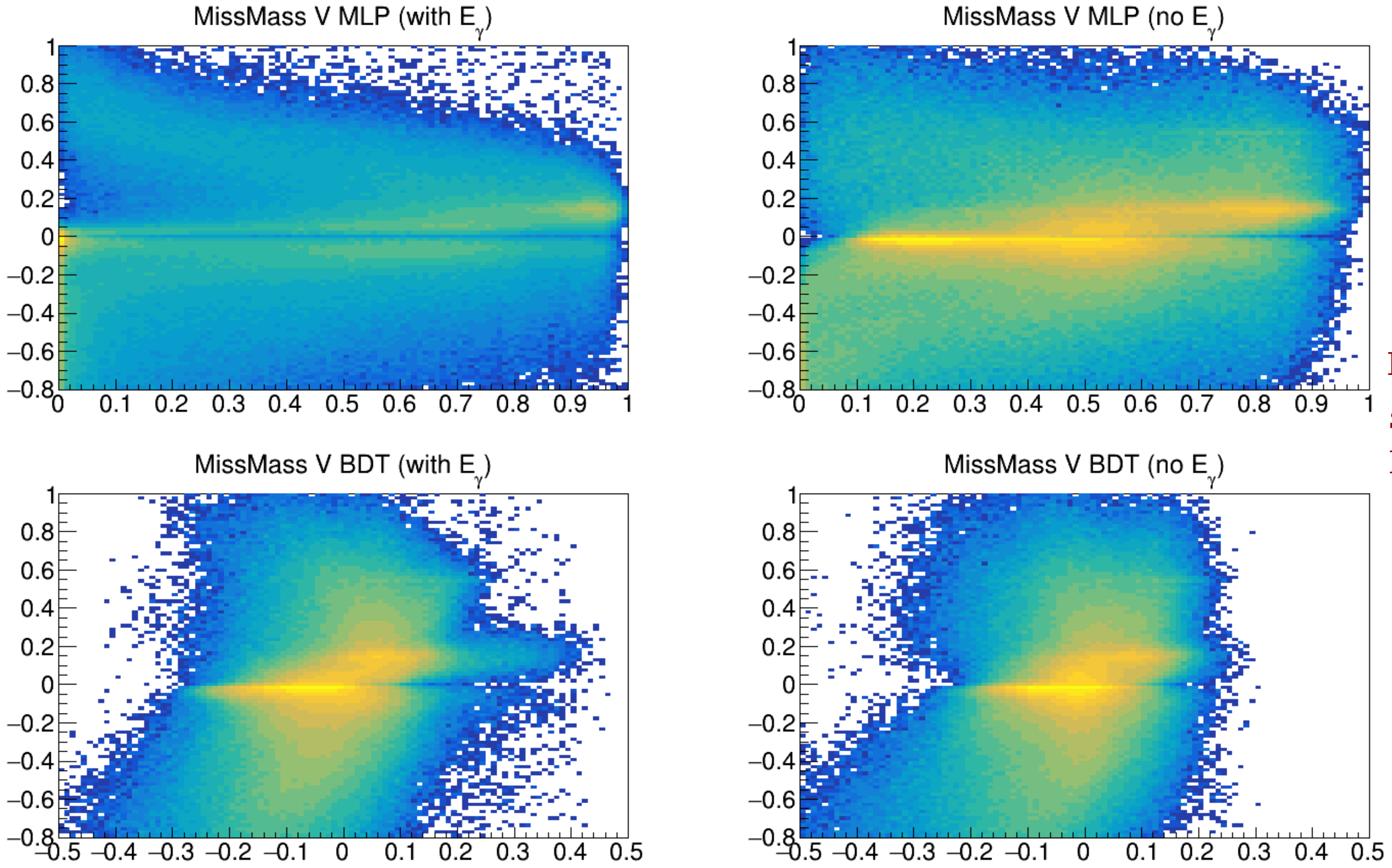
Variables - E_γ







Miss Mass V Responses

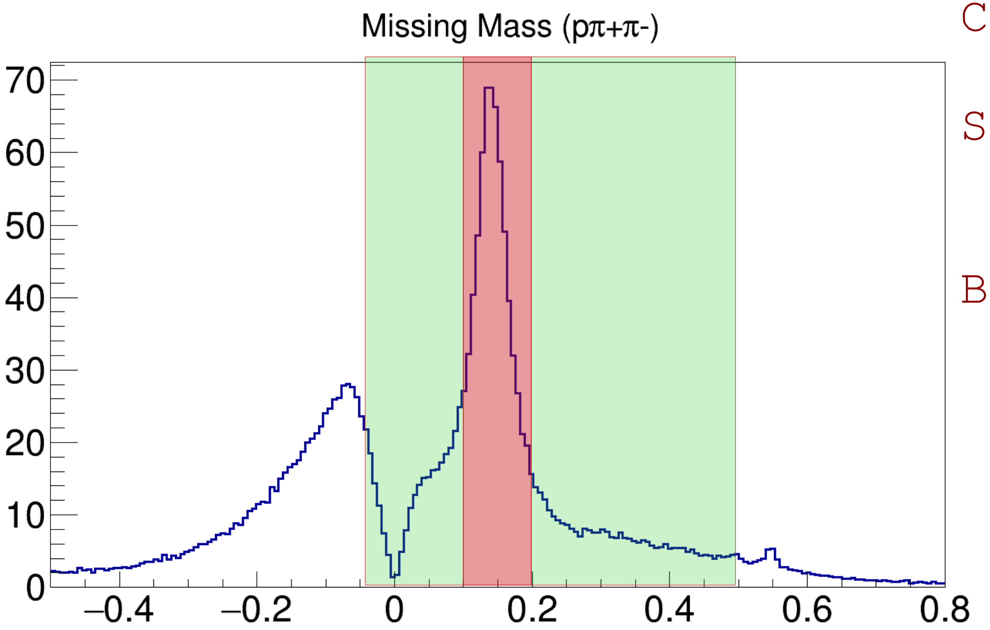


Visually MLP gives Better seperation (as with ROCs)

E_γ input gives greater seperation of 2π – 3π Particularly for MLP

E_γ input creates Correlation of Response with Exclusivity variable

Mixed events training



Cut on region with high signal density

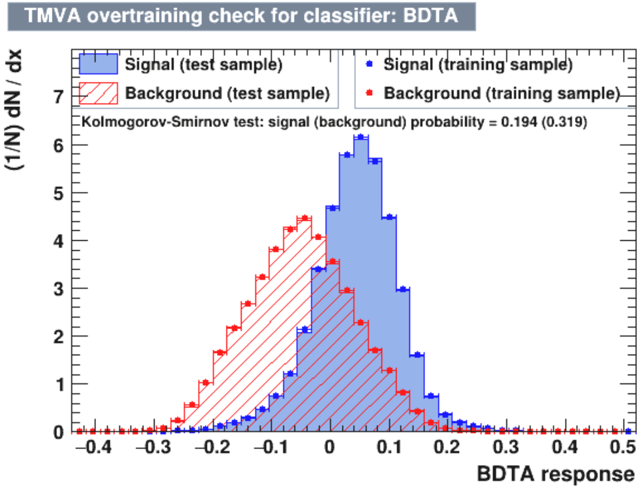
Signal = (Missing mass peak ~ π^0)

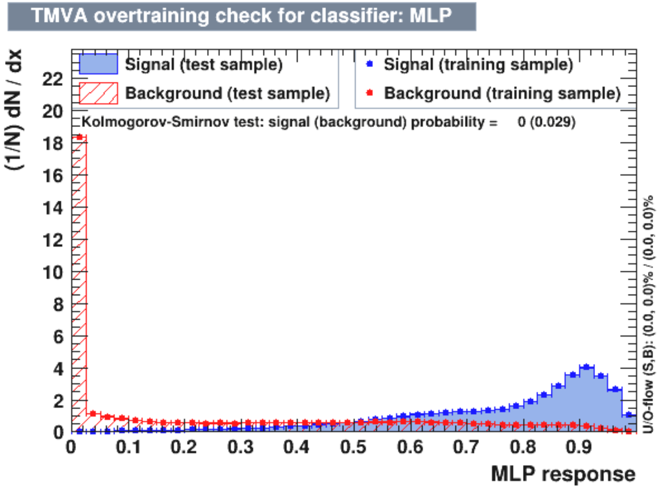
Background = sidebands inc 2π

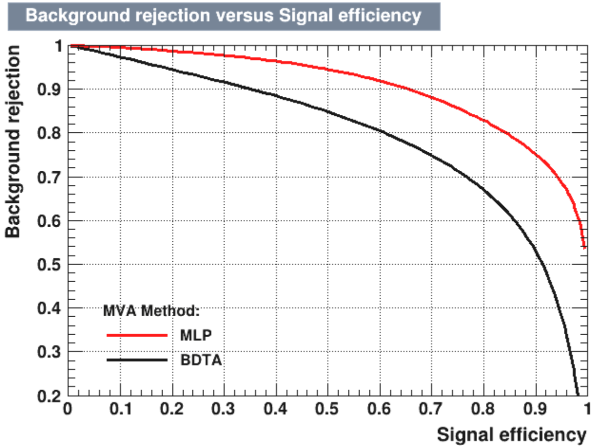
Train with Mixed Events (III)

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

Variables + E_{γ}



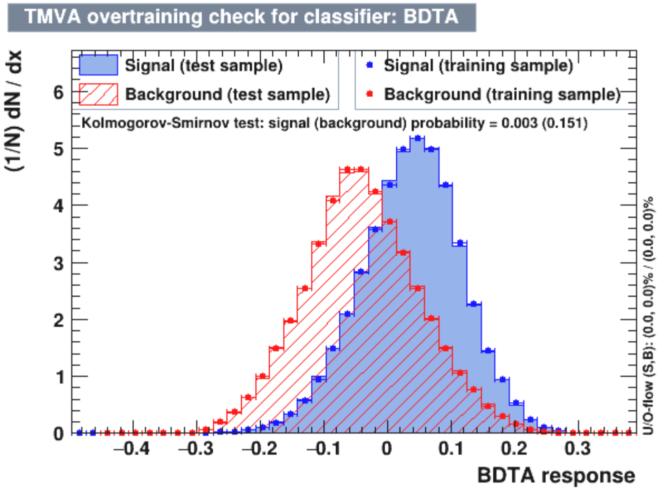


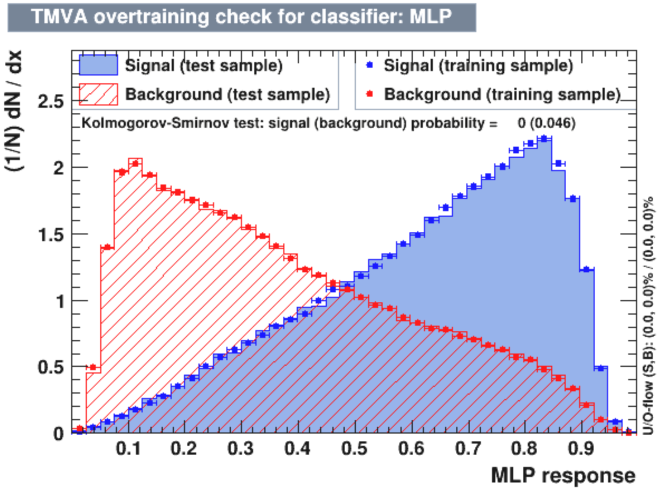


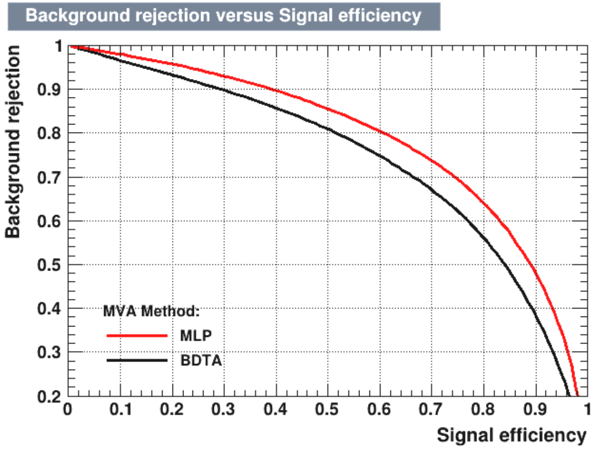
Train with Mixed Events (IV)

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

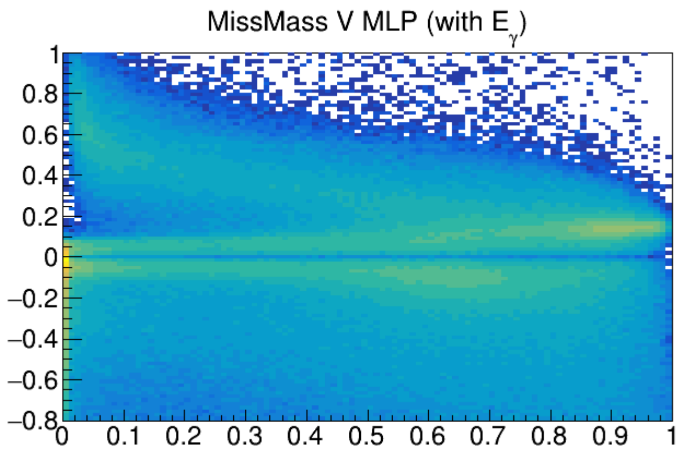
Variables - E_γ

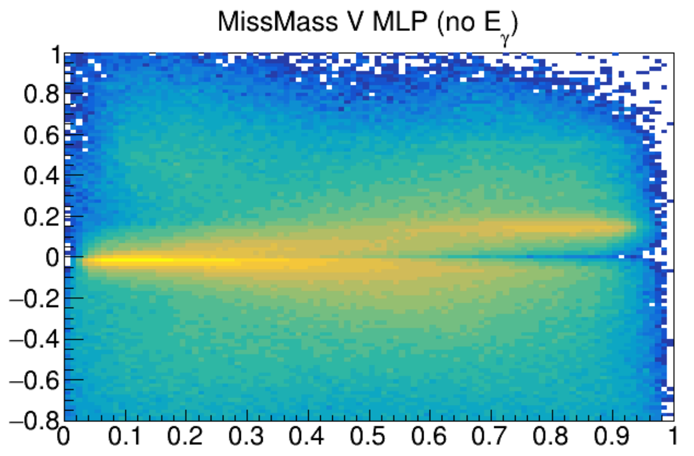


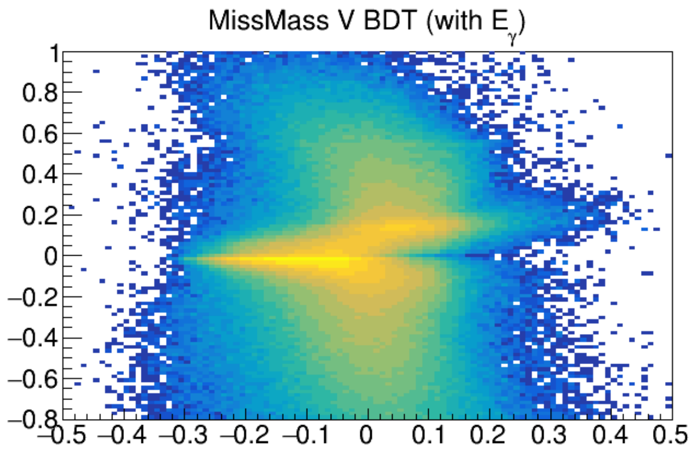


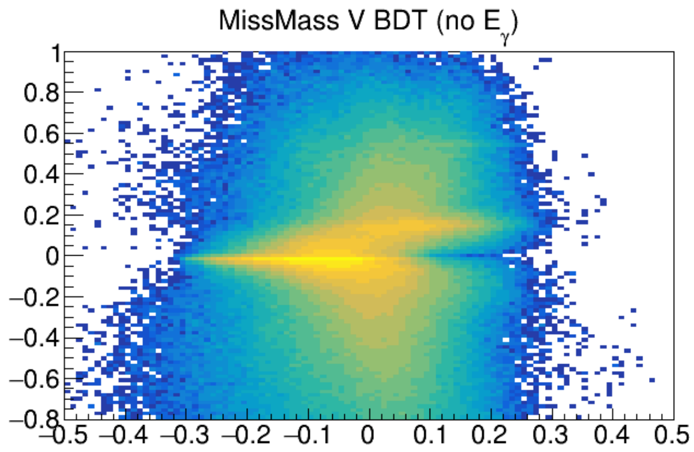


Miss Mass V Responses







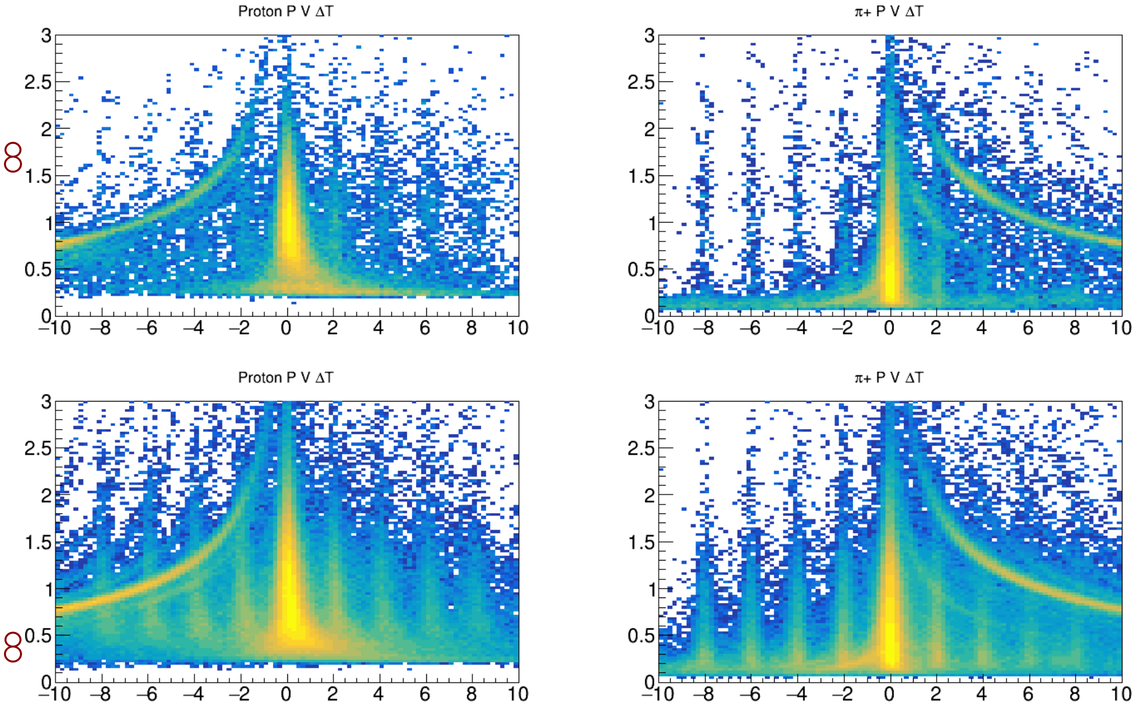


These background events give better discrimination of 2π

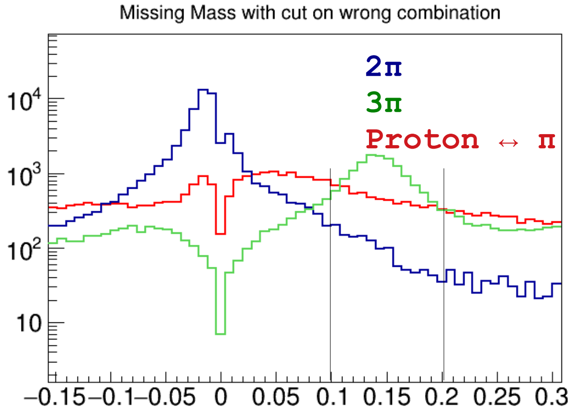
MisIDed particles

TOP:
MLP>0.8

BOTTOM
MLP<0.8



Still see significant Background from randoms and wrong combinations



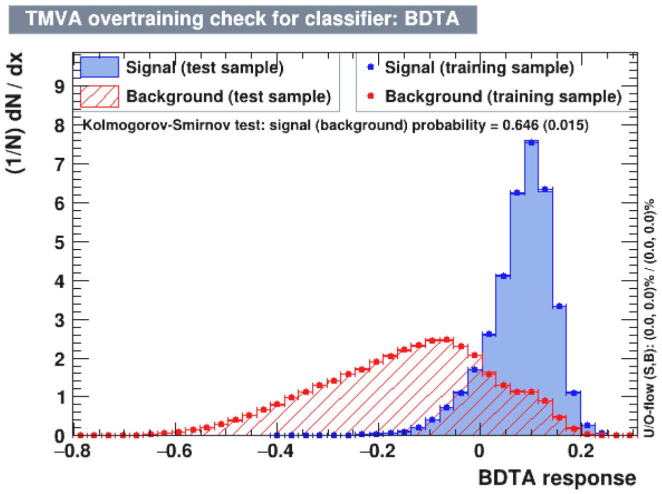
Wrong combi can give relatively high contribution to "signal" region

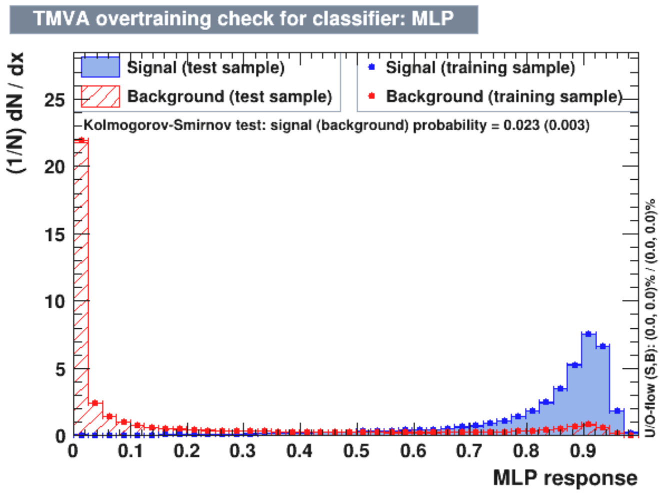
Train with Mixed Events (III)

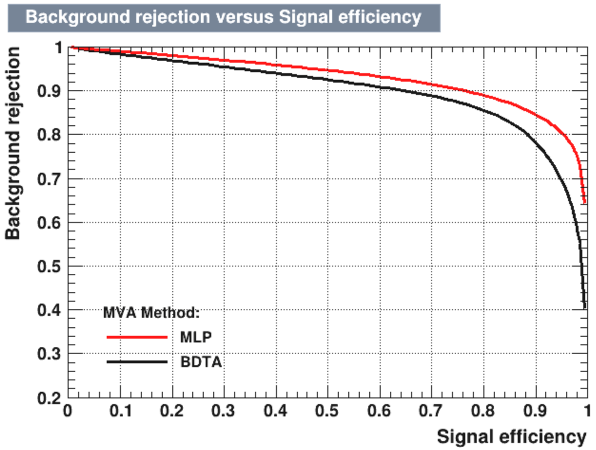
Signal : 0.1<MissMass<0.2 && **Simulated BDTA>-0.2**
Background : -0.04<MissMass<0.1 and 0.2<MissMass<0.5

Variables + E_{γ}

Simulated training was able to remove wrong combinations





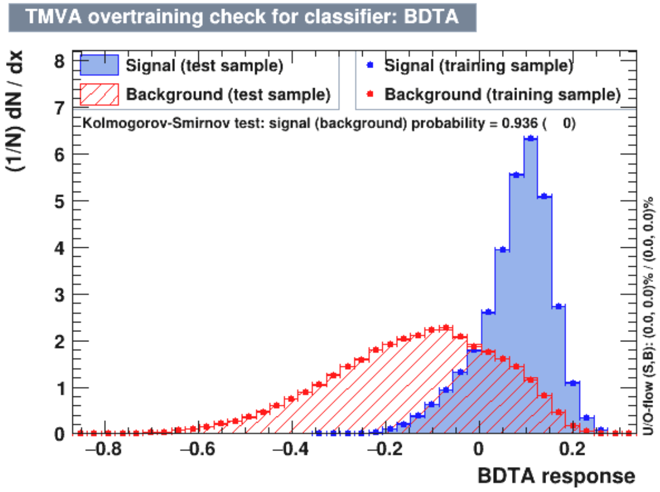


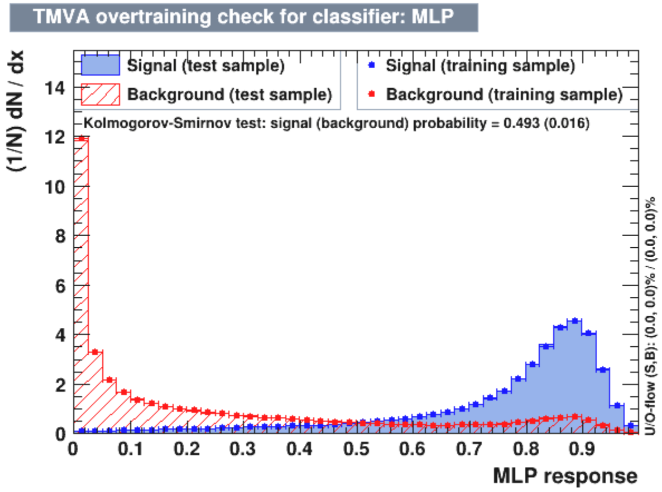
Train with Mixed Events (IV)

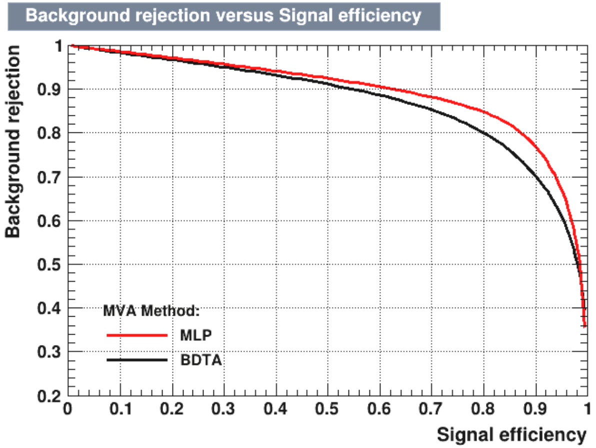
Signal : 0.1<MissMass<0.2 && **Simulated BDTA>-0.2**

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

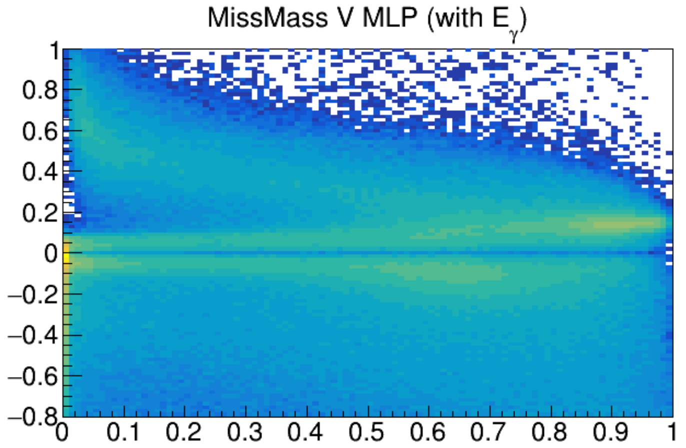
Variables - E_γ

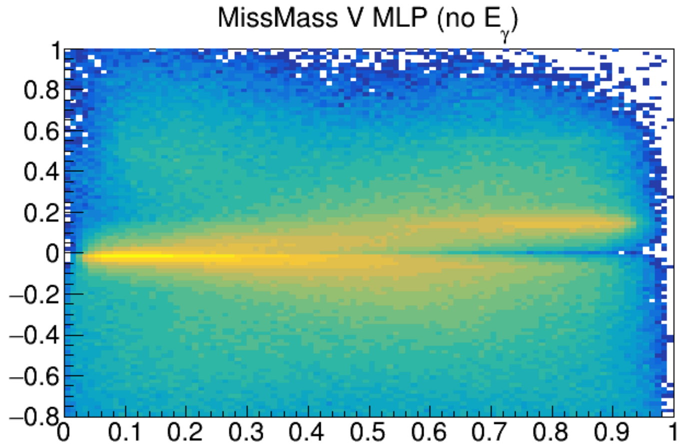




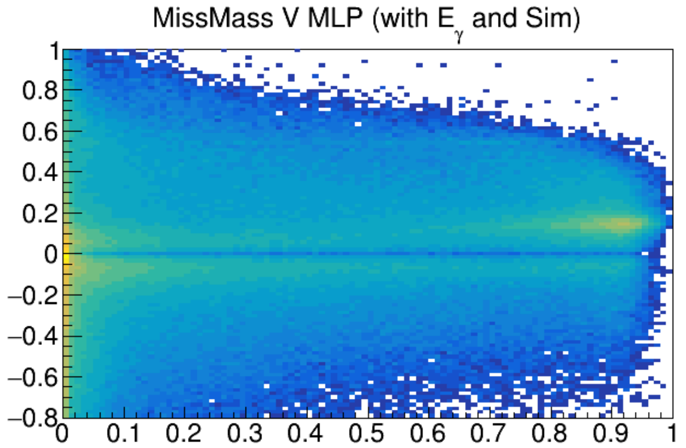


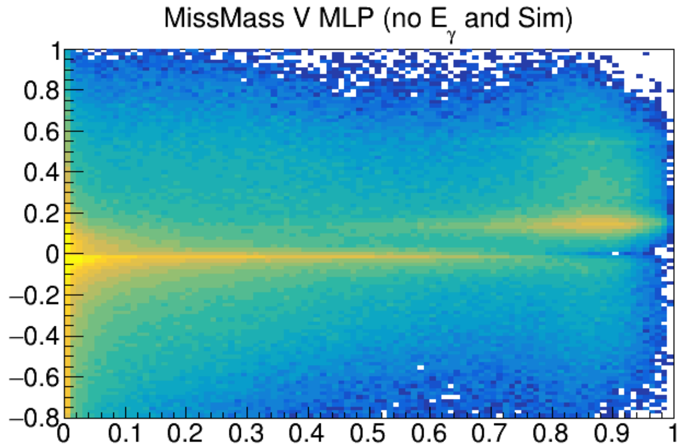
Miss Mass V Responses





TOP : Mixed Training with sidebands

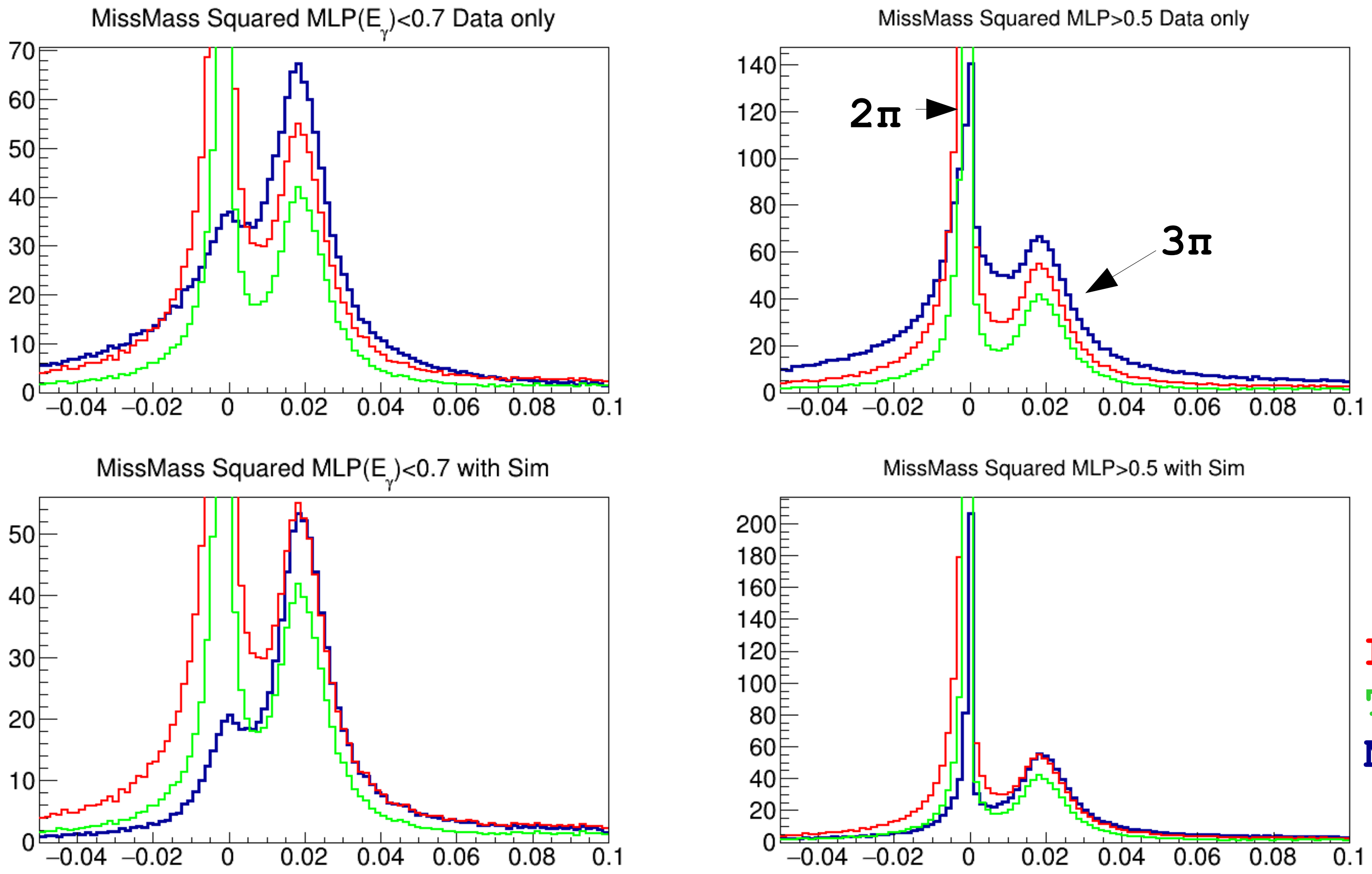




BOTTOM : additional Filter for signal Training events SimBDTA>-0.2

- less “other” (random,wrong combi) backgrounds

Effect on Missing Mass Squared



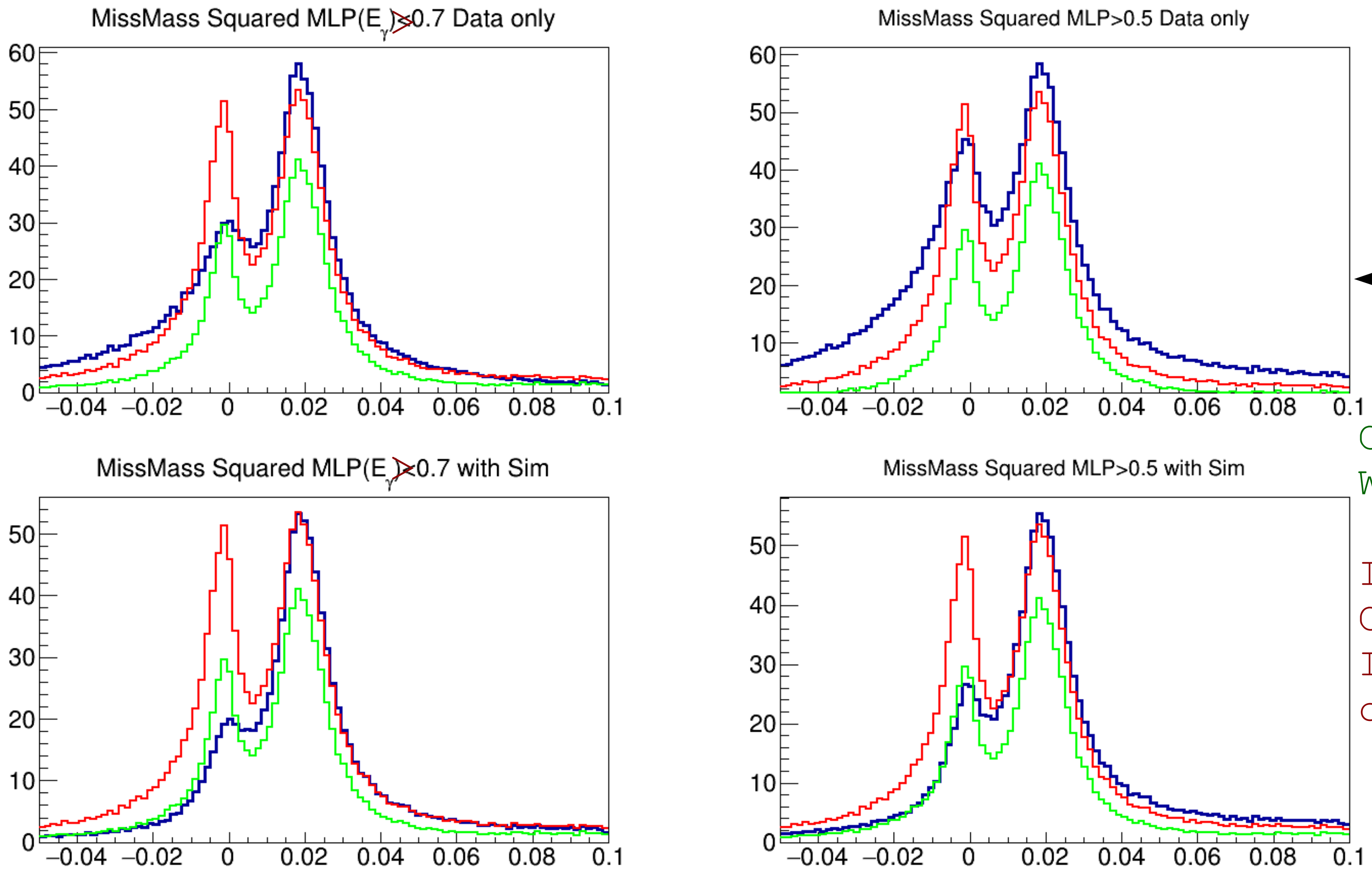
TOP : Mixed Training with sidebands

BOTTOM : additional Filter for signal Training events SimBDTA>-0.2

Only 3π => single peak ~0.02

Loose Cuts Based(3ns)
Tight Cuts Based(1ns)
MLP Cut

With extra exclusivity cut



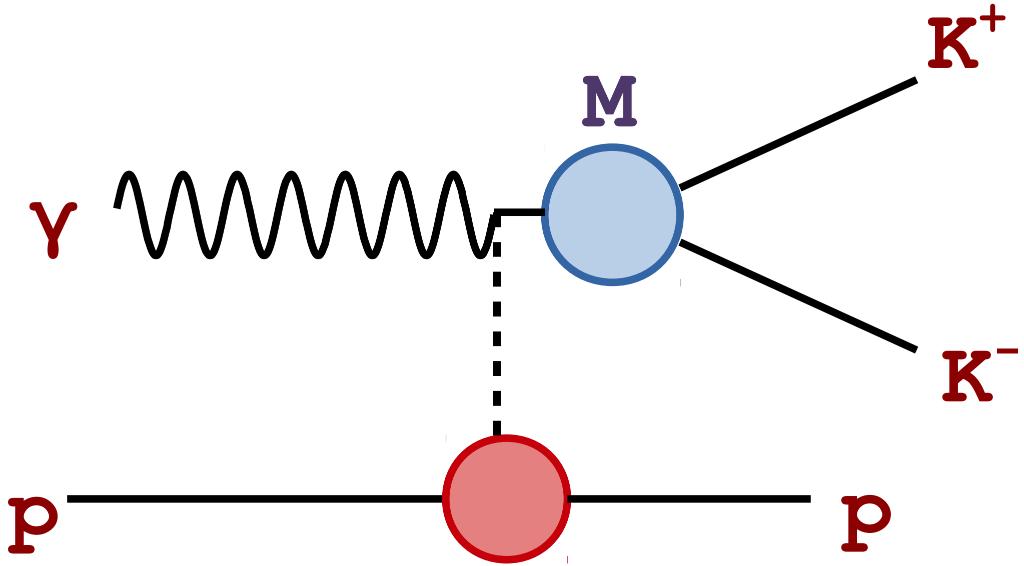
Additional exclusivity cut MissEnergy>0.1

Without SimBDT filter
Additional BG in MLP cuts

Cleaned up
With Sim BDT filter

Improved rejection of 2π
Could also be done with
Improved exclusivity cuts/kinematic fit,…

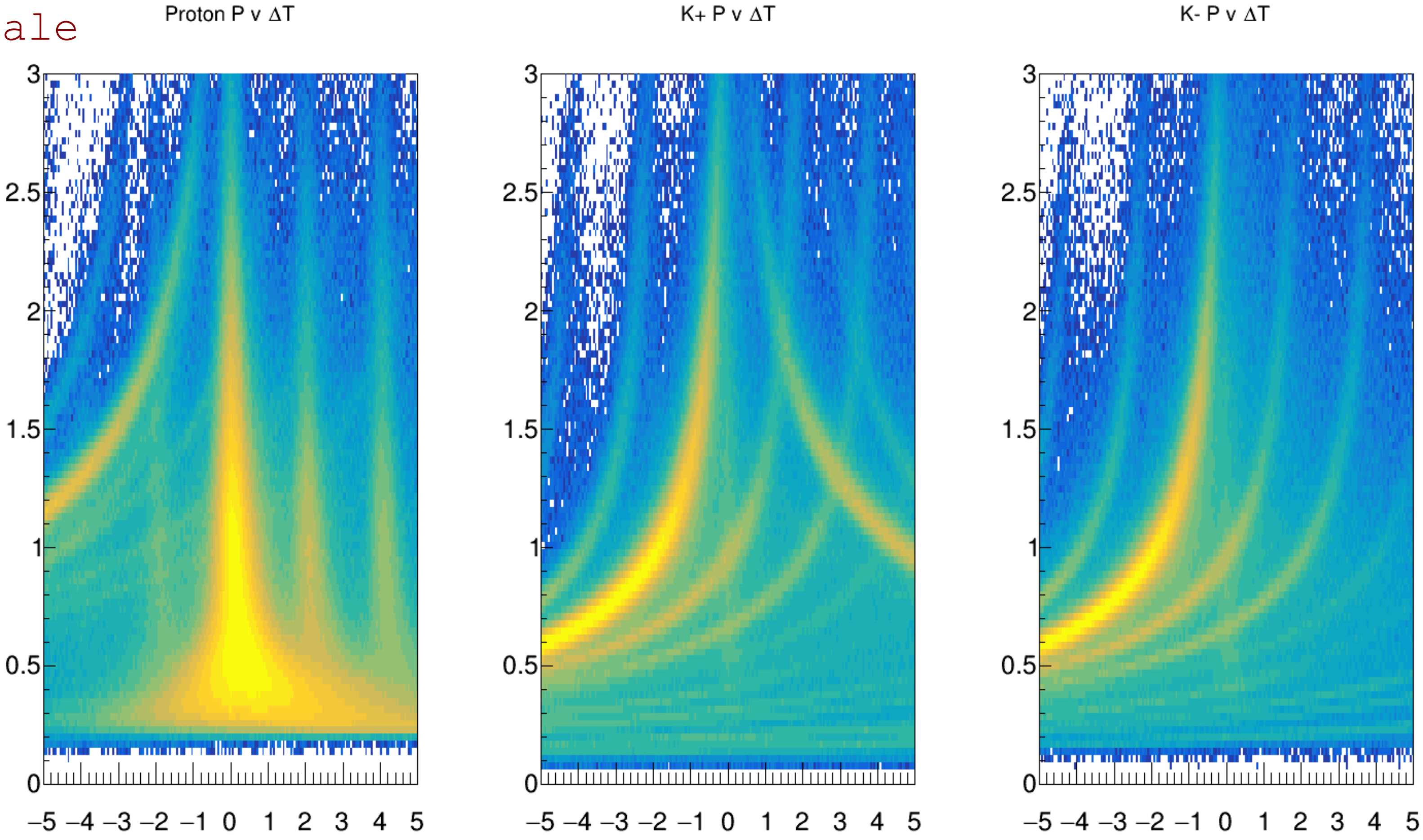
Example Reaction



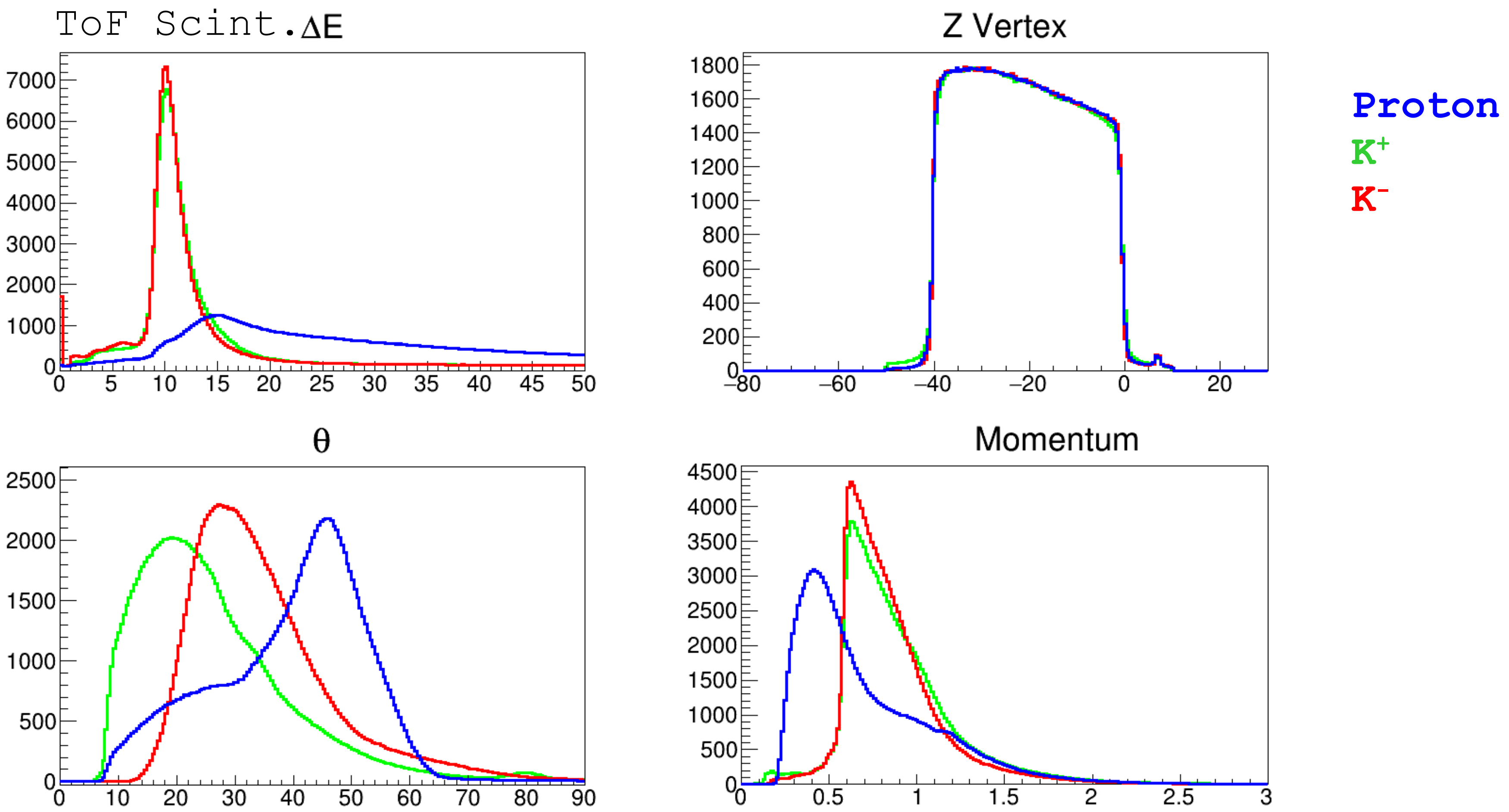
Large background from
Pions IDed as Kaons

ΔTime PID

*Log scale



Other Input Variables



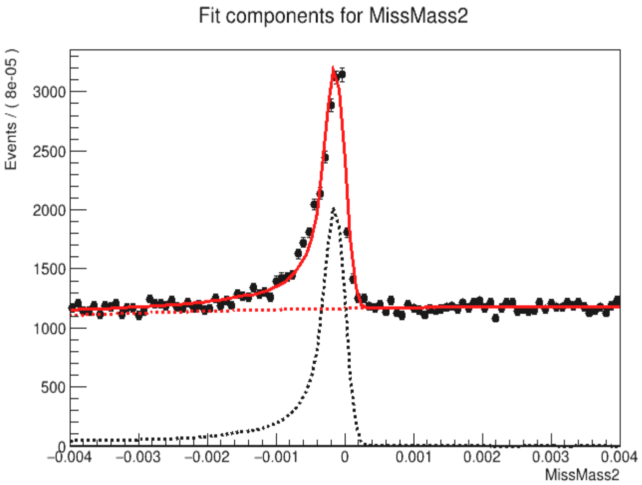
Train BDT with sWeights

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

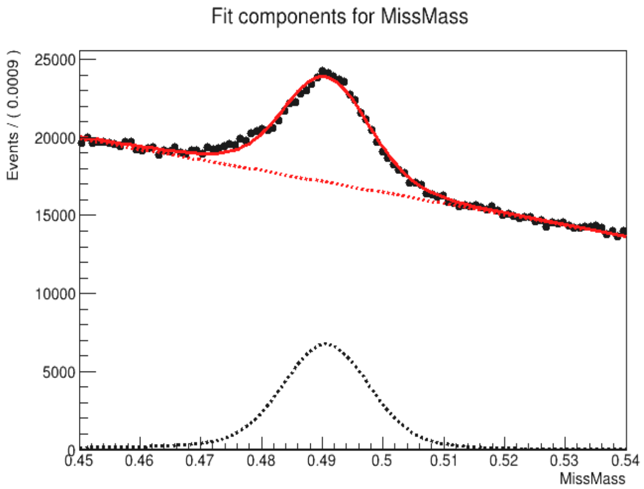
M. Pivk,F.R. Le Diberder,Nucl.Inst.Meth.A 555, 356-369, 2005

Used RooStats implementation

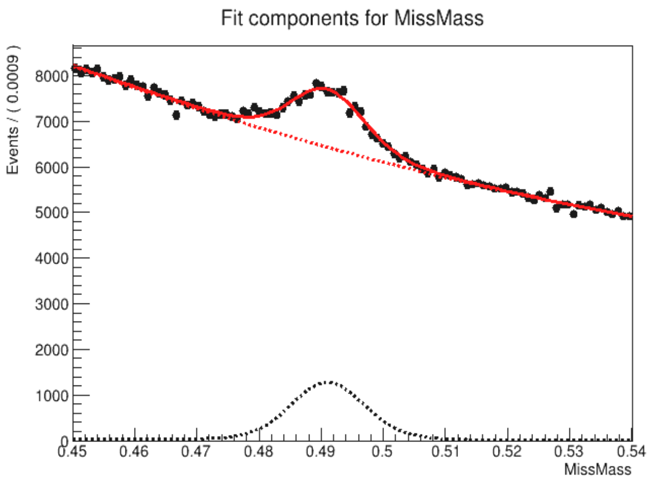
Exclusive



Missing K-



Missing K+



In TMVA BDT accept negative event weights

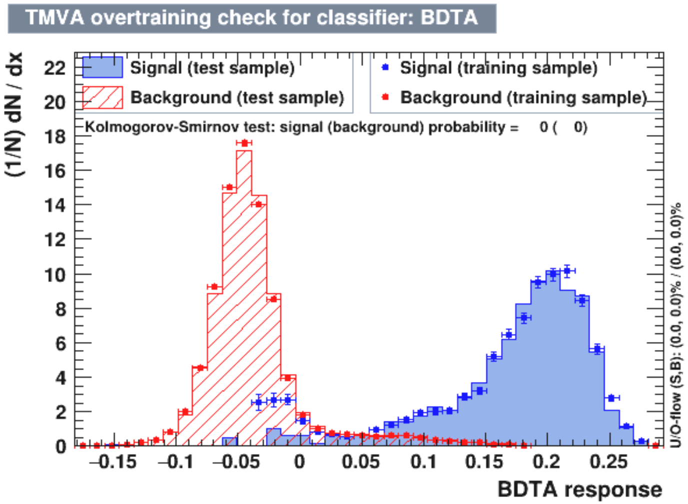
MLP does not

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

Train BDT with sWeights

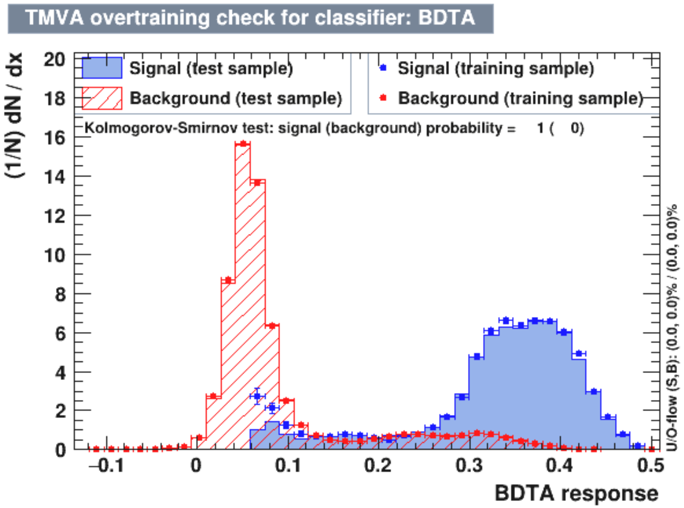
80k test and 20k train

Exclusive



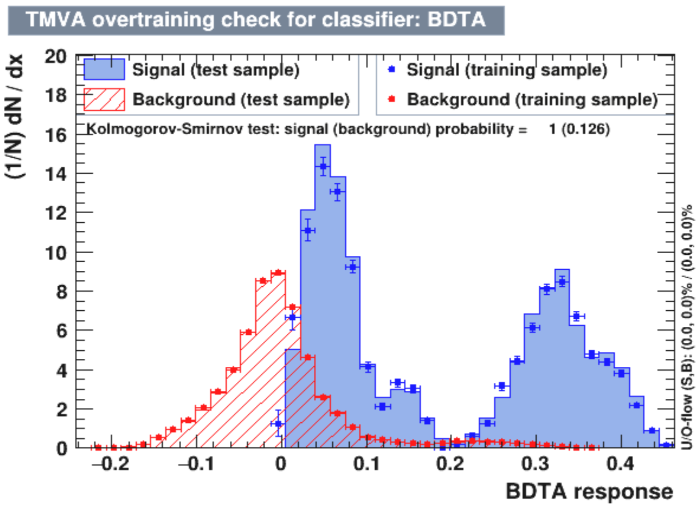
300k test and train

Missing K-



300k test and train

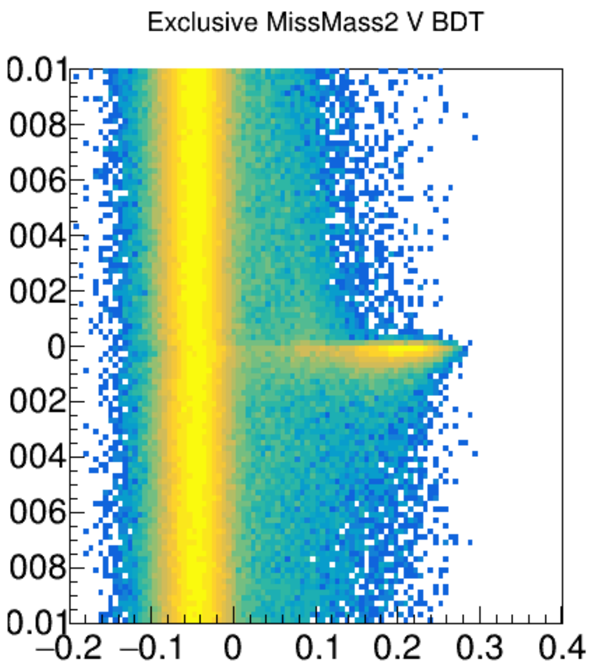
Missing K+

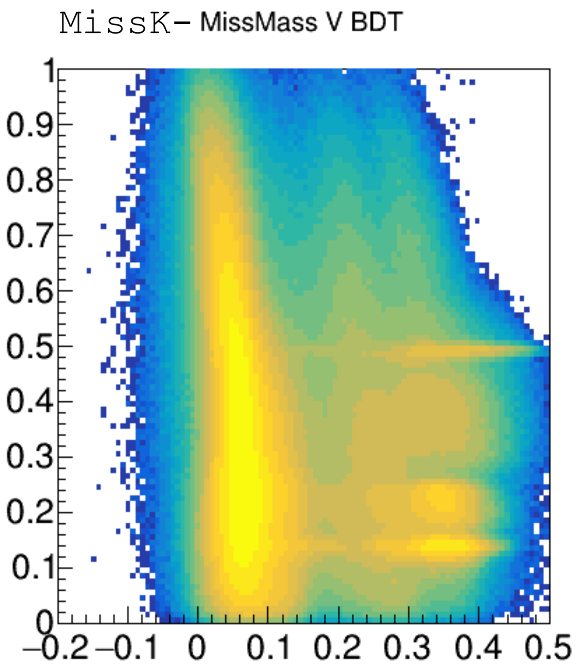


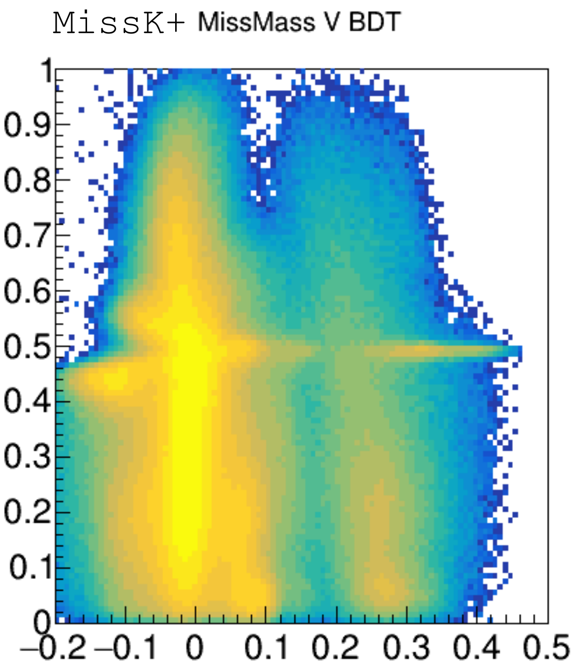
2 types of events in Missing K+ signal weights

Seperated by training

Exclusivity variables V BDT

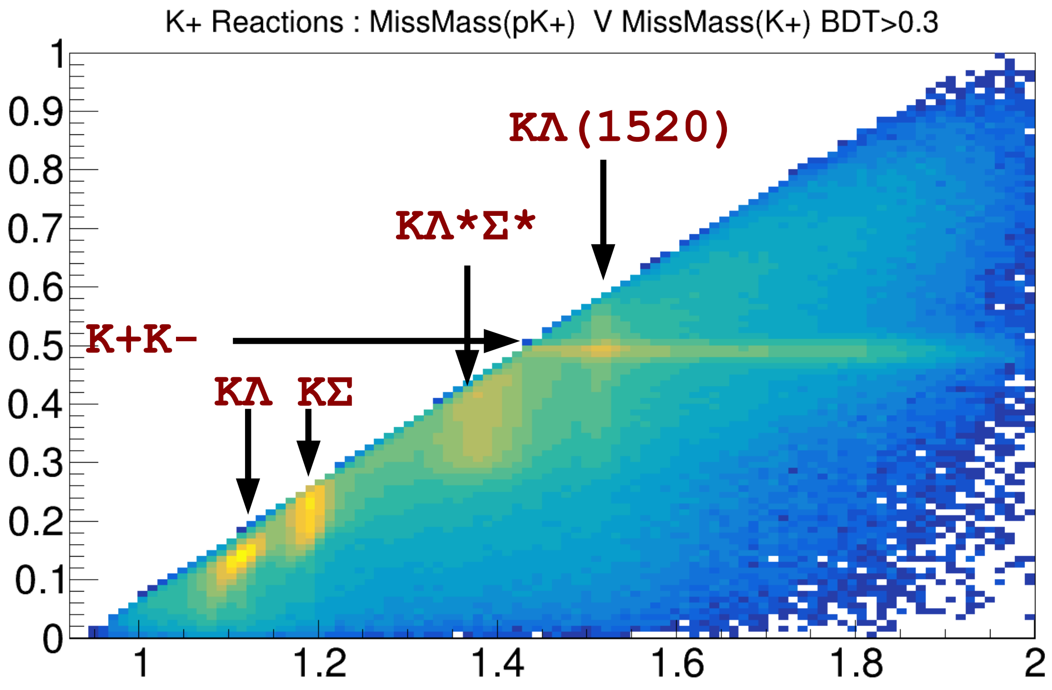




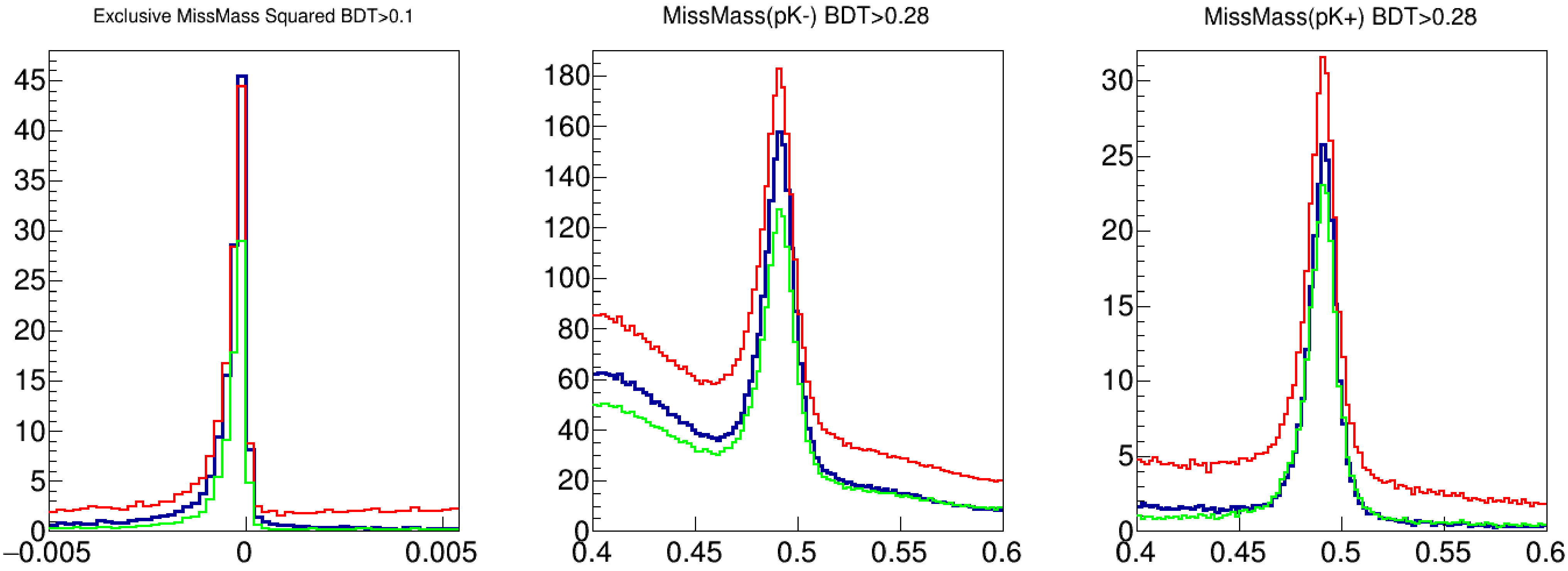


K+ production mechanims

Missing K- Topology



Event exclusivity with BDT cut

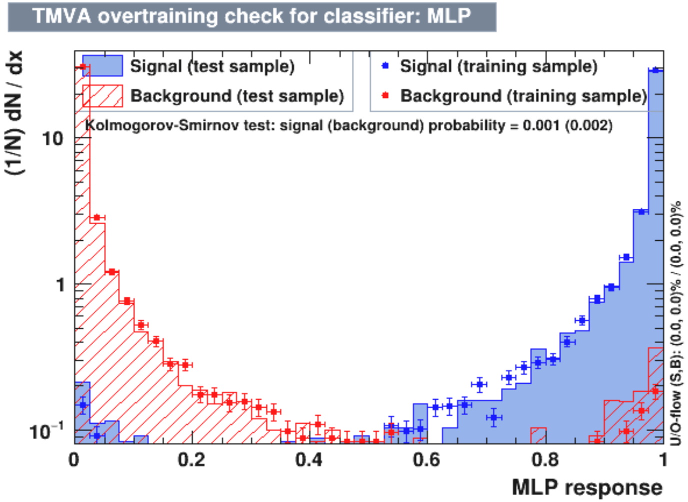


Loose Cuts Based(1ns)
Tight Cuts Based(0.5ns)
BDT Cut

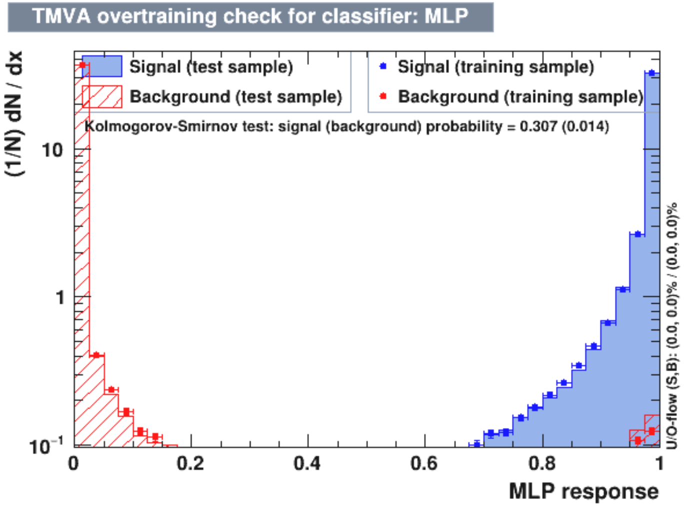
Train MLP with BDT

MLP ~ factor 10 faster to apply

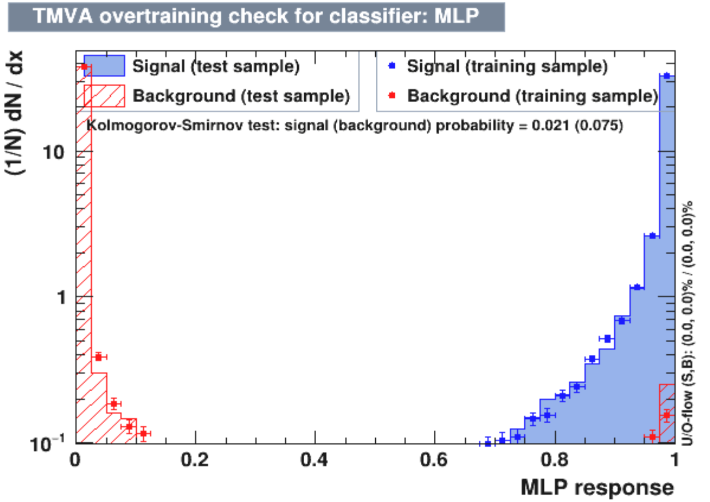
Exclusive



Missing K-



Missing K+

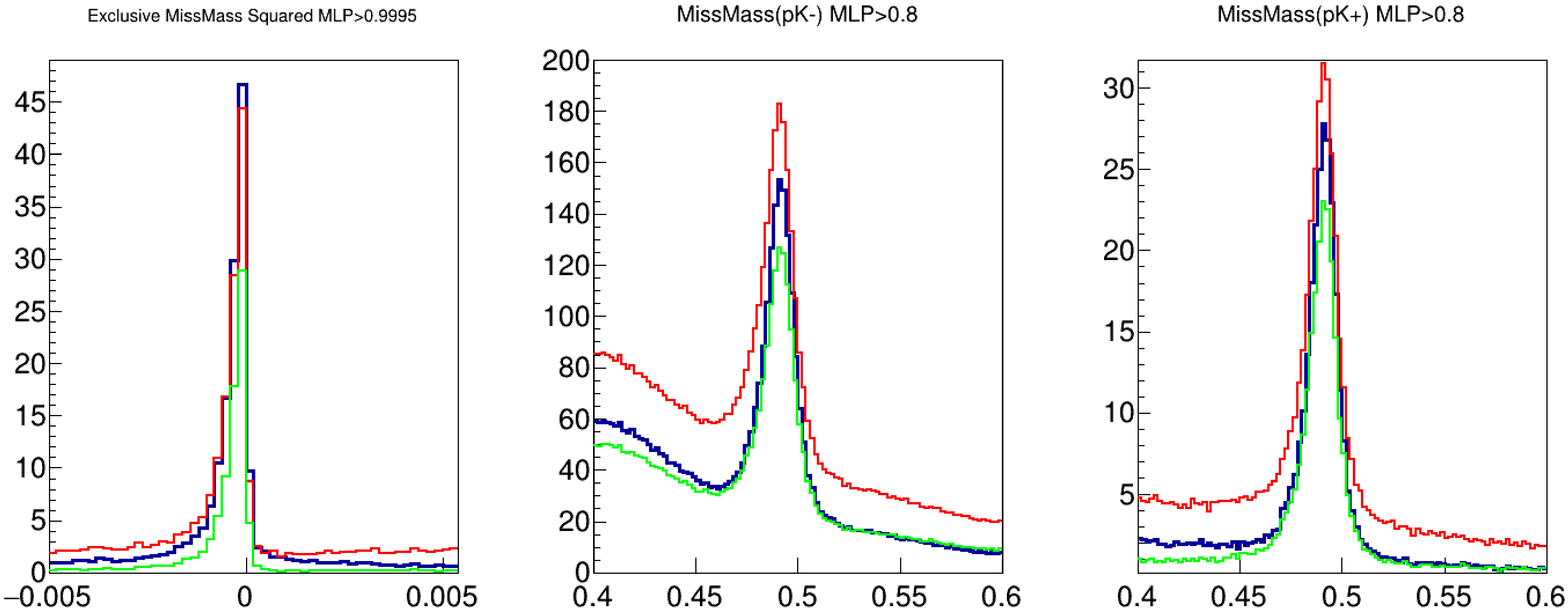


Signal BDT>0.05
Backgr BDT<0.05

Signal BDT>0.3
Backgr BDT<0.3

Signal BDT>0.3
Backgr BDT<0.3

Event exclusivity with MLP cut

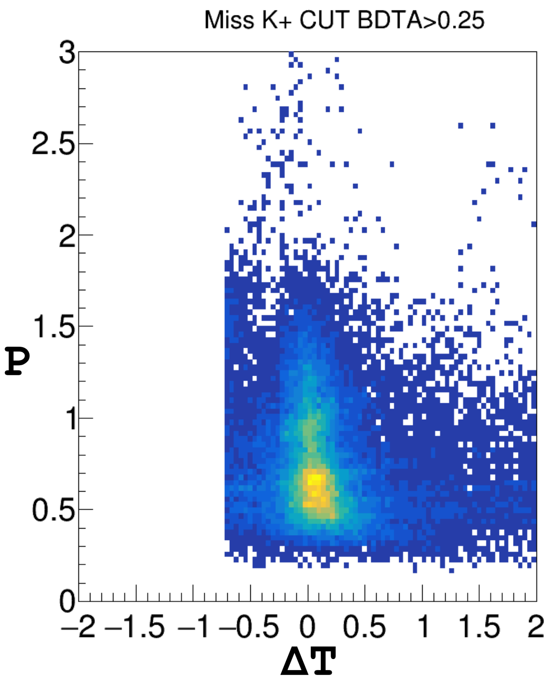


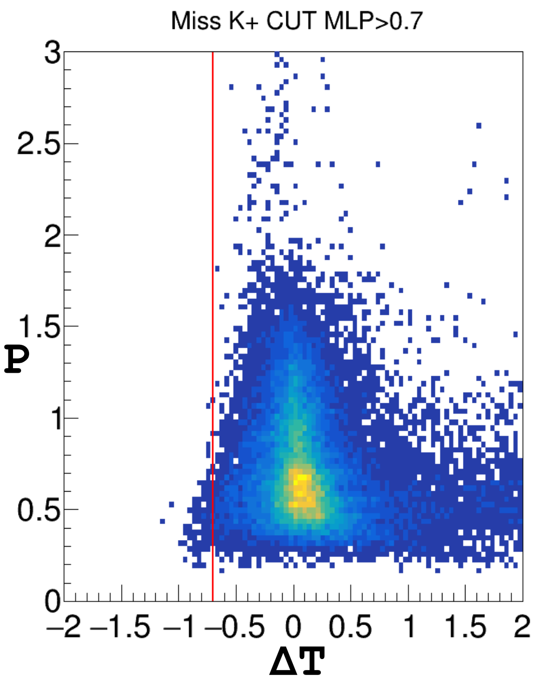
Loose Cuts Based(1ns)

Tight Cuts Based(0.5ns)

MLP Cut

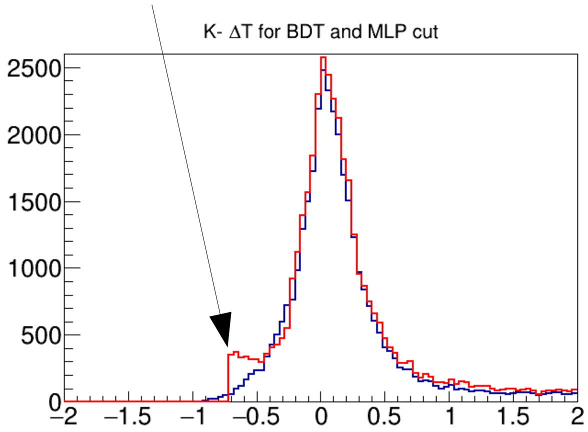
MLP and correlations



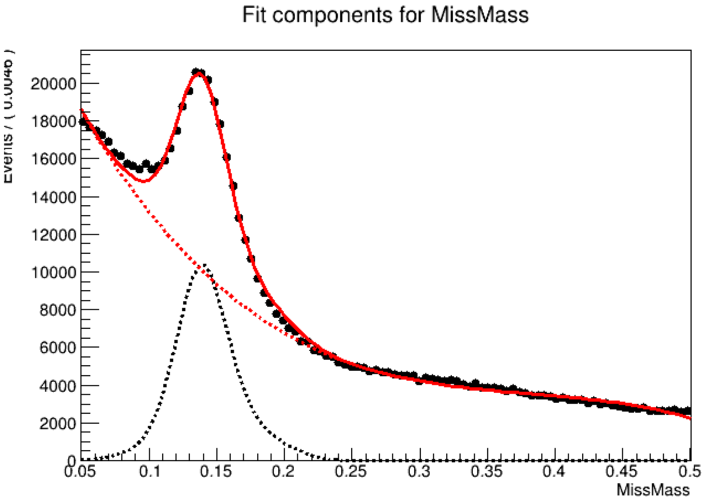


MLP better handles
Correlations
Within inputs

But small improvement
Very few events
Actually removed

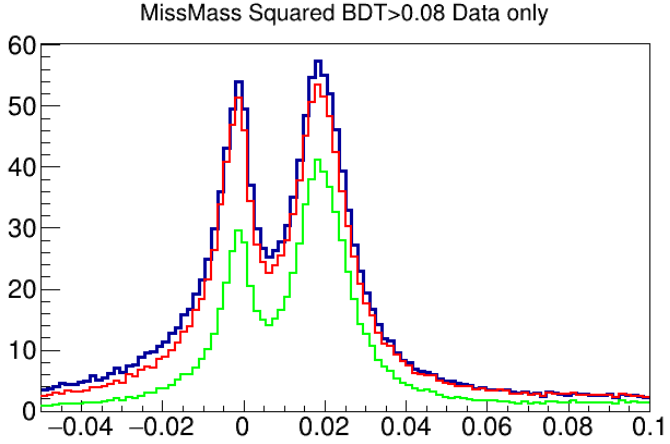


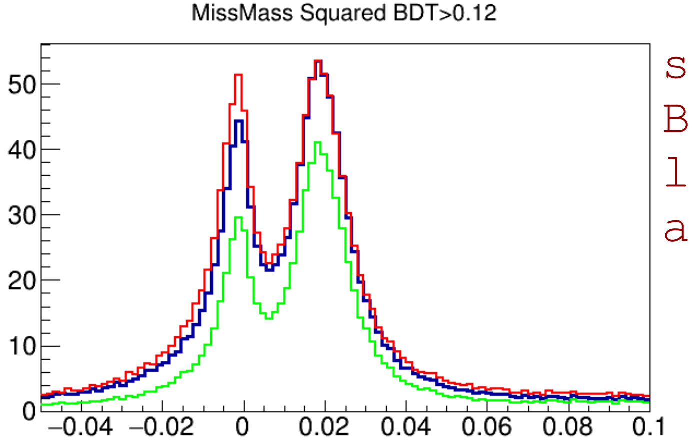
Using sPlot training with 3π



Relatively small sample of 2π background
In fit range

Beam Energy not used as input





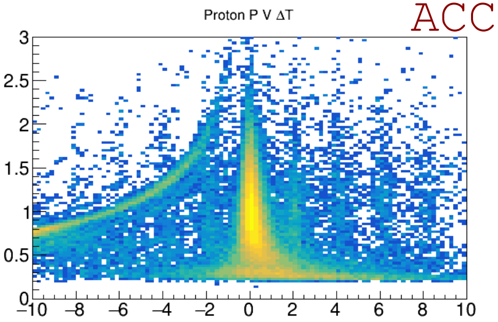
sPlot trained
BDT at
least as good
as straight cuts

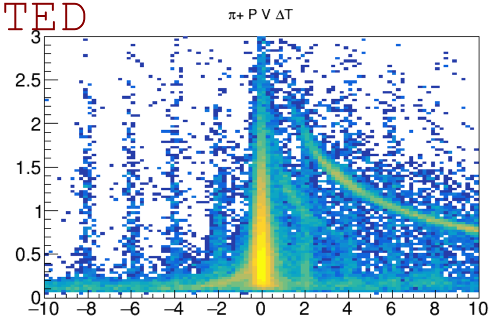
sPlot Training for 3π

Mixed Event Training

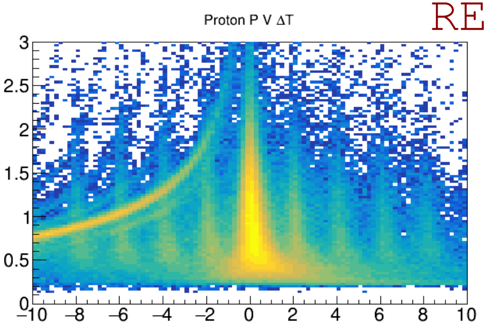
Splot Training

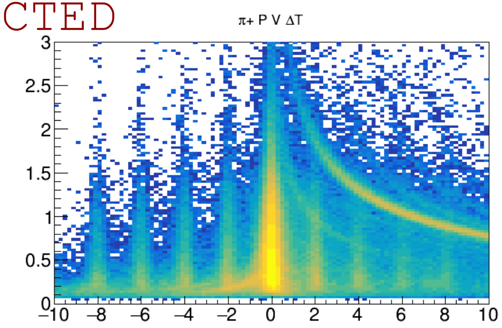
ACCEPTED



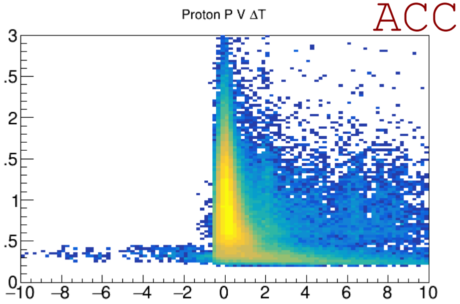


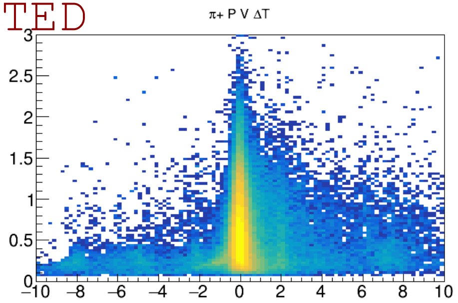
REJECTED



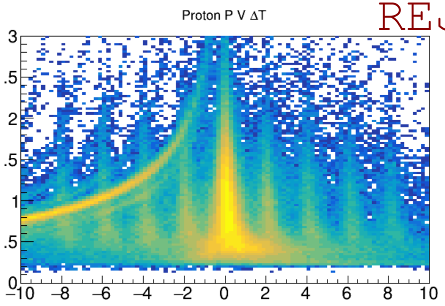


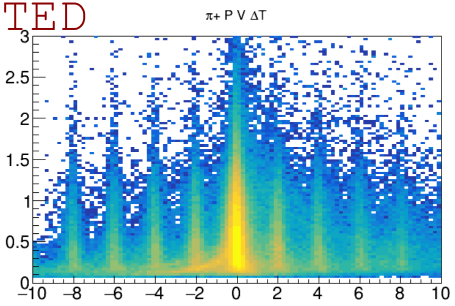
ACCEPTED





REJECTED

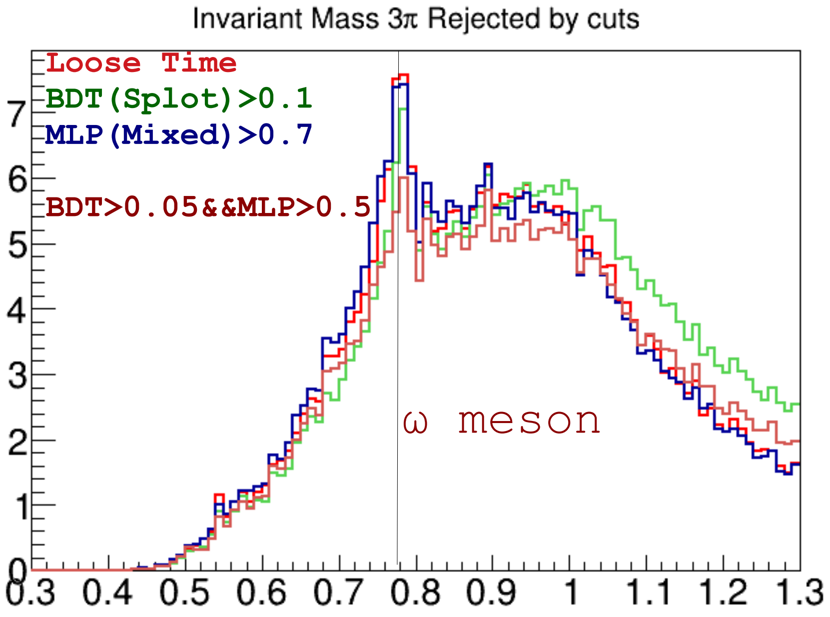


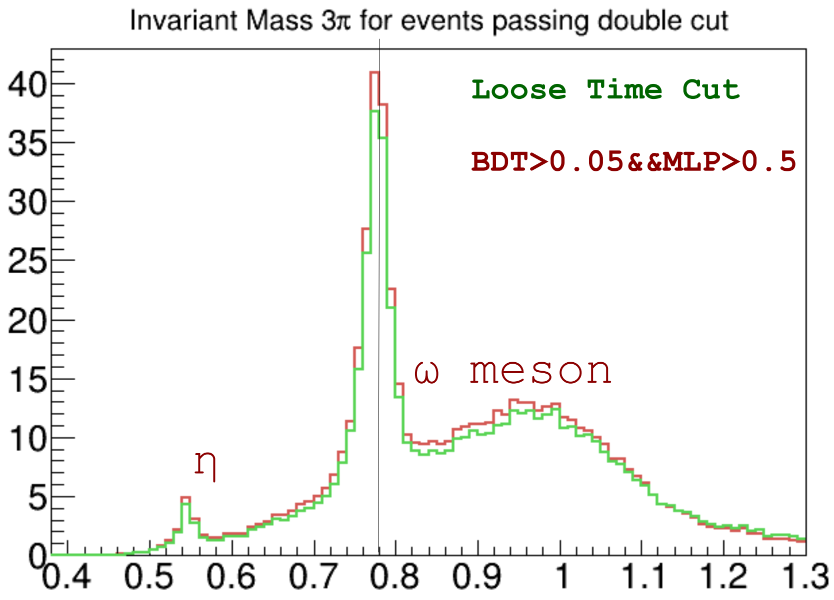


Fixes wrong combination problem
Wrong combination doesn't peak
At pi0 mass

But what about false negatives?

Check with 3π invariant mass, which should peak at ω mass

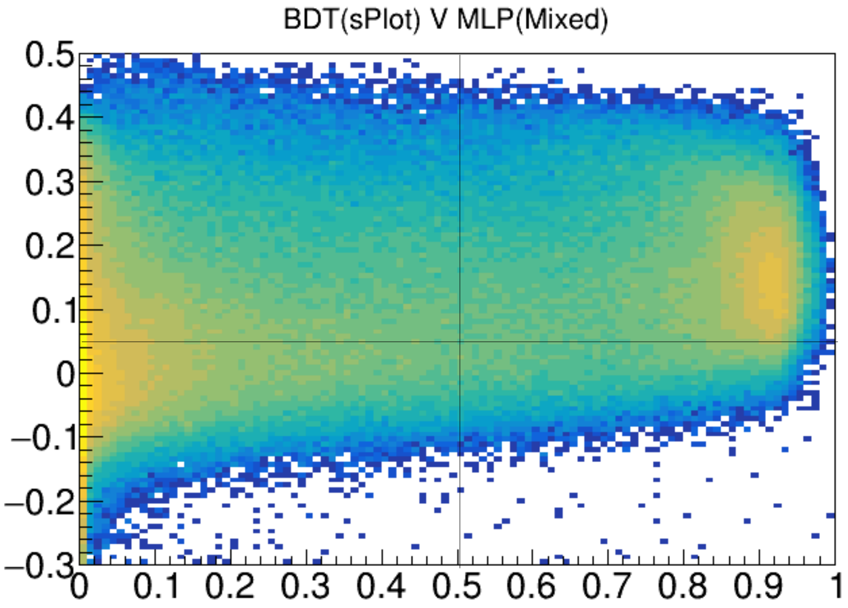


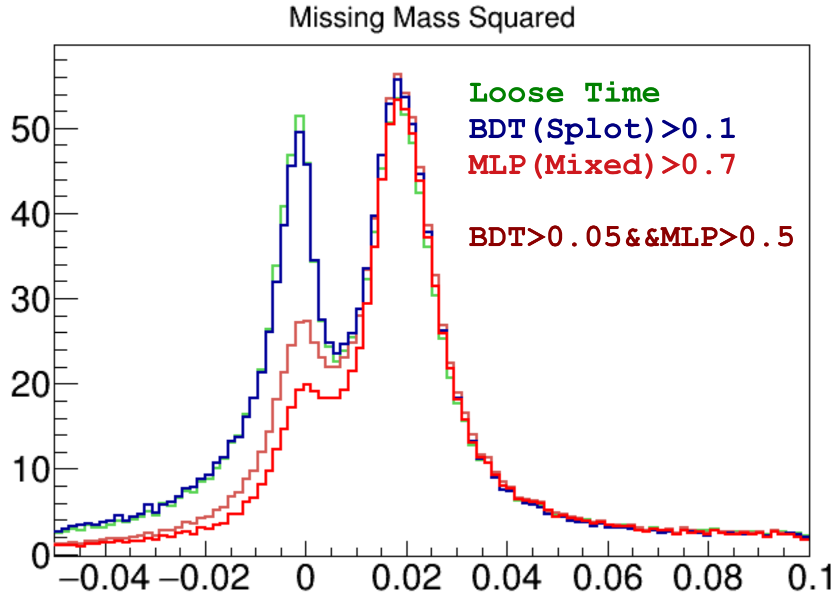


Judge from ω mass peak, we lose ~5% of events using previous cuts

Can regain these using a combination of classifiers with looser cuts

Double classifier cut





Further Reading

Boosted decision trees, sPlot training, ….

The Full Event Interpretation An Exclusive Tagging Algorithm for the Belle II Experiment Keck, T., Abudinén, F., Bernlochner, F.U. et al. Comput Softw Big Sci (2019) 3: 6. https://doi.org/10.1007/s41781-019-0021-8

T. Keck. Machine learning algorithms for the Belle II experiment and their validation on Belle data. PhD thesis, KIT, 2017. URL http://dx.doi.org/ 10.5445/IR/1000078149.

FastBDT: A Speed-Optimized Multivariate Classification Algorithm for the Belle II Experiment Keck, T. Comput Softw Big Sci (2017) 1: 2.

https://doi.org/10.1007/s41781-017-0002-8

Removing acceptance artifacts from BDT cuts

Stevens J, Williams M (2013) uboost: a boosting method for producing uniform selection efficiencies from multivariate classifiers. J Instrum 8(12). http://stacks.iop.org/1748-0221/8/i=12/a=P12013

Summary

TMVA can be used in place of standard particle ID cuts

Different methods can be used for creating signal and background samples
    Simulation, Mixed events, sPlots,...

Can benfit from multiple classifiers with different training conditions

Verification very important

Further work :

Train with simulation as signal and sPlot Background
    i.e. remove signal events when using data as background

Multiple Classifiers trained with Signal against different sources of backgrounds

- Wrong combitorial, random, physics,...
- Backgrounds from mixed events, sPlots, simulation,…

Timing Benchmarks

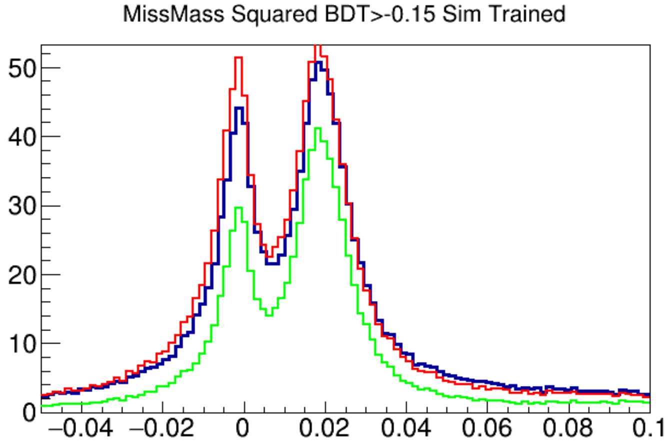
Table 1: Table of training CPU times in seconds.

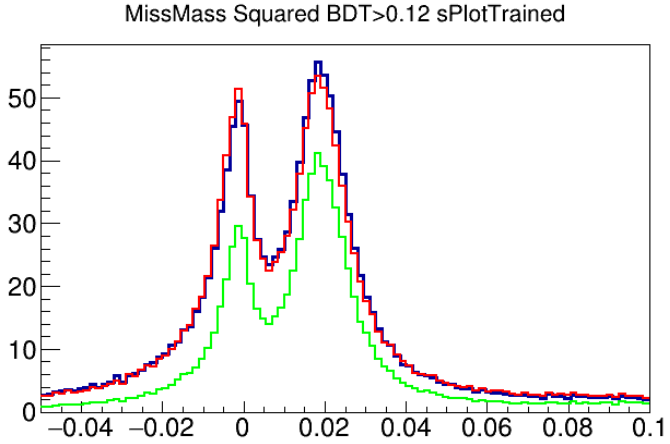
Method	12-vars
BDT	2.68
Cuts	2.41
CutsD	2.32
FDA_GA	2.40
KNN	1.93
LD	1.16
Likelihood	1.99
LikelihoodPCA	2.02
MLPBNN	19.54
RuleFit	5.33
SVM	2.23

Table 2: Table of evaluation CPU times in seconds.

Method	4-vars	8-vars	12-vars
BDT	29.57	29.33	28.51
Cuts	1.12	1.13	1.16
CutsD	1.12	1.15	1.16
FDA_GA	1.51	1.55	1.53
KNN	24.71	59.75	81.87
LD	1.38	1.47	1.59
Likelihood	2.08	2.40	2.90
LikelihoodPCA	2.92	4.25	6.32
MLPBNN	2.64	3.08	3.77
RuleFit	1.72	2.32	3.10
SVM	1.08	1.07	1.11

MLP

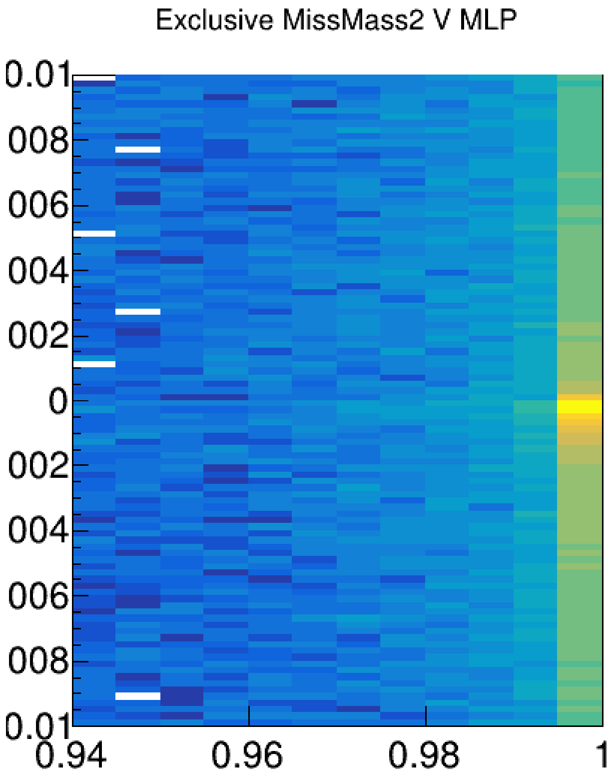




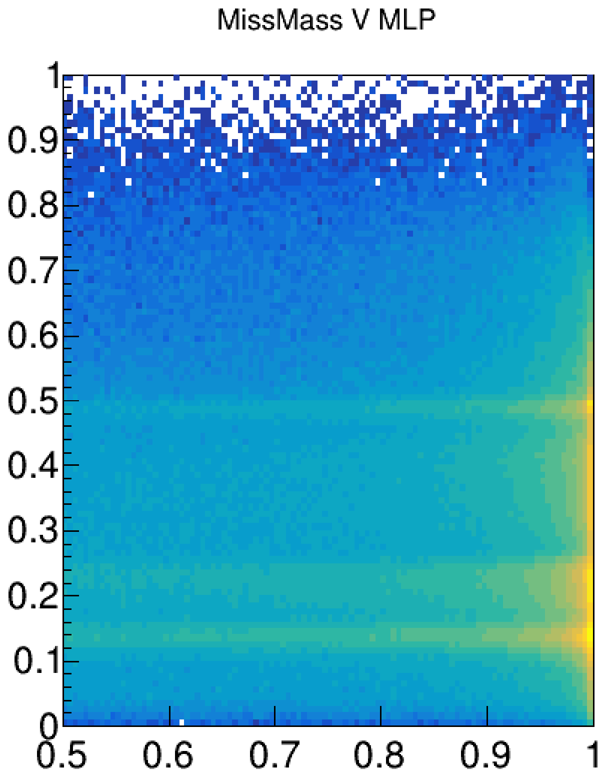
Sim trained BDT
Not quite as good as
Straight cuts
Less signal
Higher "other" background

sPlot trained
BDT only
At least as good
as straight cuts

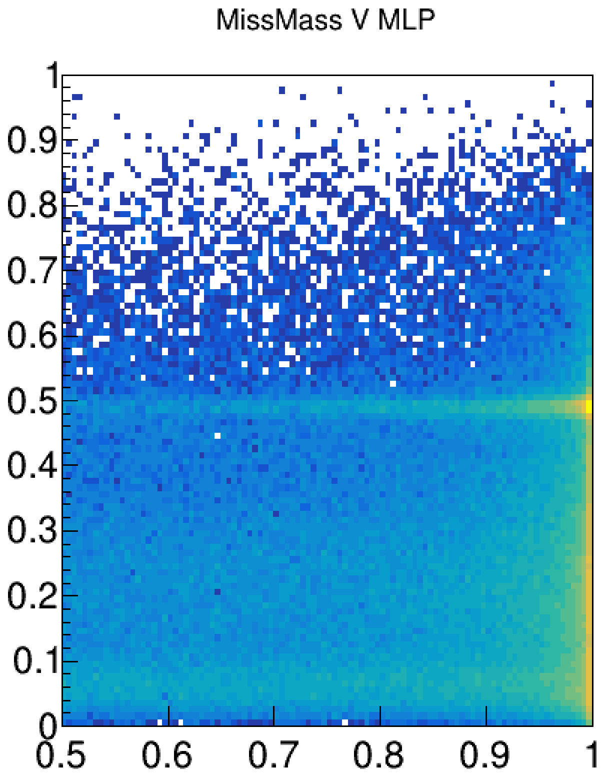
BDT → MLP



Missing K+



Missing K-



Optimised cuts based analysis