Lepton identification using TMVA methods

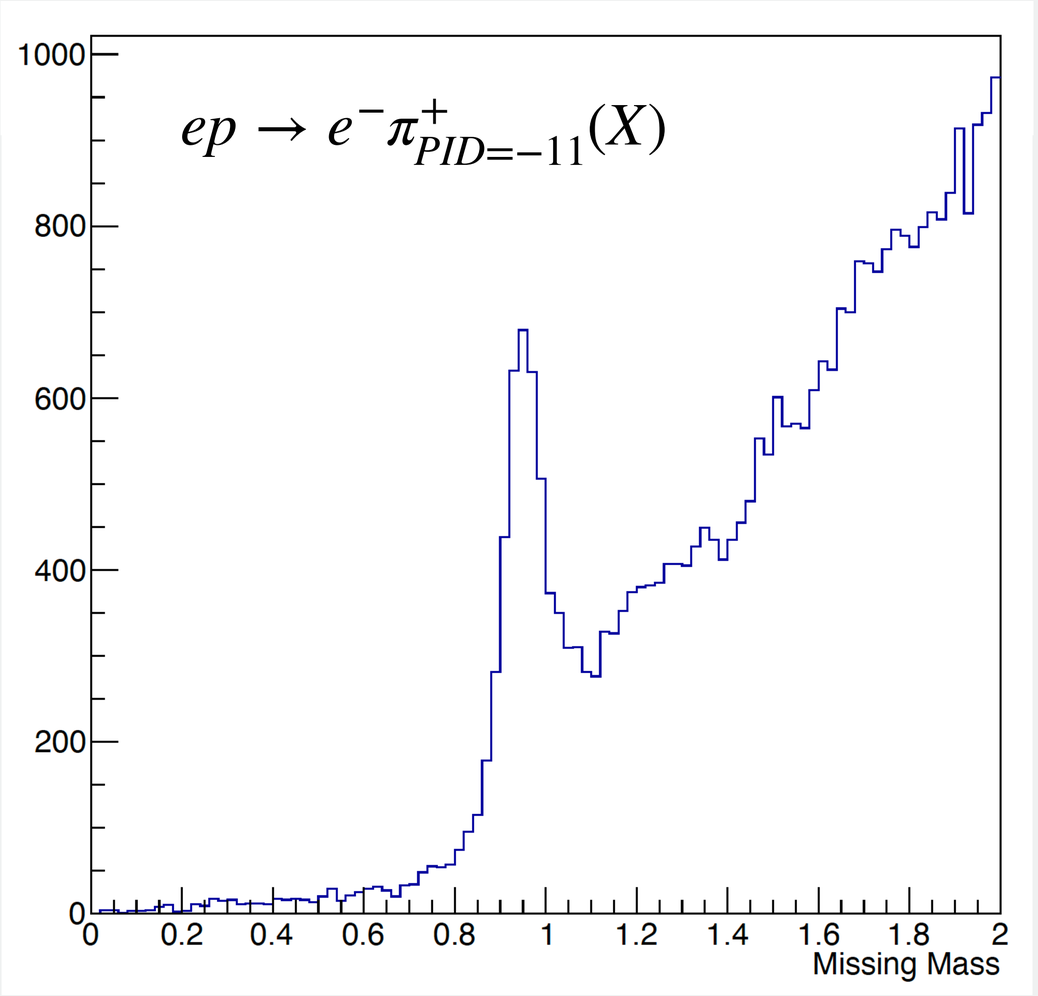
Mariana Tenorio Pita

CLAS Collaboration Meeting

March, 2024

Motivation

- Above the HTCC threshold (4.9 GeV), both pions and leptons produce a signal.
- Looking at the exclusive reaction $ep \rightarrow e^{-}e^{+}(X)$, where we assign the mass of the pion to the particle with PID=-11, we observe a clear peak in the neutron mass, due to the reaction $ep \rightarrow e^{-}\pi^{+}(n)$ where the π^{+} has been identified as a positron.
- This work has been done for Pass 1 using as variables the SF and m2 of PCAL, ECIN and ECOUT.



TMVA on ROOT

- The Toolkit for Multivariate Analysis is a ROOT integrated environment for the processing, evaluation and application of multivariate classification and regression techniques. All multivariate techniques in TMVA belong to the family of supervised learning algorithms.

- 6 Classifiers: e^+ and e^- identification on each Pass2 RGA configuration:
  - Fall 18 Inbending (10.6 Gev)
  - Fall 18 Outbending (10.6 GeV)
  - Spring 19 Inbending (10.2 GeV)

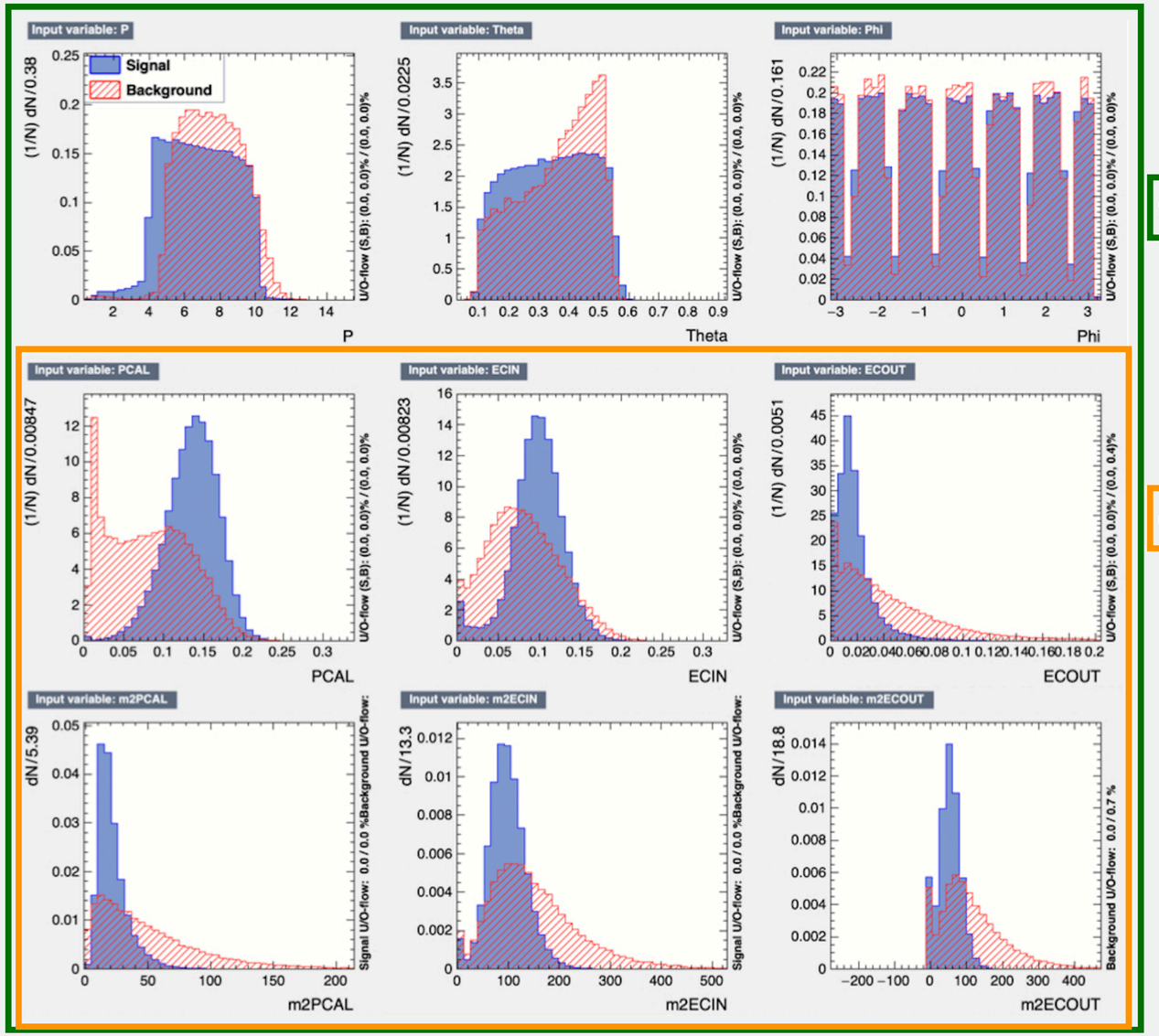
- Methods tested: Neural Networks and Boosted Decision Trees.

- Variables used: P, θ, ϕ, SF and m2 of PCAL, ECIN and ECOUT.

- 6 and 9 Variable models.

- Trained on simulation.

Training



9 variable model

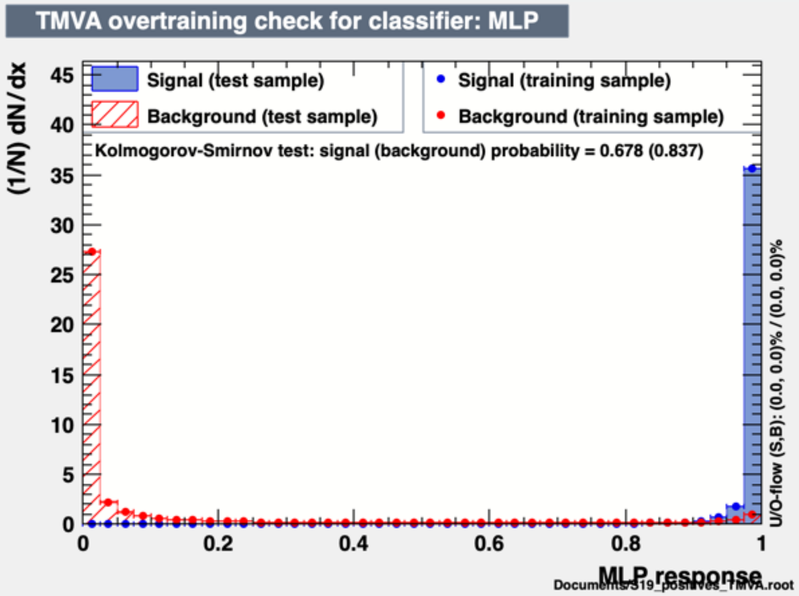
6 variable model

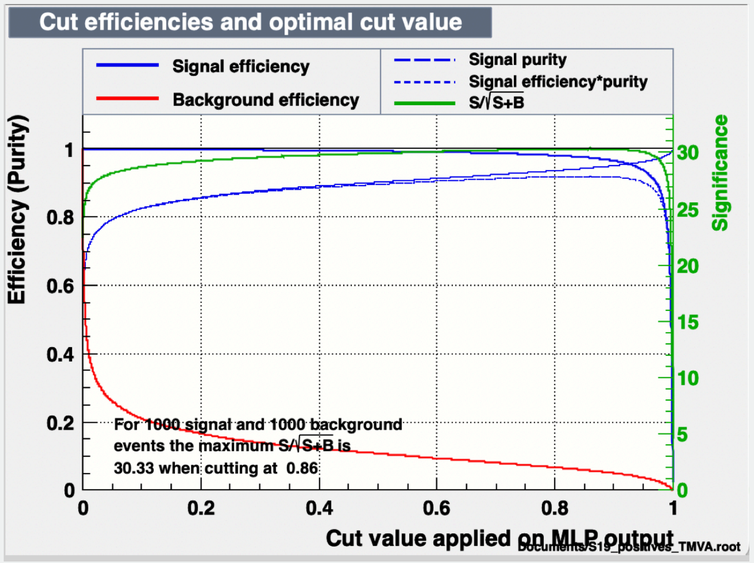
Blue: Signal e^+

Red: Background π^+

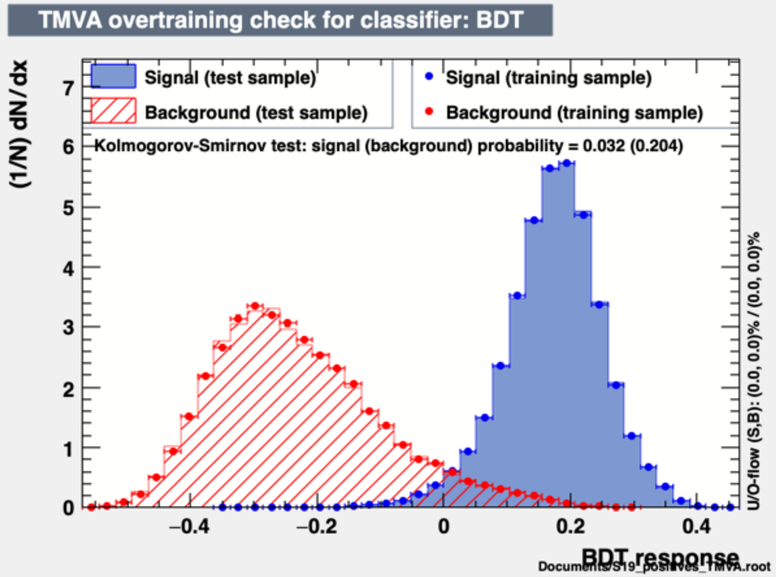
Training

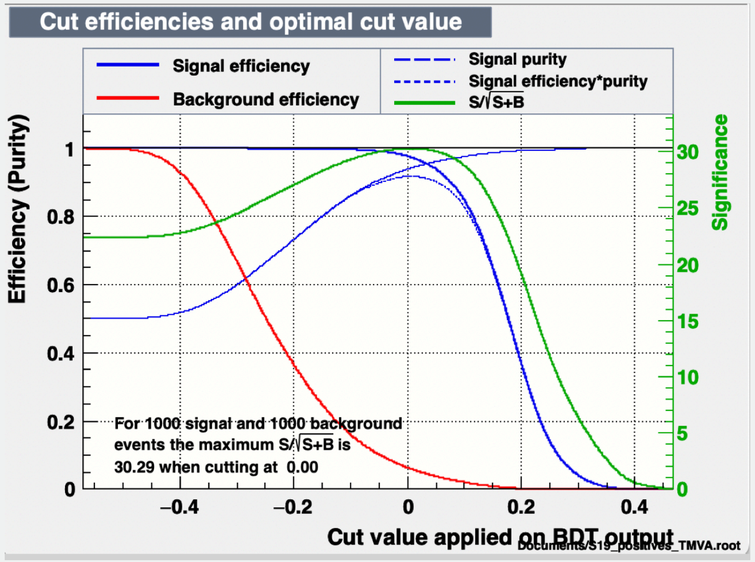
MLP



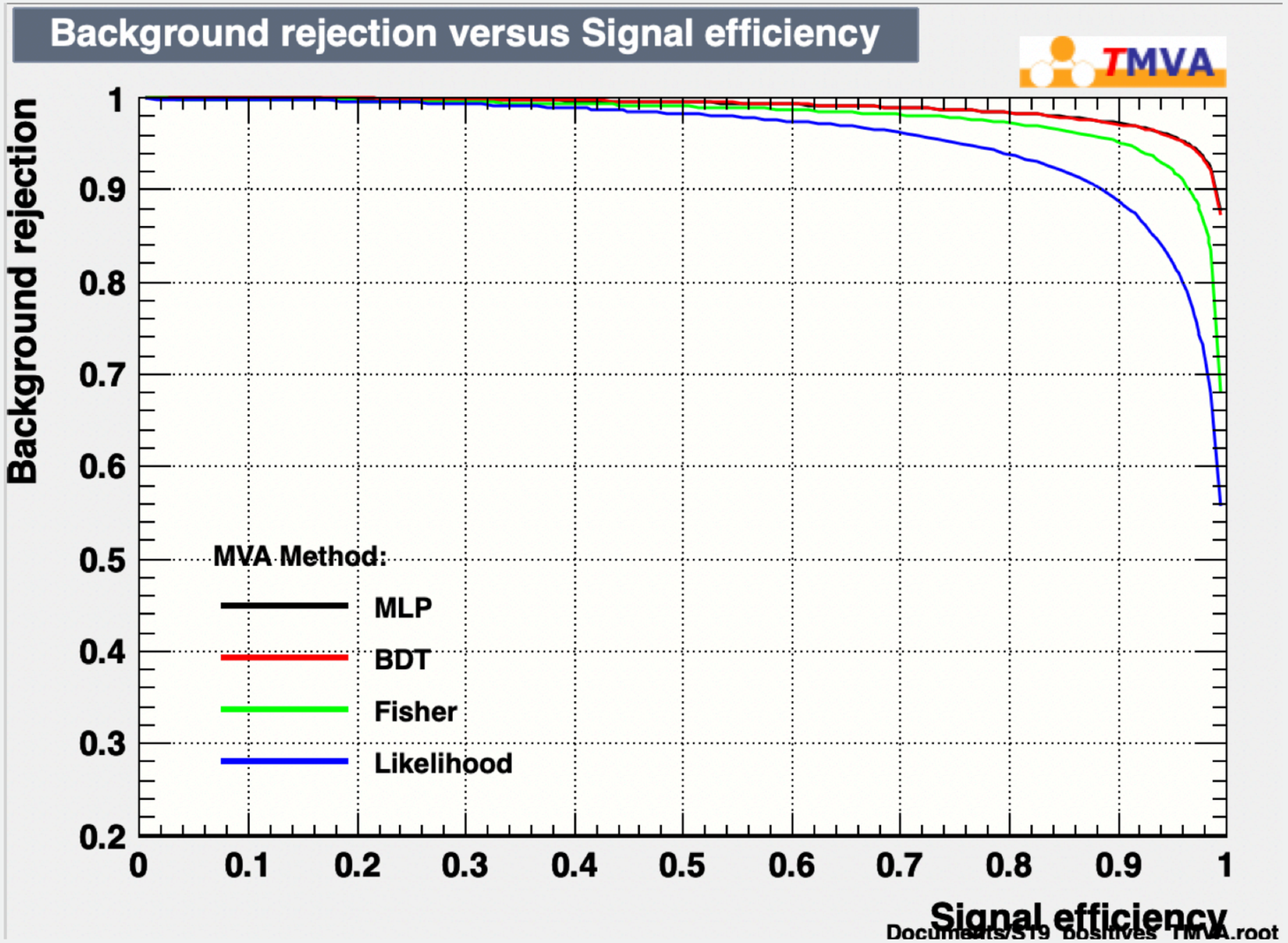


BDT





Training

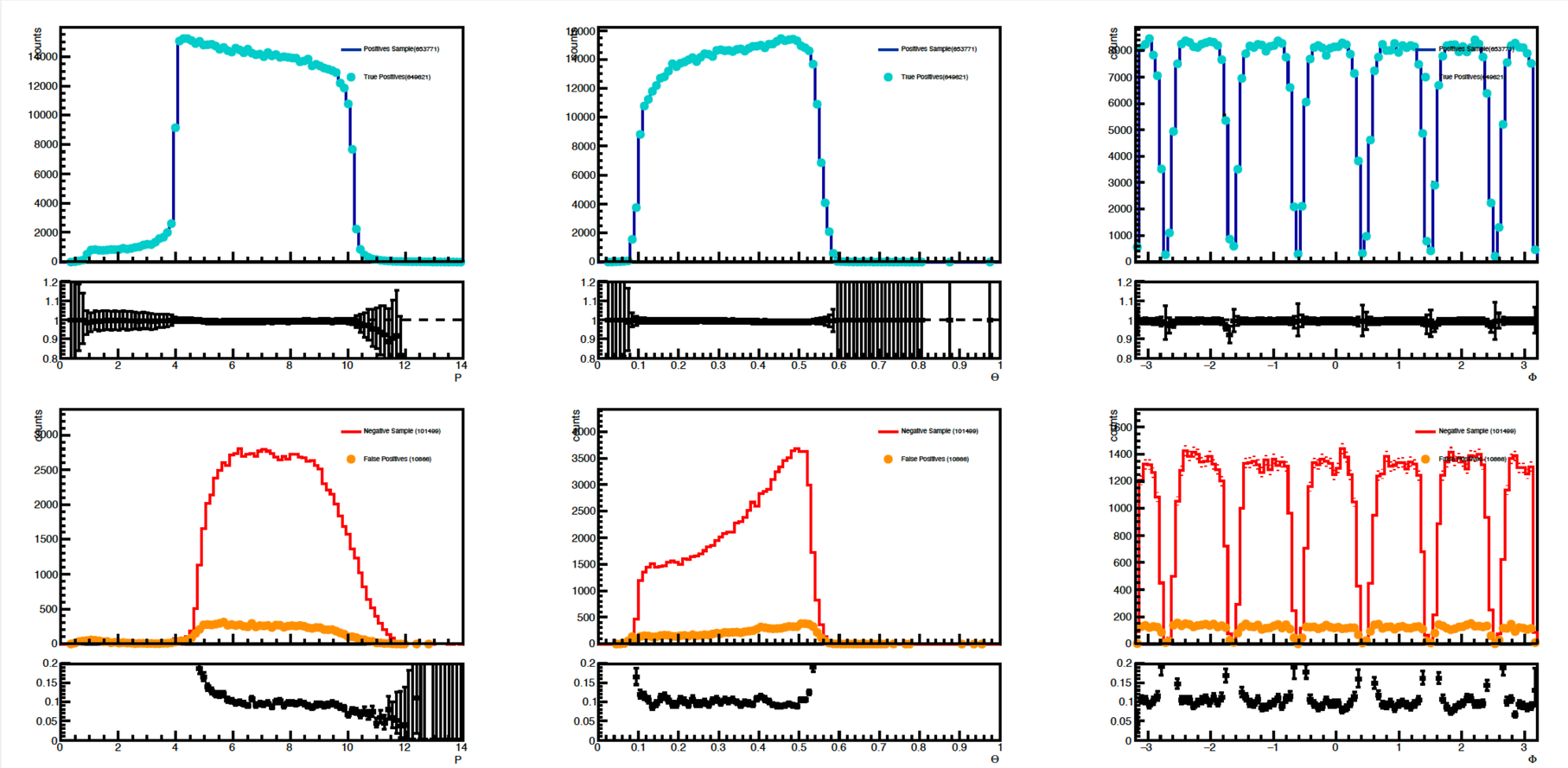

Background rejection versus Signal efficiency
TMVA
Background rejection
1
0.9
0.8
0.7
0.6
0.5
0.4
0.3
0.2
0
0.1
0.2
0.3
0.4
0.5
0.6
0.7
0.8
0.9
1
Signal efficiency
MVA Method:
MLP
BDT
Fisher
Likelihood
Documents/STS positives TMVA.root

Validation on Simulations

9 variable model

MLP	Actual e+	Actual pi+
	653,771	101,499
Predicted e+	649,244	10,158
Predicted pi+	4,527	91,341
	TPR: 99.30%	FPR: 10.10%

BDT	Actual e+	Actual pi+
	653,771	101,499
Predicted e+	649,621	10,866
Predicted pi+	4,150	90,633
	TPR: 99.36%	FPR: 10.70%



Signal Efficiency and Background reduction for P, θ and ϕ

Spring 2019

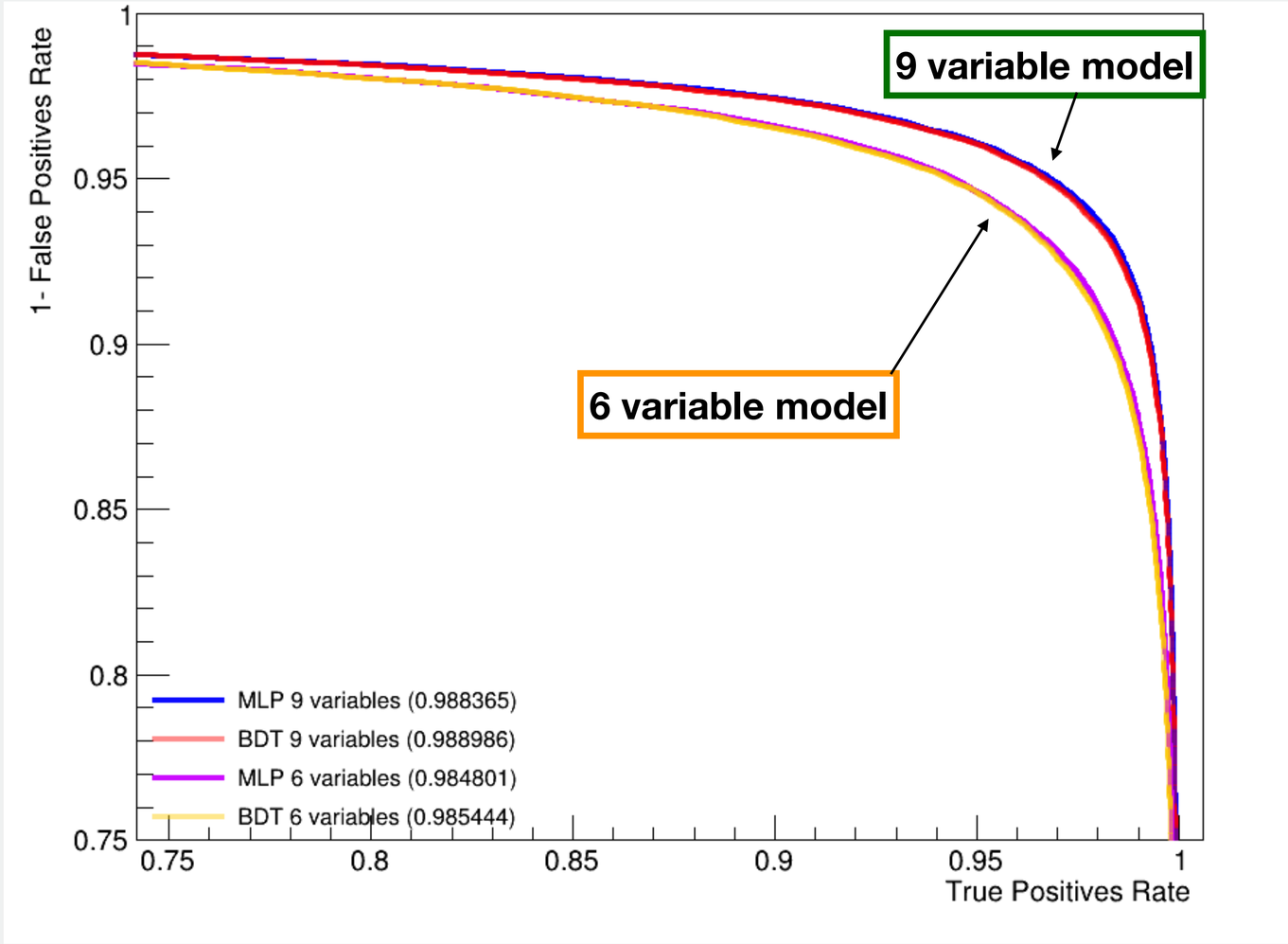
Signal: e^+

Background: π^+

cut=-0.06

Validation on Simulations

	True Positives Rate	False Positives Rate
MLP 9	99.30%	10.10%
BDT 9	99.36%	10.70%
MLP 6	99.06%	12.61%
BDT 6	99.02%	12.86%

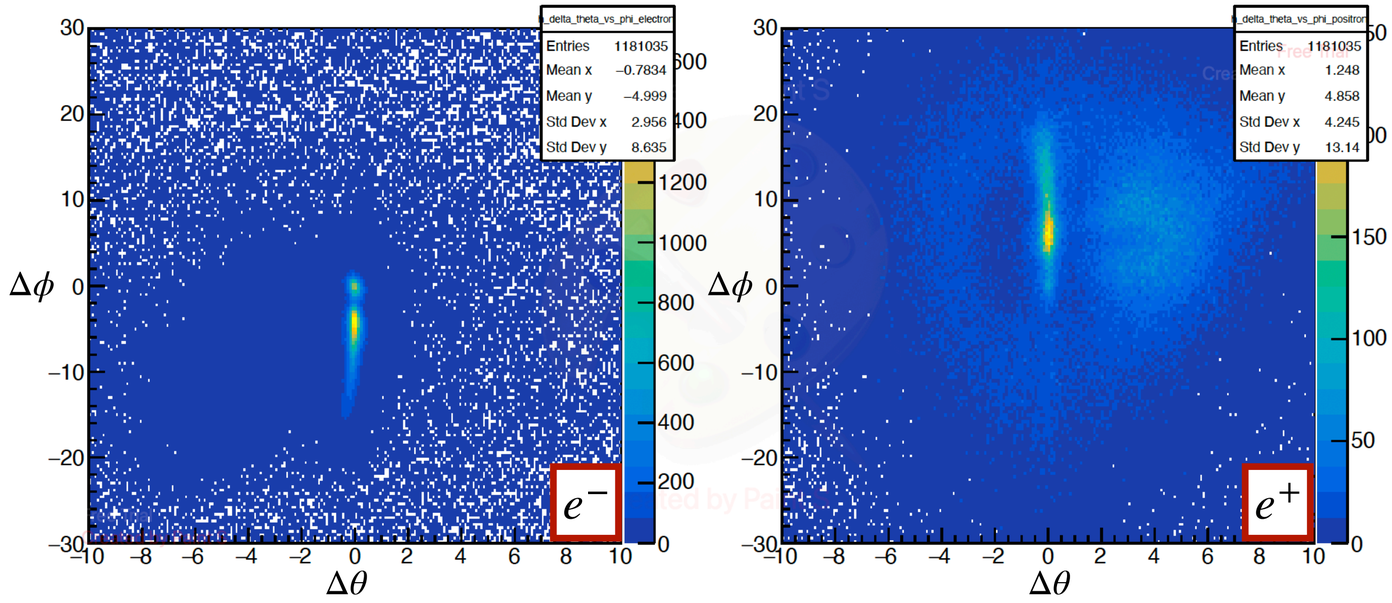


ROC Curve obtained from the validation of the models on simulations. Spring 2019

Signal: e^+ Background: π^+

Validation on Data-Signal Efficiency

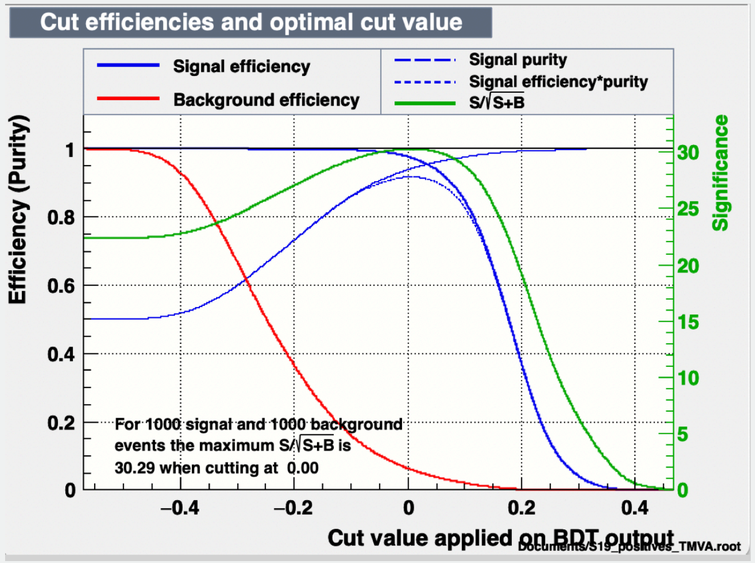
- For signal efficiency we look at the reaction $ep \rightarrow e\gamma$
- Leptons propagating from the target lose energy by radiating photons.
- We can identify as the radiating photons those where $\Delta\theta = \theta_\gamma - \theta_l \approx 0$
- By selecting events where $|\Delta\theta| \approx 0$ degrees we will assure that the particles selected are indeed leptons.

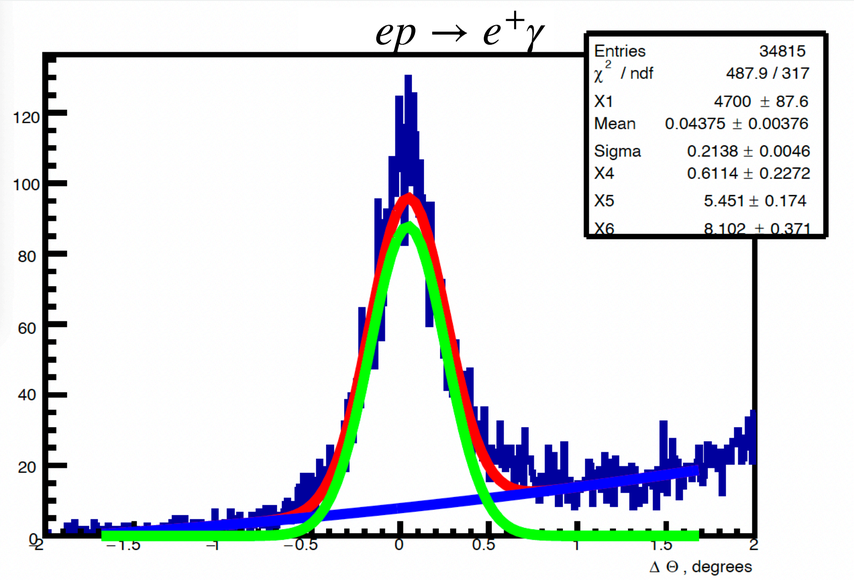


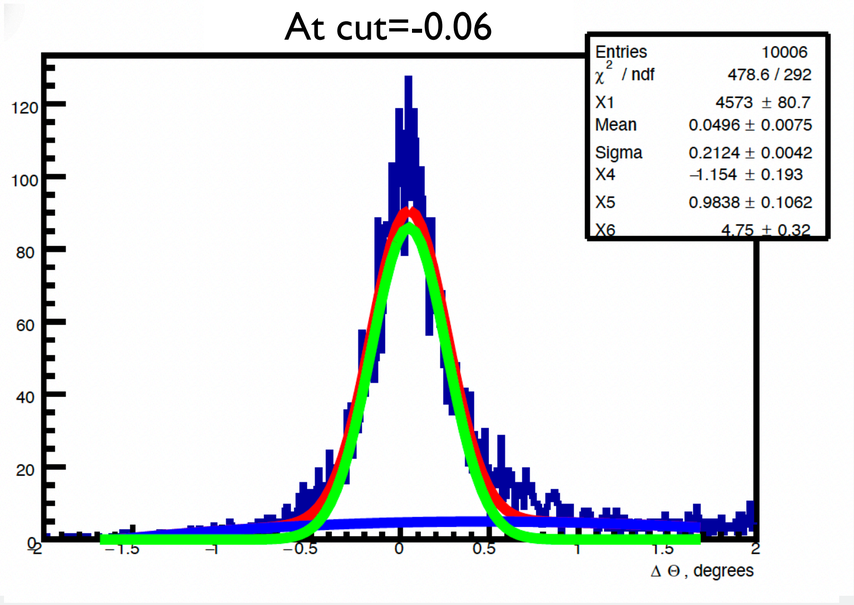
$\Delta\phi$ vs $\Delta\theta$ distributions for electrons (left) and positrons (right). Spring 2019 Pass2 data set

Validation on Data - Signal Efficiency e^+

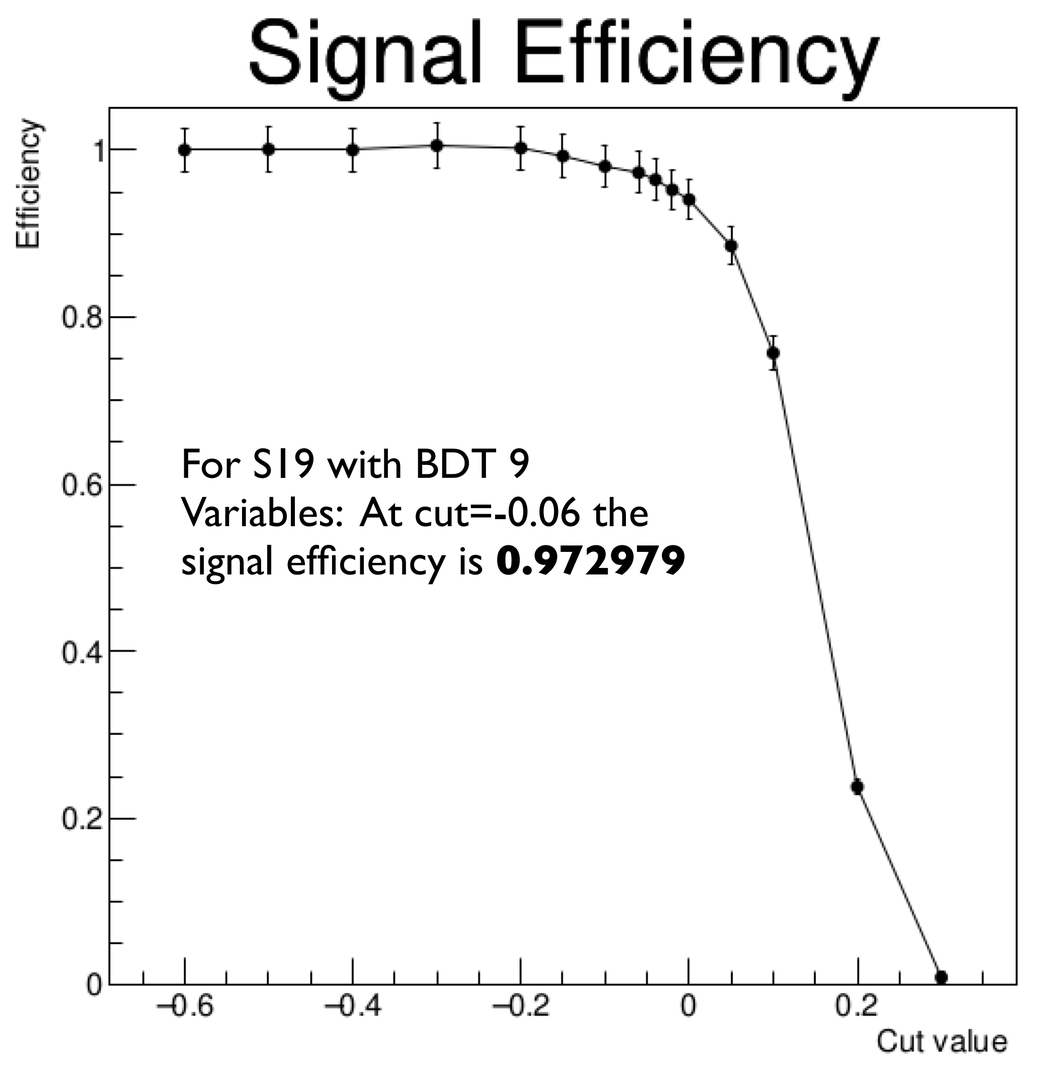
- We select $1\ e^+$ with $P > 4.5$ GeV and the associated photon.
- By selecting events where $|\Delta\theta| \approx 0$ degrees we will assure that the particles selected are indeed leptons.
- We apply different cuts for lepton ID from -0.6 to 0.4 for BDT to observe the effect on the signal.







Validation on Data - Signal Efficiency e^+



Signal Efficiency	Spring 2019
BDT 9	97.30%
BDT 6	97.98%

Signal Efficiency	Fall 2018 Inbending
BDT 9	99.75%
BDT 6	98.72%

Signal Efficiency	Fall 2018 Outbending
BDT 9	90.48%
BDT 6	89.29%

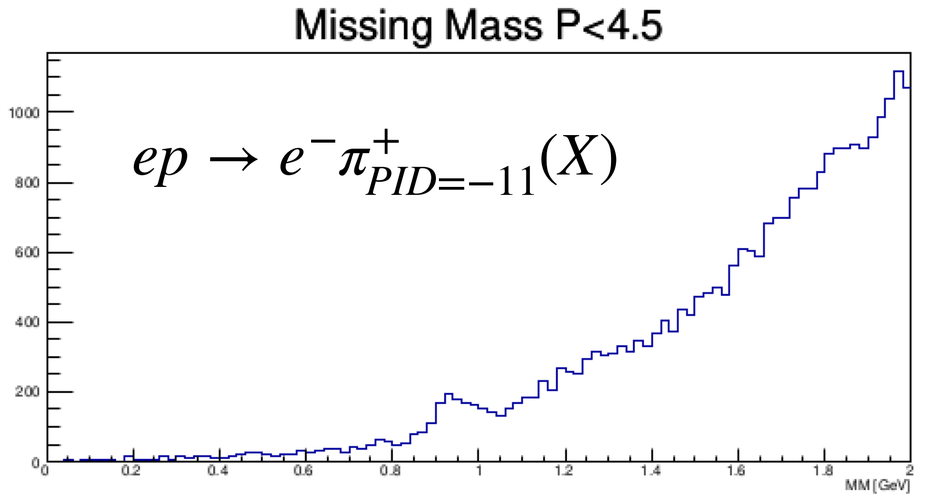
Validation on Data - Background Suppression

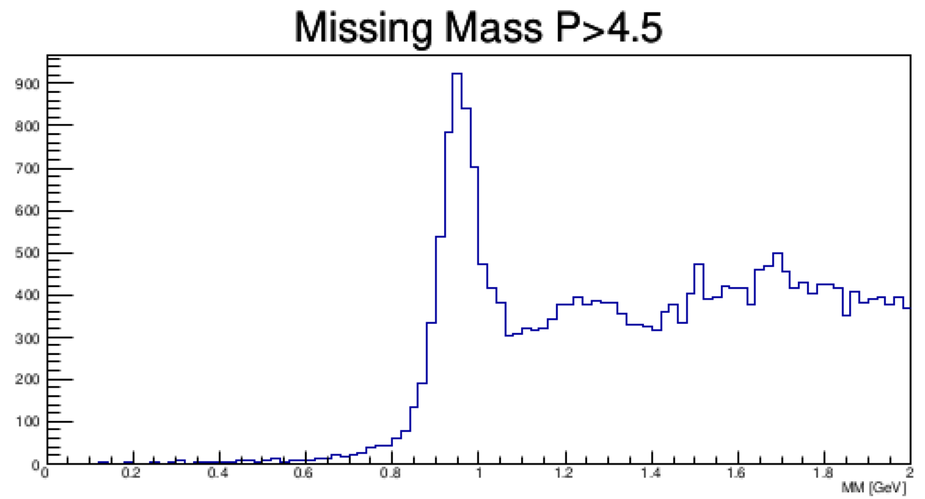
- For background suppression for positrons we can look at the reaction $ep \rightarrow e^{-}\pi^{+}(n)$

  → PID=-11

  Mass of pion

- We are looking at the background suppression at $P > 4.5$ GeV





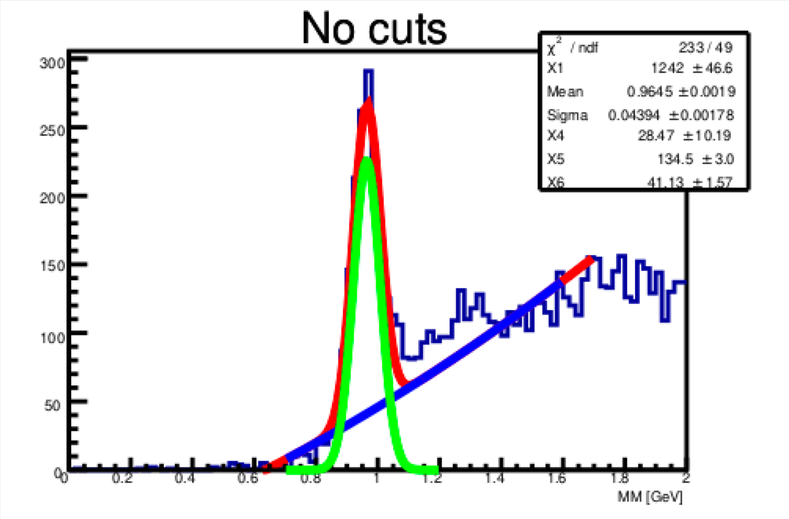
Validation on Data - Background Suppression

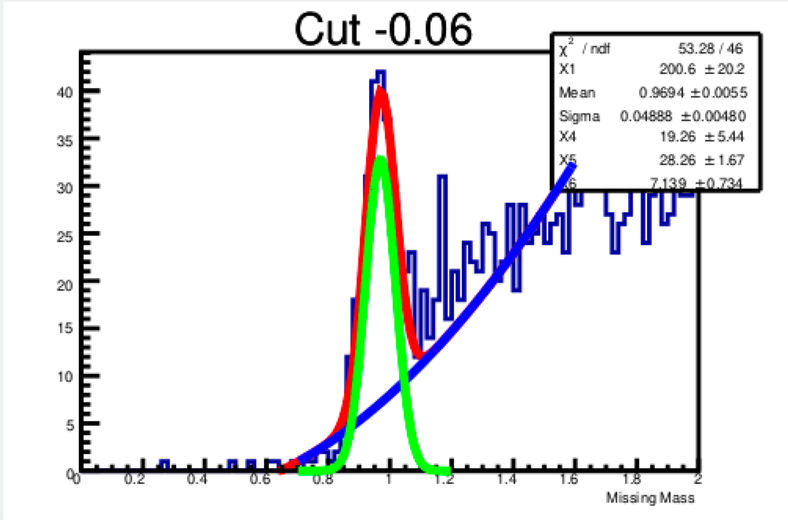
- For background suppression for positrons we can look at the reaction $ep \rightarrow e^{-}\pi^{+}(n)$

  → PID=-11

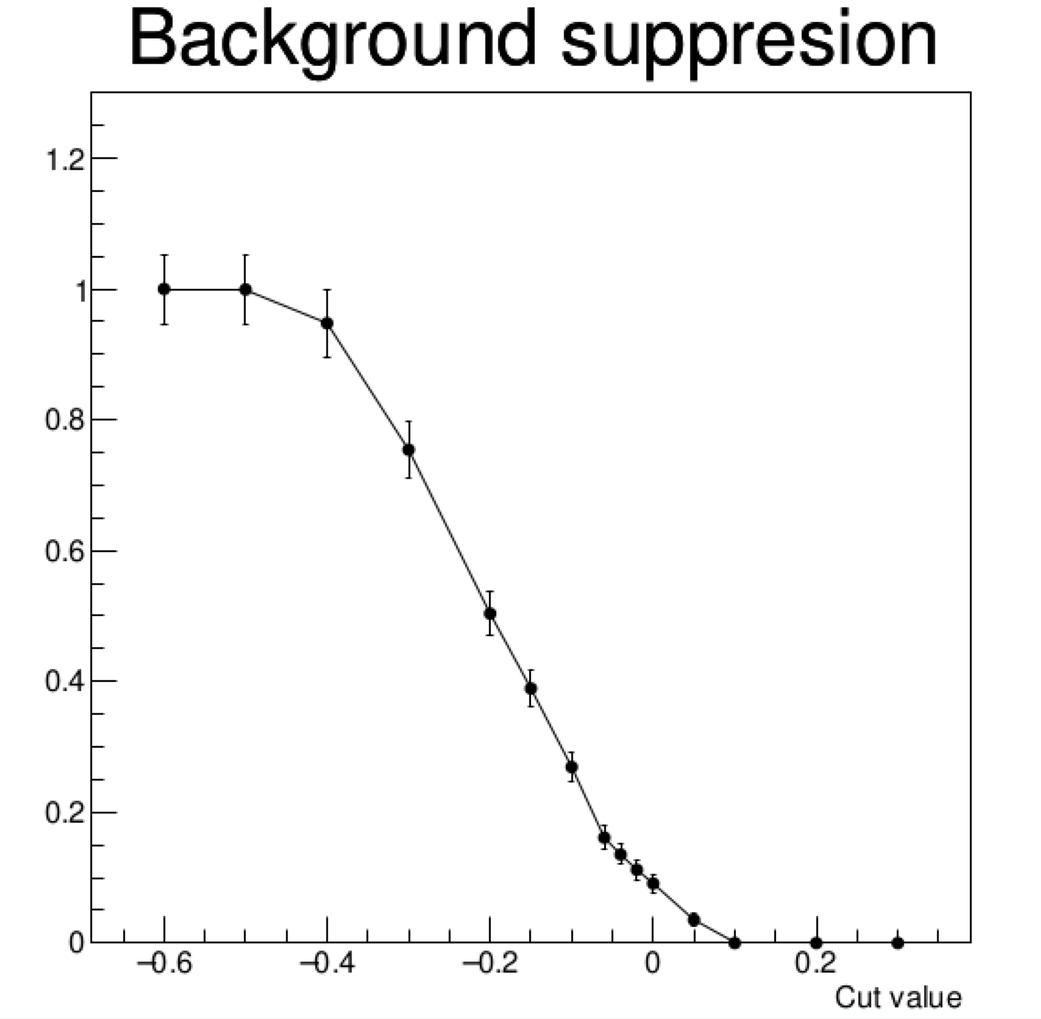
  Mass of pion

- We are looking at the background suppression at $P > 4.5$ GeV





Validation on Data - Background Suppression e^+

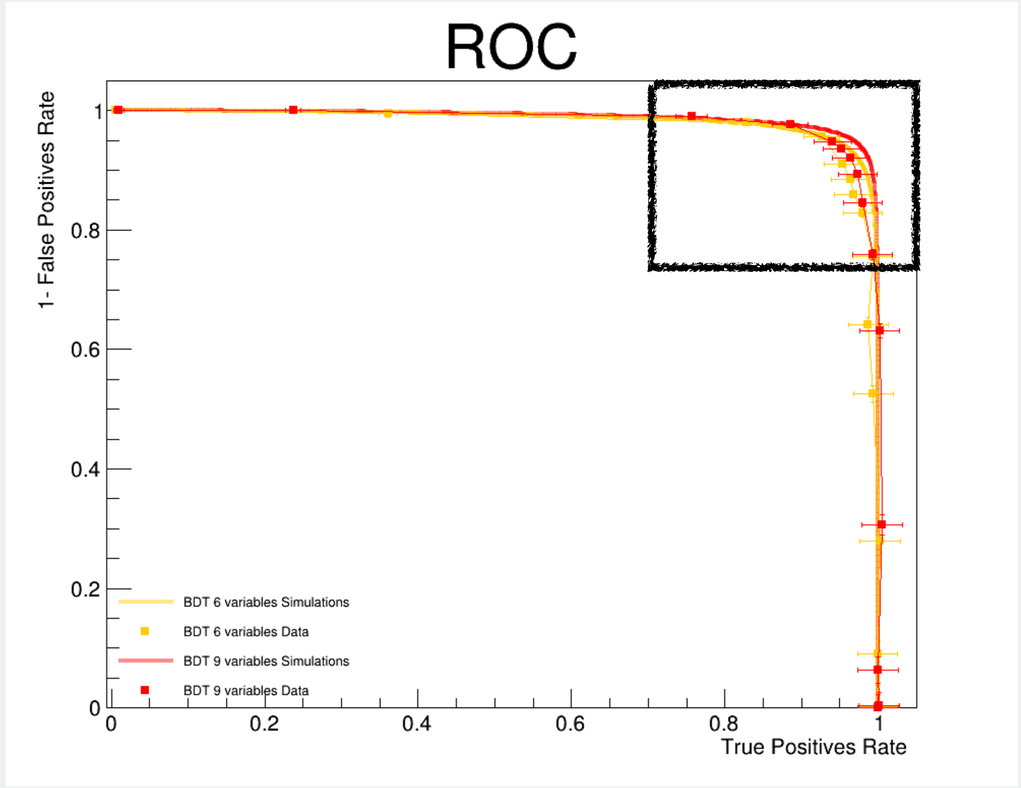


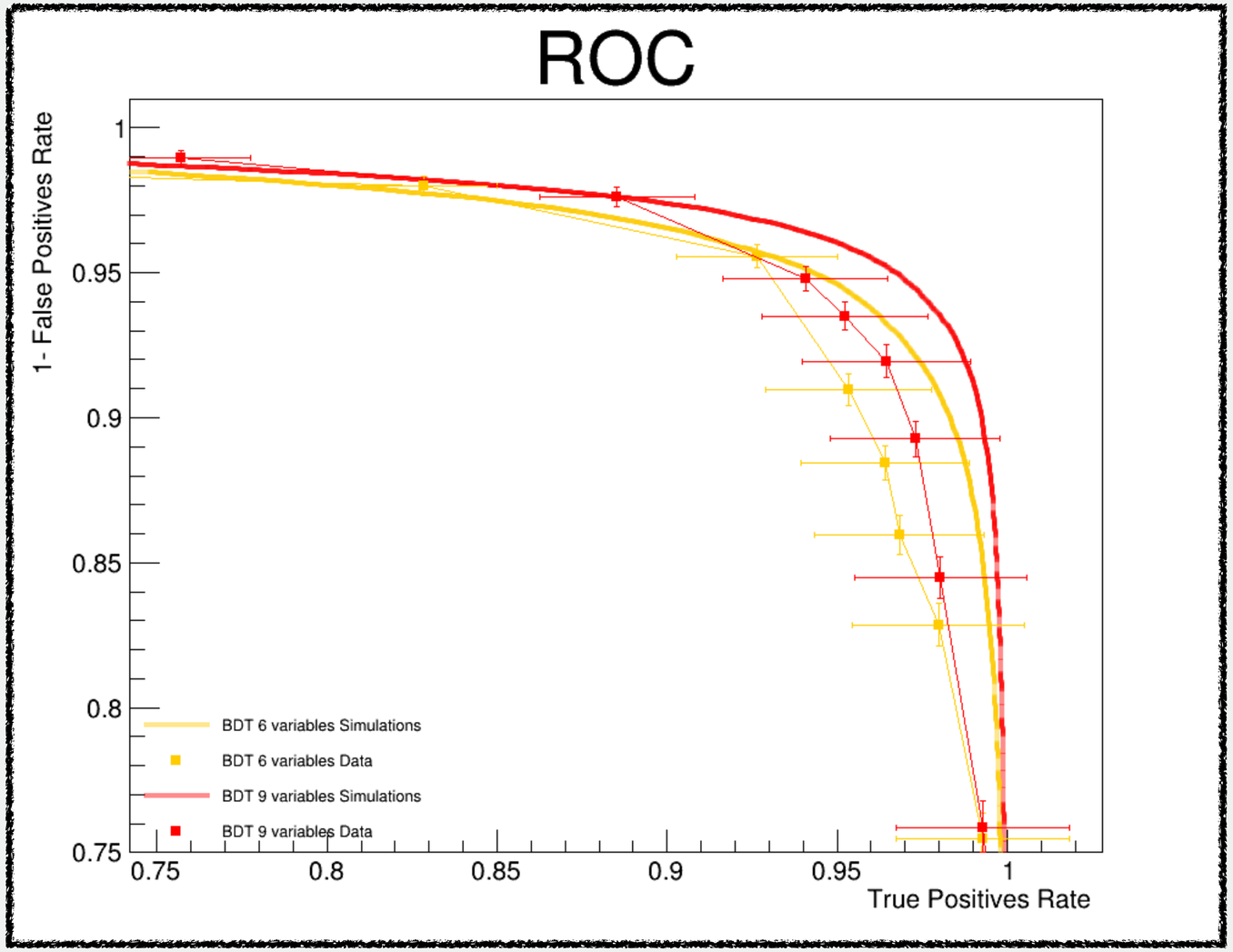
Background Suppression	Spring 2019
BDT 9	10.73%
BDT 6	17.15%

Background Suppression	Fall 2018 Inbending
BDT 9	16.12%
BDT 6	30.38%

Background Suppression	Fall 2018 Outbending
BDT 9	20.21%
BDT 6	32.66%

Validation on Data - ROC Curve





Conclusion

- The training and the validation was done for each configuration.
- The models seem to work as expected with over 90% of signal efficiency and background suppression at 10%.
- This work will be implemented in Iguana.

Signal Efficiency	Spring 2019
BDT 9	97.30%
BDT 6	97.98%

Background Suppression	Spring 2019
BDT 9	10.73%
BDT 6	17.15%

Data

	True Positives Rate	False Positives Rate
BDT 9	99.36%	10.70%
BDT 6	99.02%	12.86%

Simulations

Thank you!

Validation on Data - e^+

Spring 19

Signal Efficiency	**-0.06**	**0.0**	**0.05**
BDT 9	97.30%	94.06%	88.53%
BDT 6	97.98%	94.40%	92.64%

Fall 18 Inbending

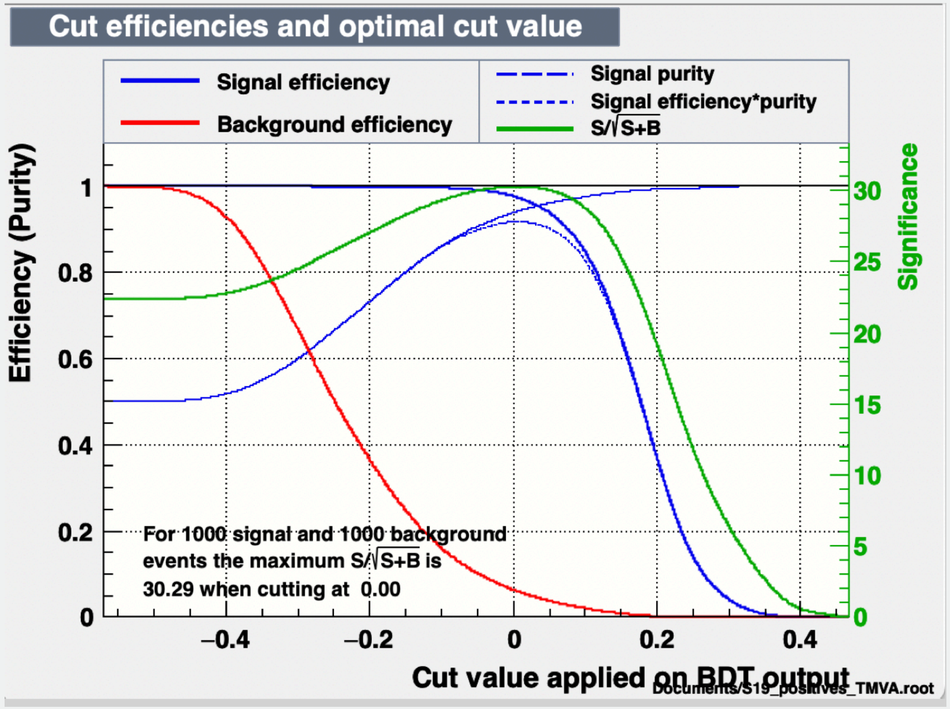
Signal Efficiency	**-0.06**	**0.0**	**0.05**
BDT 9	99.75%	97.85%	91.71%
BDT 6	98.72%	98.54%	95.31%

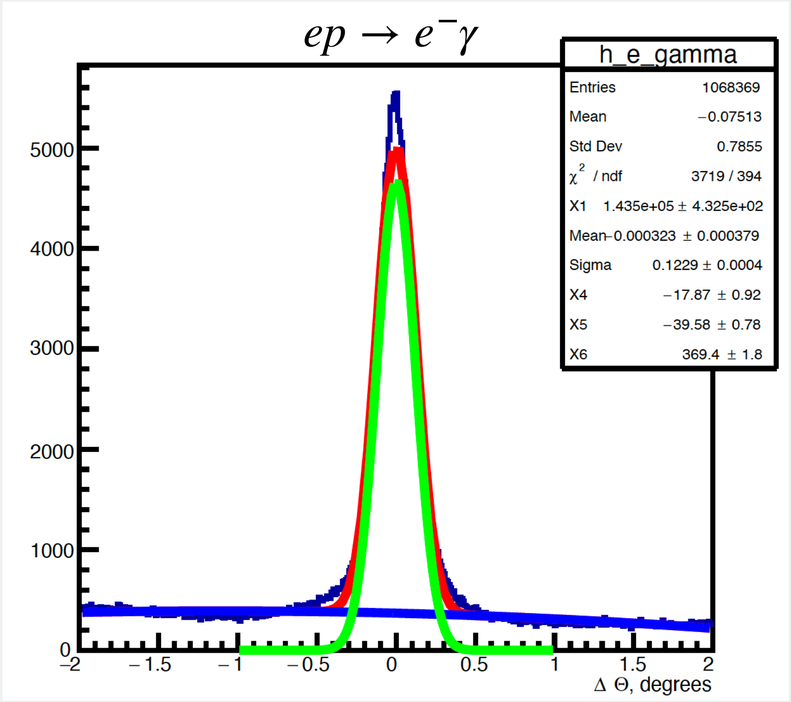
Fall 18 Outbending

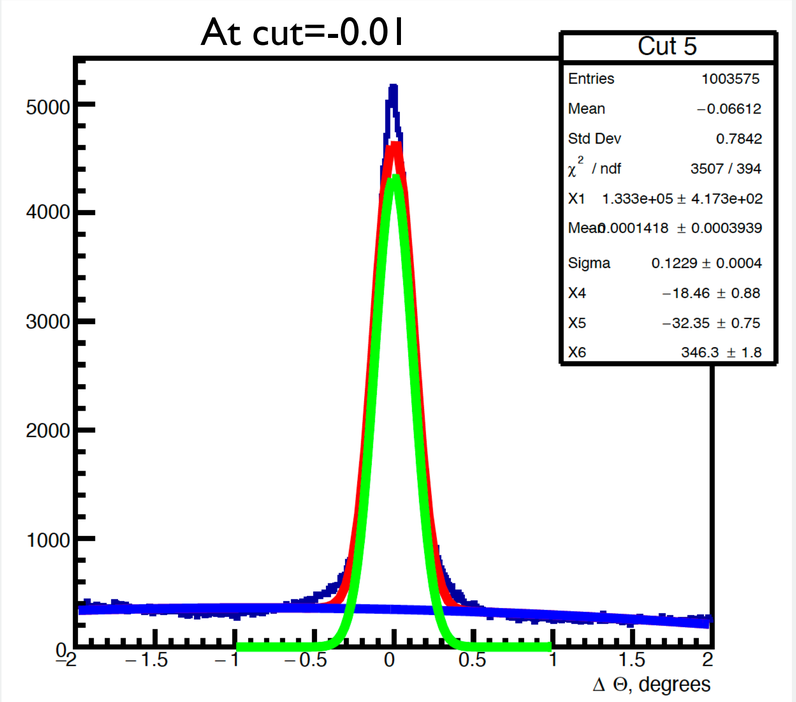
Signal Efficiency	**-0.06**	**0.0**	**0.05**
BDT 9	90.48%	90.48%	85.71%
BDT 6	89.29%	91.67%	89.29%

Validation on Data - Signal Efficiency e^-

- We select $1\ e^-$ with $P > 4.5$ GeV and the associated photon.
- We apply different a range of from -0.6 to 0.4 for BDT to observe the effect on the signal



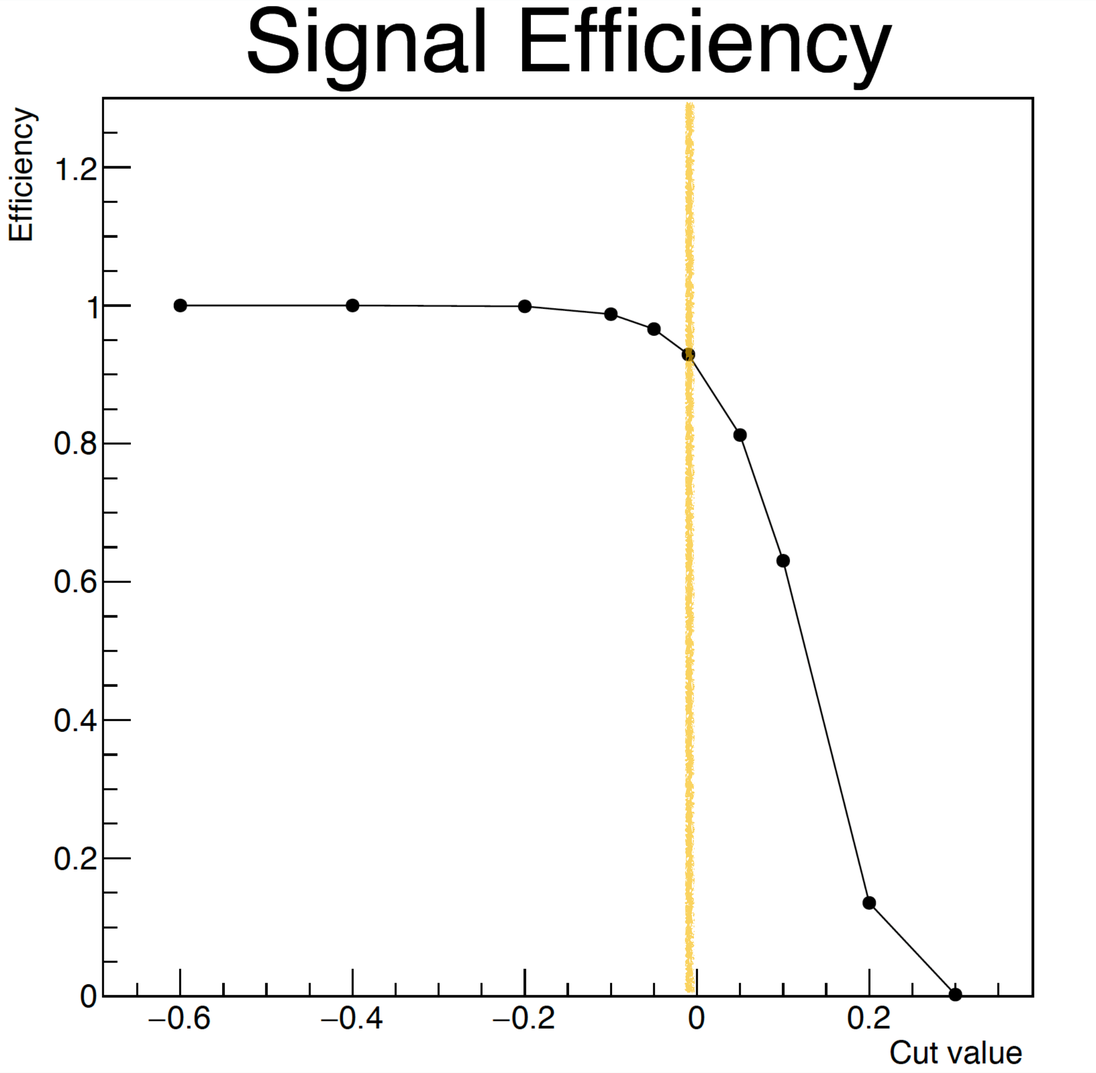




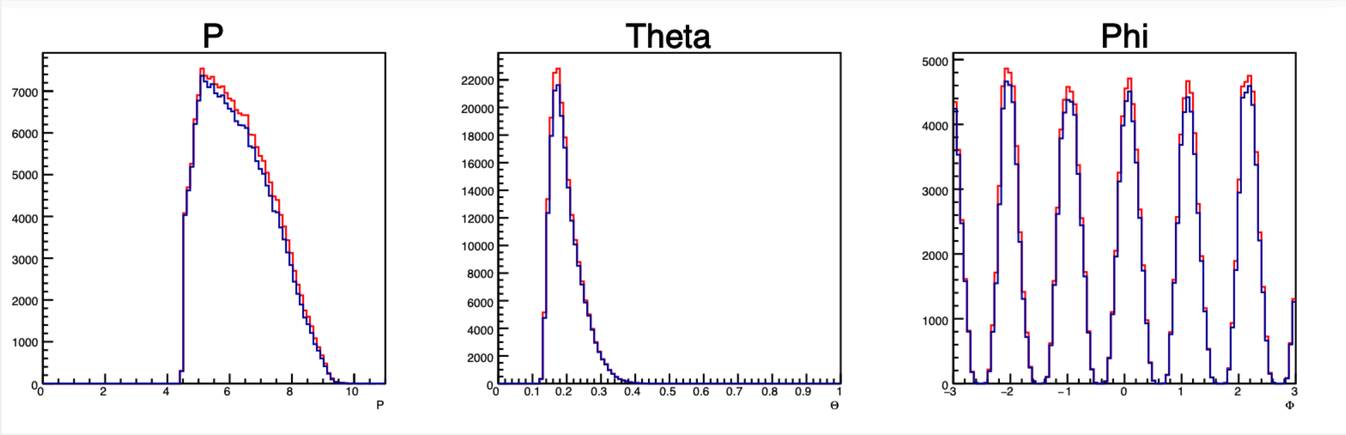
For S19 with BDT 9 Variables:

At -0.01 the signal efficiency is **0.9290**

Validation on Data - Signal Efficiency e^-

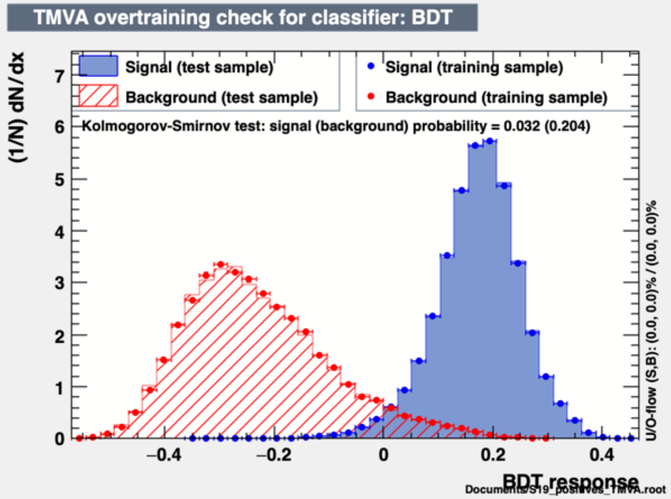


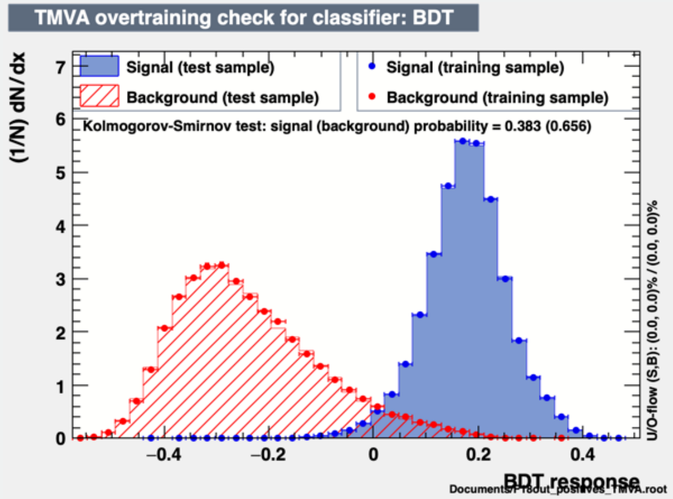
Signal Efficiency	Spring 2019	Fall 2018 Inbending	Fall 2018 Outbending
BDT 9	92.90%	94.08%	91.99%

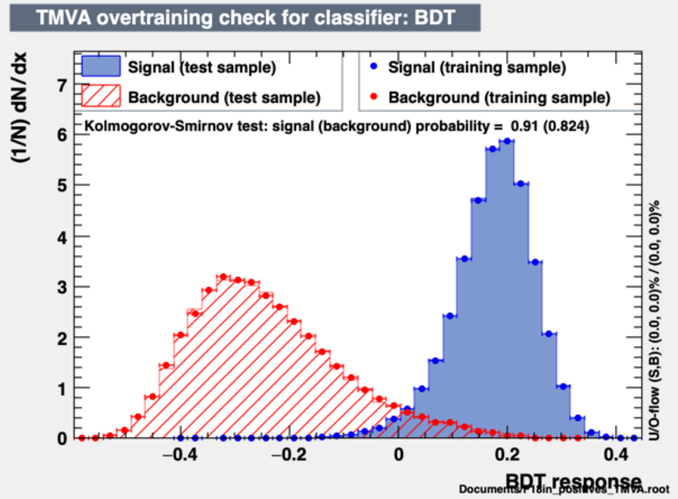


Training

9BDT







6BDT