Quarks

u	c	t	γ
d	s	b	g
ν_e	ν_μ	ν_τ	Z
e	μ	τ	W

Leptons

Ⓗ

Hadron $\begin{cases} \text{baryon : 3 quarks} \\ \text{meson : 2 quarks} \end{cases}$

$c = \hbar = 1$

克莱因-戈尔登方程 (Klein-Gordon)

$$i\frac{\partial\psi}{\partial t} = \hat{H}\psi \quad \text{and} \quad E^2 = \vec{p}^2 + m^2$$

$$\cancel{\left(i\frac{\partial\psi}{\partial t}\right)^2 =}$$

$$\left(i\frac{\partial}{\partial t}\right)^2\psi = \hat{H}^2\psi = (-\vec{\nabla}^2 + m^2)\psi$$

$$(\Box + m^2)\psi = 0 \quad , \quad \text{where } \Box = \frac{\partial^2}{\partial t^2} - \vec{\nabla}^2$$

Two problems: ① $E < 0$ ② $|\psi|^2 < 0$ 概率 = 负

$$\mathcal{L} = \frac{1}{2}\partial_\mu\phi\partial^\mu\phi - \frac{1}{2}m^2\phi^2$$

$$\partial_\mu\left(\frac{\partial L}{\partial(\partial_\mu\phi)}\right) - \frac{\partial L}{\partial\phi} = 0$$

$$(\partial_\mu\partial^\mu + m^2)\phi = 0$$

狄拉克方程 (Dirac)

$$(i\gamma^\mu\partial_\mu - m)\psi(x) = 0 \quad \left(\partial_\mu = \frac{\partial}{\partial x_\mu}\right)$$

$\Rightarrow$ Lorentz covariant Lagrangian

$\rightarrow \mathcal{L} = \bar{\psi}(i\gamma^\mu\partial_\mu - m)\psi$, where $\bar{\psi} \equiv \psi^\dagger\gamma^0$

for free Dirac field ψ

$$\psi = \begin{pmatrix} \psi_\uparrow \\ \psi_\downarrow \\ \psi^\dagger_\uparrow \\ \psi^\dagger_\downarrow \end{pmatrix} \quad \text{Dirac spinor}$$

由 扫描全能王 扫描创建

$\mathcal{L} = \bar{\psi}(i\gamma^\mu\partial_\mu - m)\psi$ 相位变换（规范变换）（类比理力中的平移、旋转对称性）

① ~~改变相位~~：$\psi \to e^{-iw}\psi$，$\bar{\psi} \to e^{iw}\bar{\psi}$，$w$ 为常数与 x 无关

此时拉氏量 $\mathcal{L}$ 不改变，满足规范不变性

$e^{\pm iw}$ 构成一个可交换群 abelian。$e^{-iw_1}e^{-iw_2} = e^{-iw_2}e^{-iw_1}$

称为 $U(1)$，一阶酉群，一阶酉矩阵构成的群 $U^\dagger = U^{-1}$

$\mathcal{L}$ 具有全局 $U(1)$ 规范不变性 invariant under global $U(1)$ transformation

② 局域 $U(1)$ 变换

$\psi \to e^{-iw(x)}\psi$，$\bar{\psi} \to e^{iw(x)}\bar{\psi}$，$w(x)$ 与 x 有关

取 $w(x)$ 极小，有 $\psi \to \psi + \delta\psi = \psi - iw(x)\psi$ （或反括代入）

$\bar{\psi} \to \bar{\psi} + \delta\bar{\psi} = \bar{\psi} + iw(x)\bar{\psi}$

此时 $\mathcal{L}$ 不满足规范不变性

$$\mathcal{L} \to \mathcal{L} + \delta\mathcal{L} = \mathcal{L} + \bar{\psi}\gamma^\mu[\partial_\mu w(x)]\psi$$

引入向量场 vector field $A_\mu(x)$

相互作用 $-e\bar{\psi}\gamma^\mu A_\mu\psi$

新的拉氏量为 $\mathcal{L} = \bar{\psi}(i\gamma^\mu(\partial_\mu + ieA_\mu) - m)\psi$，~~满足规范不变性~~

令 $A_\mu \to A_\mu + \delta A_\mu = A_\mu + \frac{1}{e}[\partial_\mu w(x)]$，$\mathcal{L}$ 满足规范不变性

类比：
$$\begin{cases} \nabla\cdot\vec{B} = 0 \\ \nabla\times\vec{E} = -\frac{\partial\vec{B}}{\partial t} \end{cases} \Rightarrow \begin{cases} \vec{B} = \vec{\nabla}\times\vec{A} \\ \nabla\times\vec{E} = -\frac{\partial}{\partial t}(\nabla\times\vec{A}) \end{cases}, \quad \nabla(\vec{E} + \frac{\partial\vec{A}}{\partial t}) = 0$$

$$\Rightarrow \vec{E} + \frac{\partial\vec{A}}{\partial t} = -\nabla V$$

$\vec{A}' = \vec{A} + \nabla\lambda$

$\Rightarrow$ [illegible]

由 扫描全能王 扫描创建

加上场与场的耦合项　　vector field　$\mathcal{L} = -\frac{1}{4}F^{\mu\nu}F_{\mu\nu} + \frac{1}{2}m^2A^\mu A_\mu$ 各个分量满足KG方程

$$\mathcal{L} = -\frac{1}{4}F_{\mu\nu}F^{\mu\nu} + \bar{\psi}(i\gamma^\mu(\partial_\mu + ieA_\mu) - m)\psi,$$

其中 $F_{\mu\nu} = \partial_\mu A_\nu - \partial_\nu A_\mu$

质量项不满足不变性，$\delta\mathcal{L} = \frac{2m^2}{e}A^\mu\partial_\mu w(x) \neq 0$.

$\boxed{\text{vector boson must be massless!}}$

引入 covariant derivative D_μ, as $D_\mu = \partial_\mu + ieA_\mu$

so $\mathcal{L} = -\frac{1}{4}F_{\mu\nu}F^{\mu\nu} + \bar{\psi}(i\gamma^\mu D_\mu - m)\psi$　(QED)

推广到非阿贝尔群：　(杨-米斯 Yang-Mills theory)

$SU(n)$，需要 n^2-1 个实参数

$e^{-iw^aT^a}$　w^a $a \in \{1, \dots, n^2-1\}$，$T^a$ 称之 generator，

↓ traceless，n 阶 Hermitian 厄米矩阵

这时 $e^{-iw_1^aT^a}e^{-iw_2^aT^a} \neq e^{-iw_2^aT^a}e^{-iw_1^aT^a}$

因为 $[T^a, T^b] \neq 0$

$\psi \to (e^{-iw^aT^a})\psi$

为了保持规范不变性，引入 n^2-1 个新 vector particle A_μ^a.

取 $n=3 \longrightarrow$ QCD，8个胶子 (n^2-1)

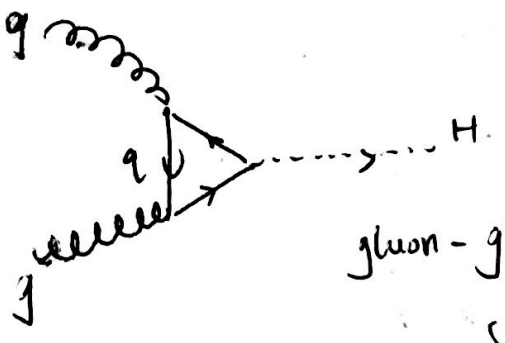
$$\mathcal{L} = -\frac{1}{4}F_{\mu\nu}F^{\mu\nu}$$

$+ i\bar{\psi}\not{D}\psi + h.c.$ $\longrightarrow$ 规范粒子与 Higgs 场 ϕ 耦合　(27 free parameters)

$+ \psi_i y_{ij} \psi_j \phi + h.c.$ $\longrightarrow$ fermion 粒子与 Higgs 耦合 (电弱)　Higgs 质量

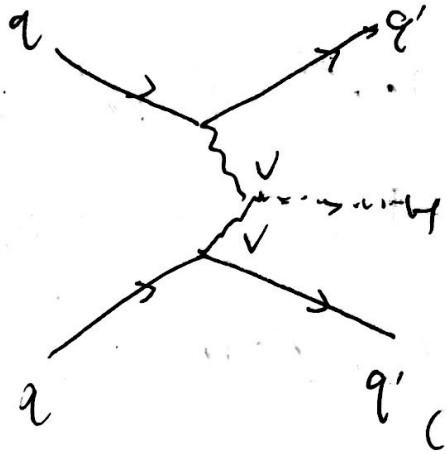
由 扫描全能王 扫描创建

两种产生过程



gluon-gluon-fusion

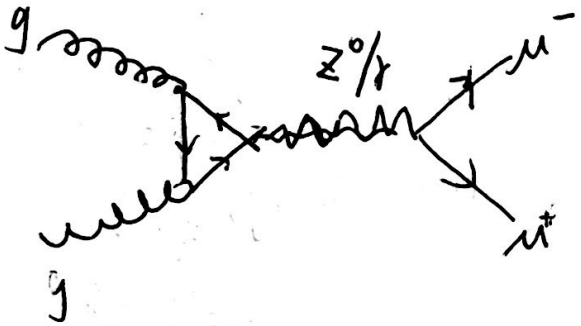
(ggF)



(vector boson fu

VBF

Drell-Yan (DY)



由 扫描全能王 扫描创建

Select ggF events with MVA method in $H \to \mu\mu$ Analysis

Haifeng Li, Yidi Qi, Dmitri Tsybychev

Stony Brook University

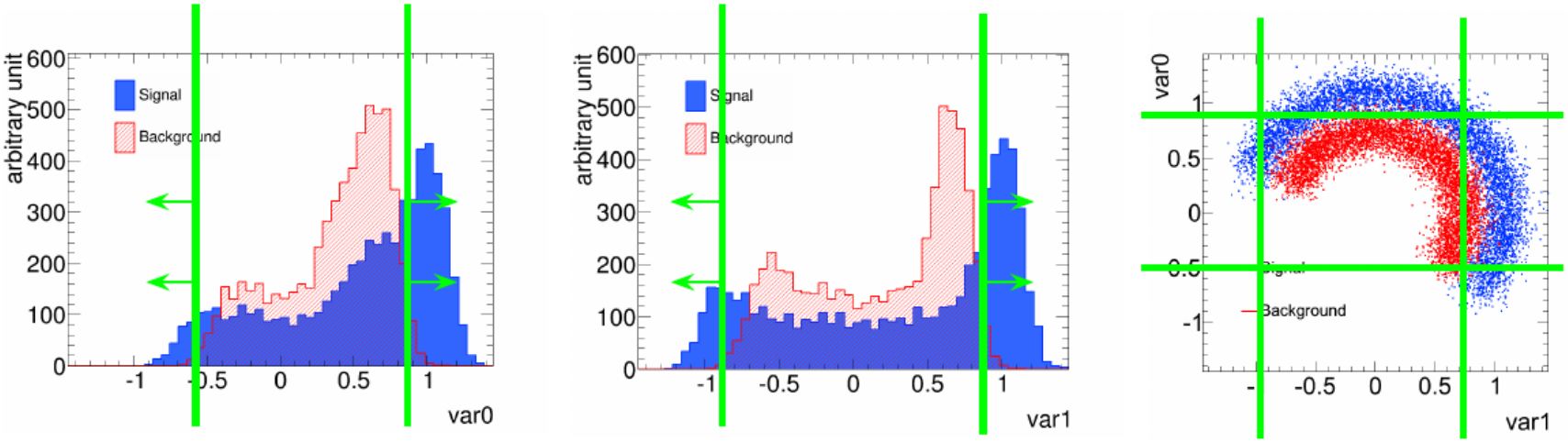
August 16, 2017

Introduction

- MVA is a powerful tool to discriminate signal and background
- Currently, in our analysis, cut based method is used in ggF region for categorization
- Try MVA method instead and compare the performence

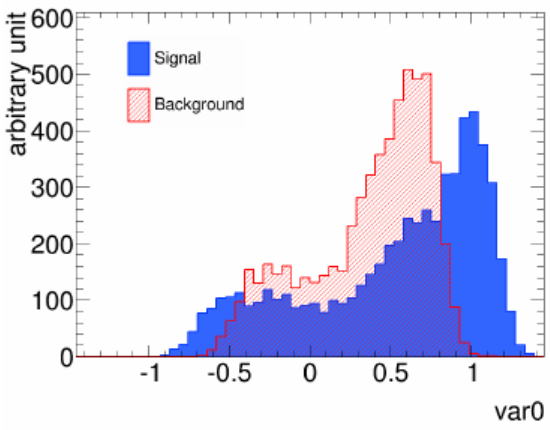
Introduction to Multivariable Analysis

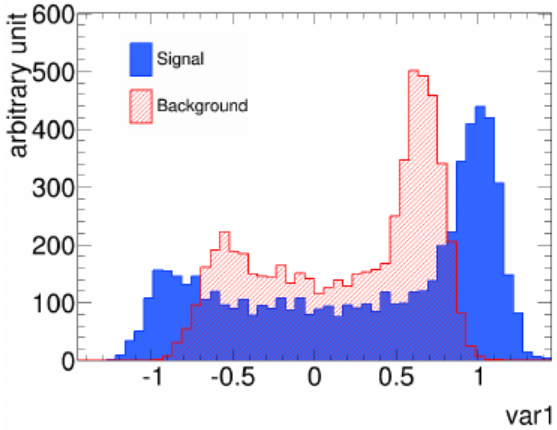
- What we want - Better Signal vs Background
- Why multivariable? Why not simply cut on variables?
  - Correlations unknown

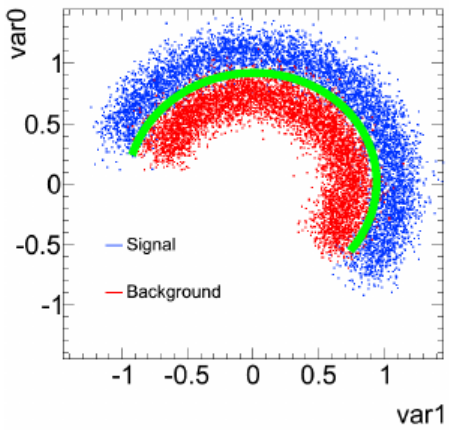


Introduction to Multivariable Analysis

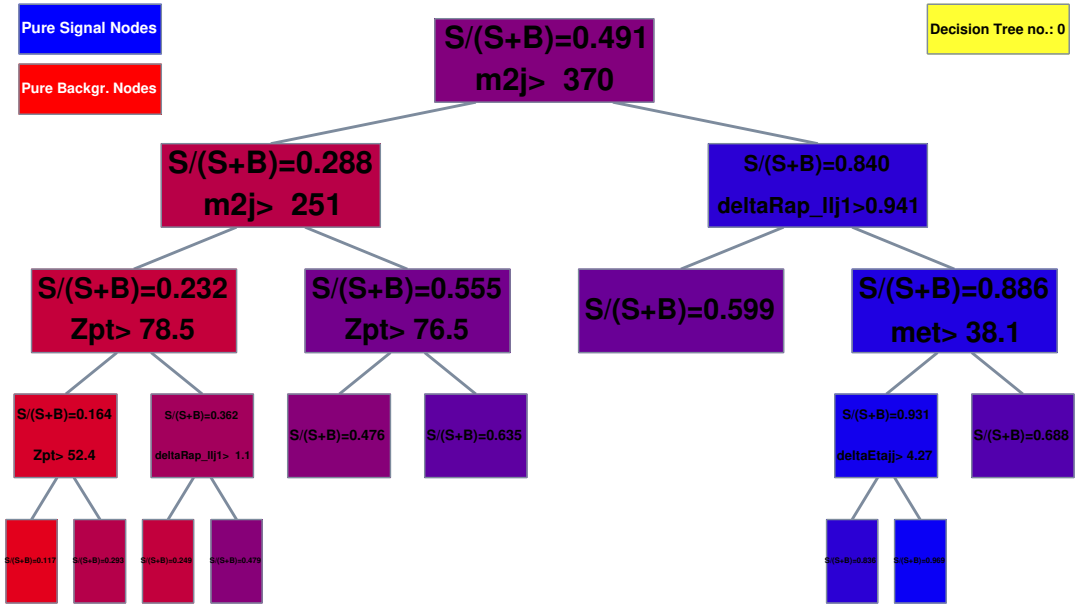
- What we want - Better Signal vs Background
- Why multivariable? Why not simply cut on variables?
  - Correlations unknown







Boosted Desicion Tree



- Desicion Tree: Classifier with a set of questions (nodes)
- Boosting: A single tree is weak, use a bunch of trees, combine weak learners into a strong learner
  - Boosting Method: AdaBoost, Gradient Boosting, Random forest...
- BDT is like a function: Event $\Rightarrow f\,(\mathrm{Event}) \Rightarrow$ Value
  Then use the values to seperate signal and background

ggF Categories

- Use BDT to select VBF events, the rest are classified into ggF region
- In ggF region, base on:
  - $p_{\mathrm{T}}^{\mu\mu}$: low (< 15 GeV), medium (15 - 50 GeV) and high(> 50 GeV)
  - Eta: Central: $|\eta_{\mu 1}| < 1$ and $|\eta_{\mu 2}| < 1$, Forward: the rest
- We used to separate the events into into 6 Categories:
  - LowPtll Central
  - MidPtll Central
  - HighPtll Central
  - LowPtll Forward
  - MidPtll Forward
  - HighPtll Forward
- Differenitate between $Z/\gamma^* \to \mu\mu$ and $H \to \mu\mu$
- $p_{\mathrm{T}}^{\mu\mu}$ shows strong discrimination
- Keep the eta Categories and put $p_{\mathrm{T}}^{\mu\mu}$ into MVA

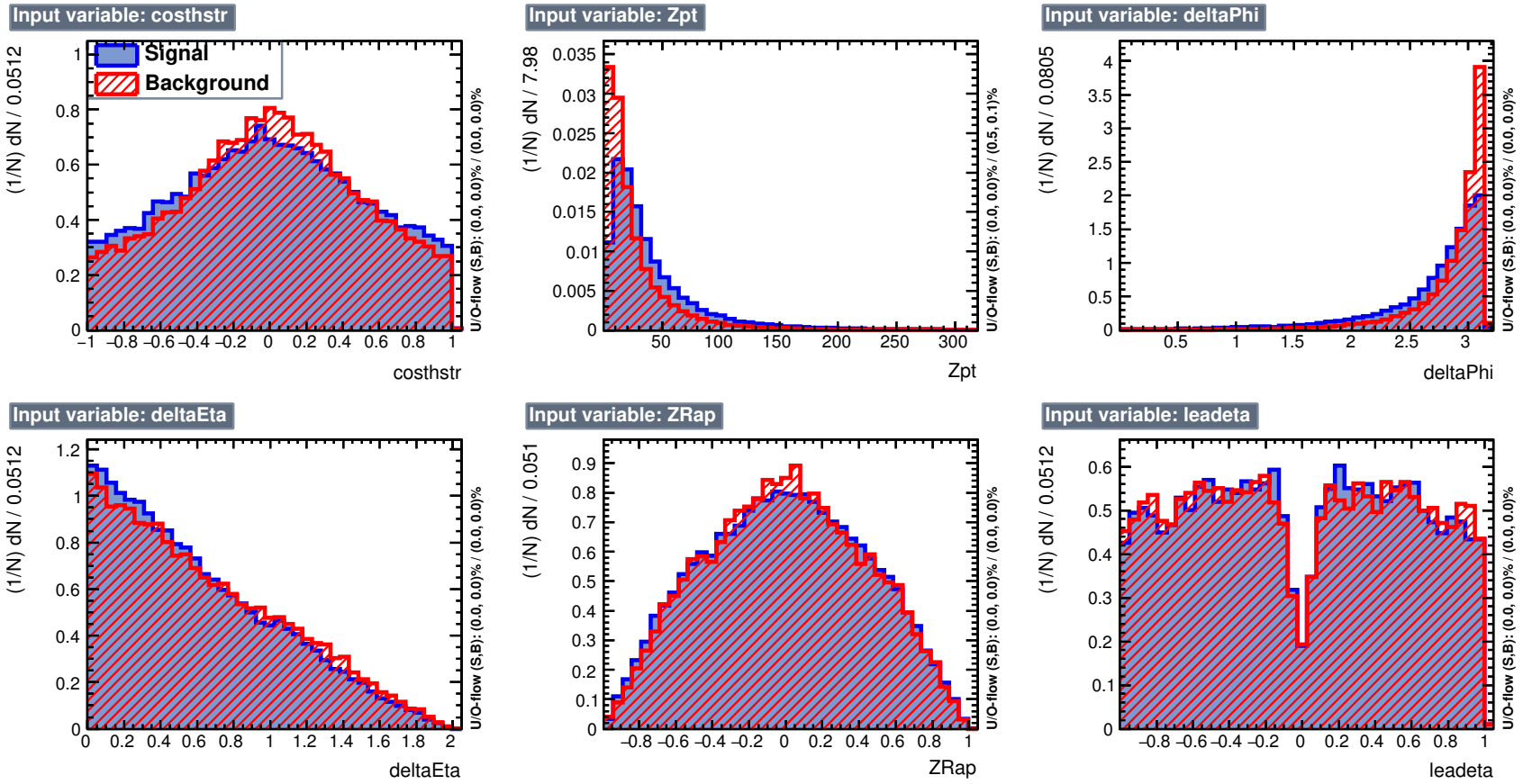
MVA Implementation

- Use 125 GeV ggF signal
- Only use Powheg Z + jets as background
- Use half of the events for training and half for testing
- $110 < m_{\mu\mu} < 150$
- Split the events into two regions, and train them separately:
  - Central: $|\eta_{\mu 1}| < 1$ AND $|\eta_{\mu 2}| < 1$
  - Forward: $|\eta_{\mu 1}| > 1$ OR $|\eta_{\mu 2}| > 1$

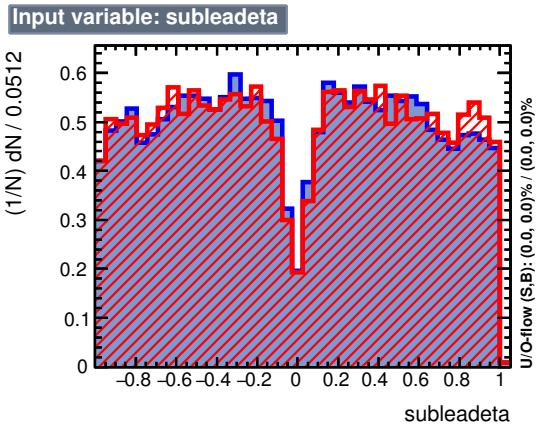
MVA Implementation

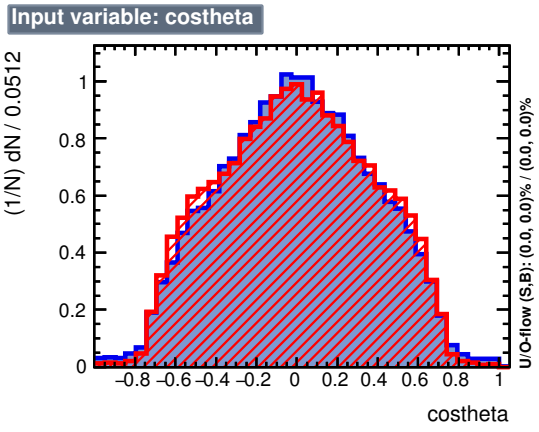
- Use following variables (9 overall):
  - Leading Eta
  - Subleading Eta
  - Ptll
  - yll
  - deltaEta
  - deltaPhi
  - $\cos\theta$ - angle between the highest pt muon in the dimuon rest frame with respect to z axis
  - $\cos\theta^*$ - angle between the minus muon in the dimuon rest frame with respect to dimuon vector
  - ϕ^* - $\Delta\phi$ between the highest pt muon in the dimuon rest frame with respect to dimuon vector
- Avoid variables having strong correlation with $m_{\mu\mu}$, like leading pt
- Also tried $p_{\mathrm{T}}^{\mu\mu}/m_{\mu\mu}$, but almost the same distributon with $p_{\mathrm{T}}^{\mu\mu}$

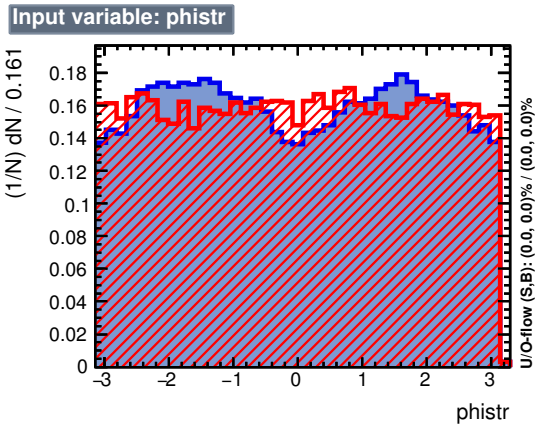
Input Variables Distribution - Central



Input Variables Distribution - Central







Correlation Matrix - Central

Correlation Matrix (signal)

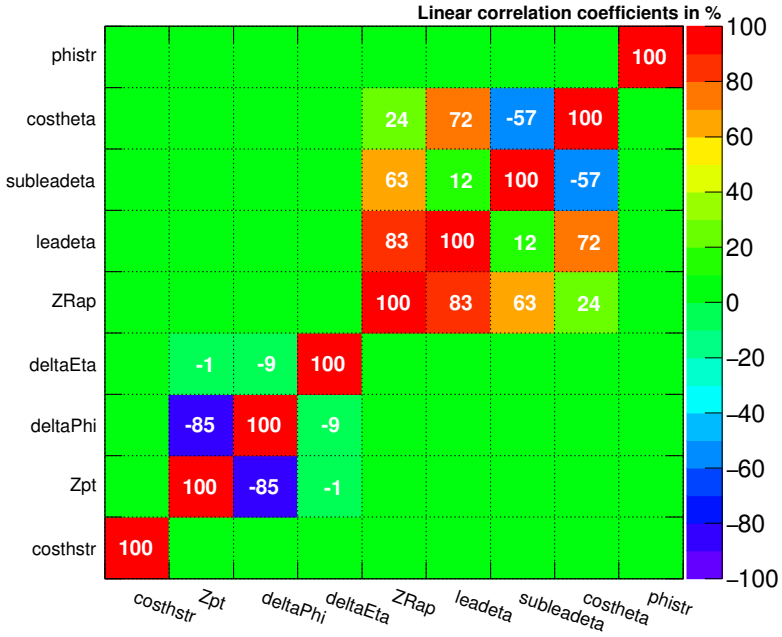


Figure: Signal

Correlation Matrix (background)

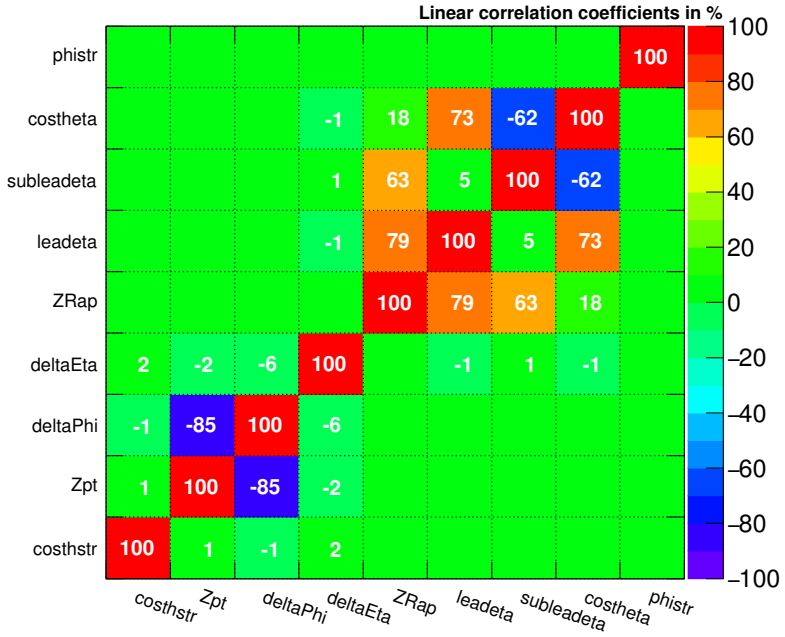
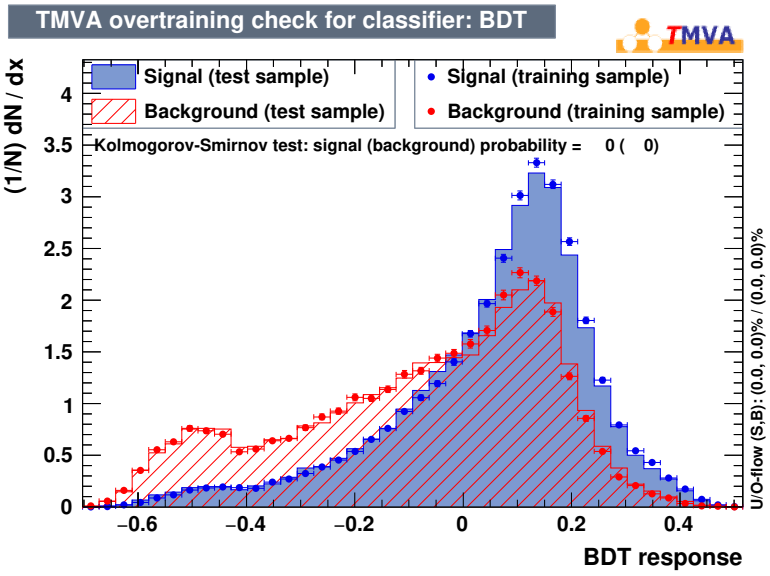
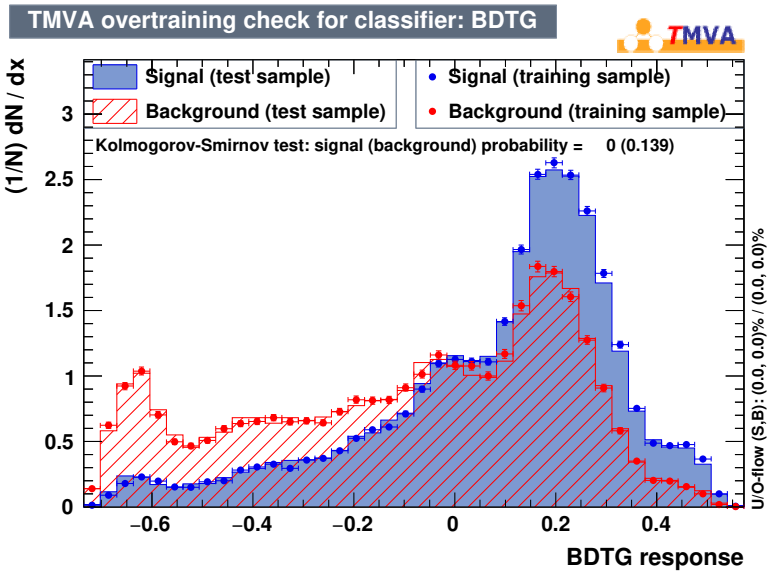
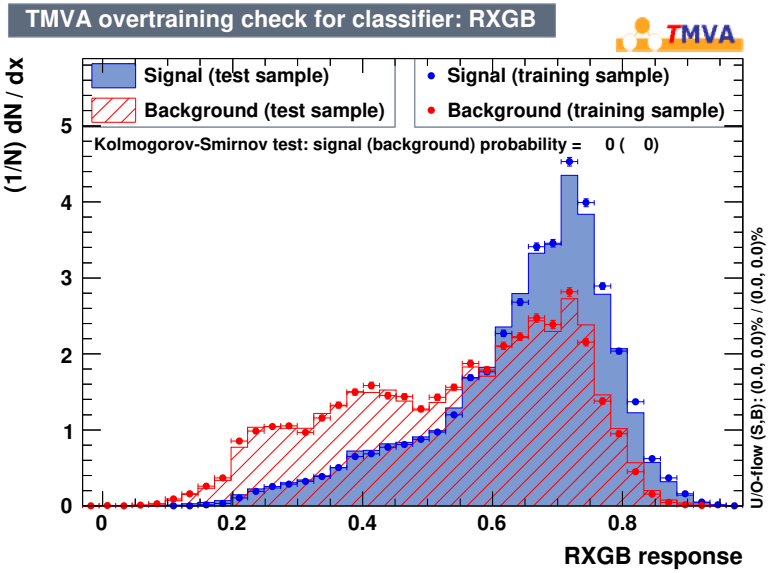


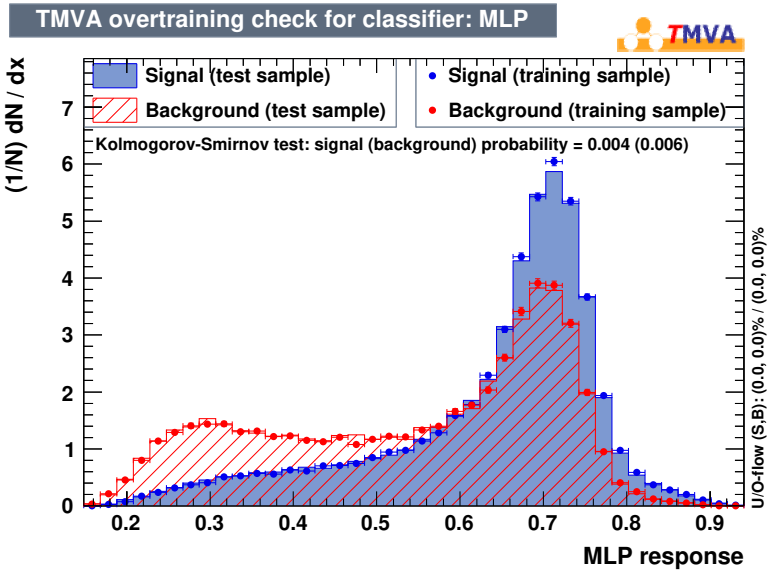
Figure: Background

MVA Output - Central

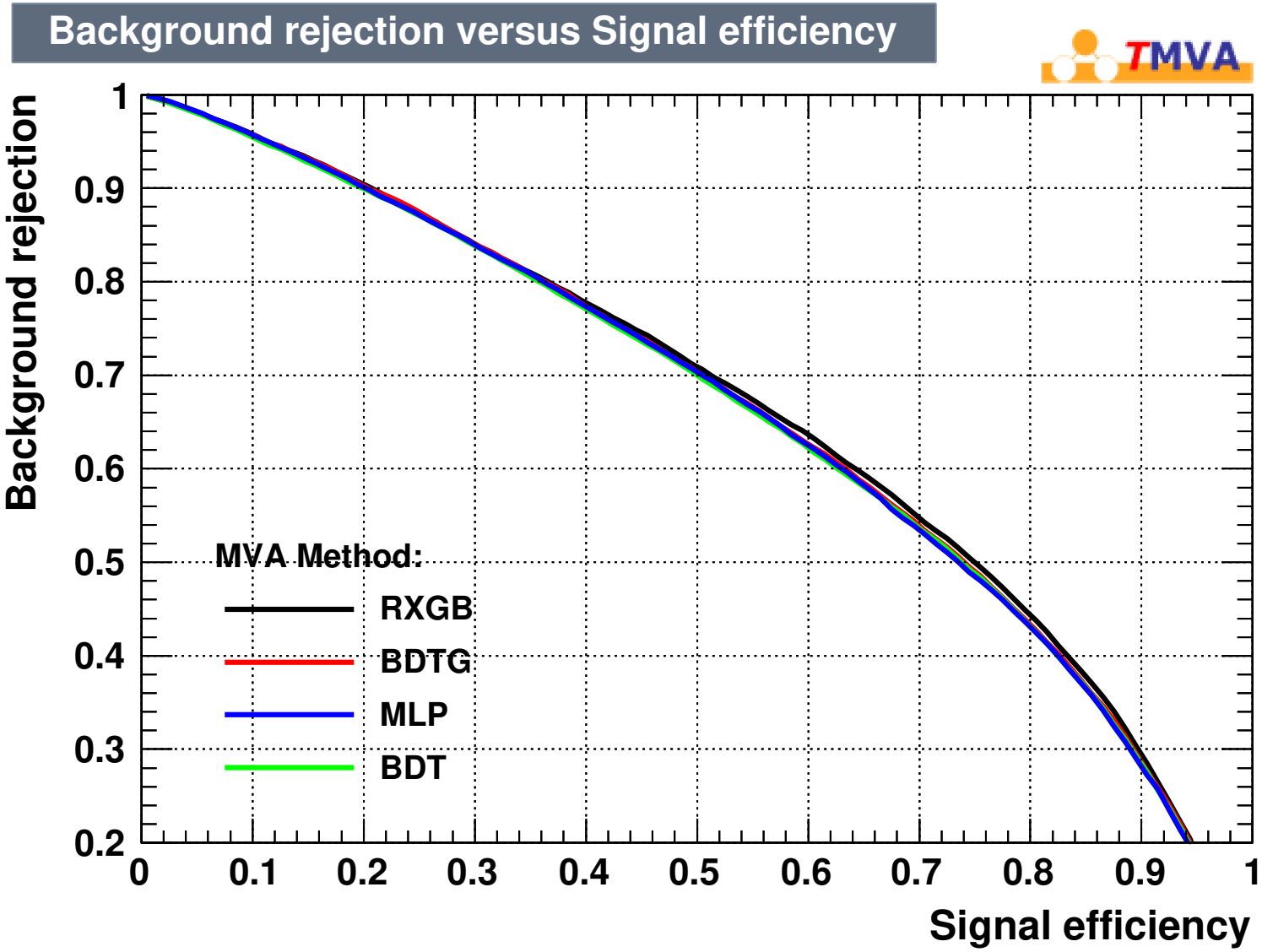




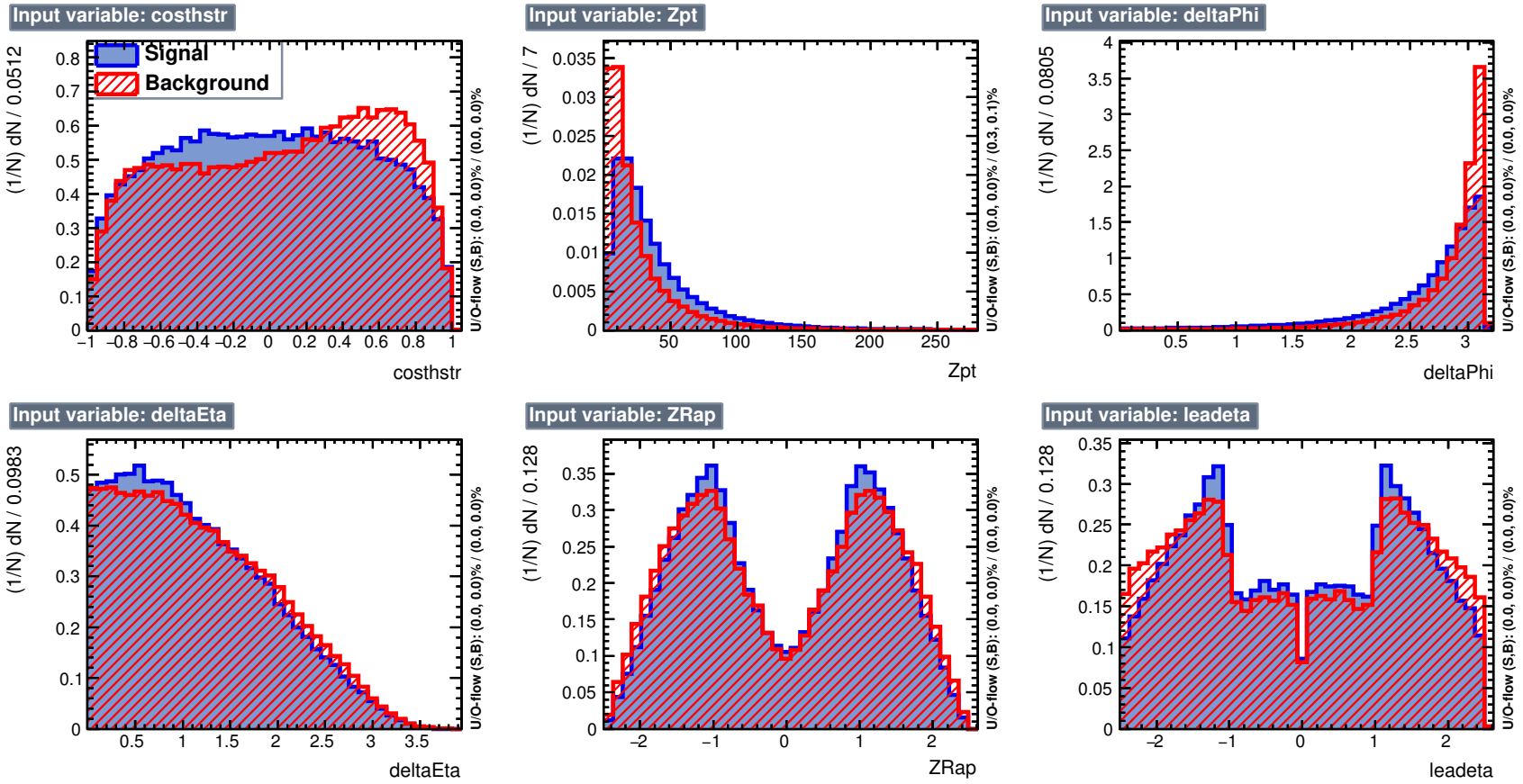




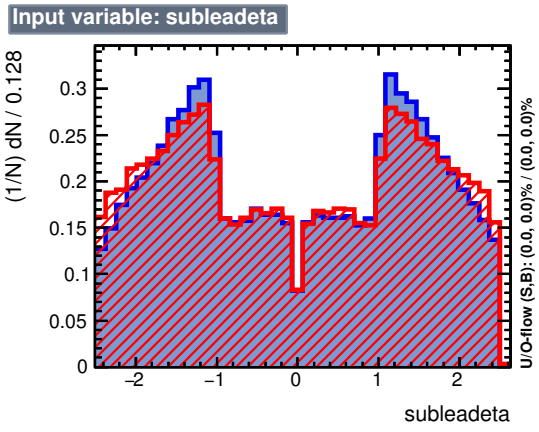
ROC Curve - Central

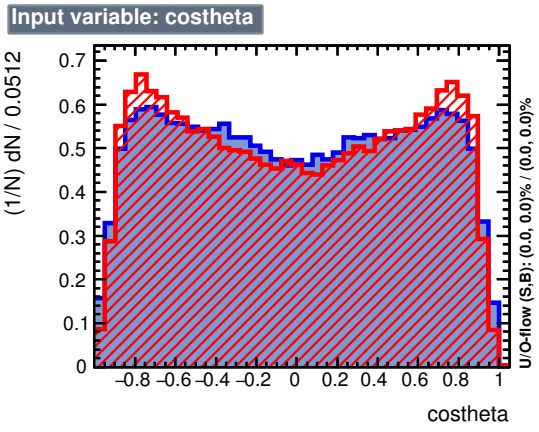


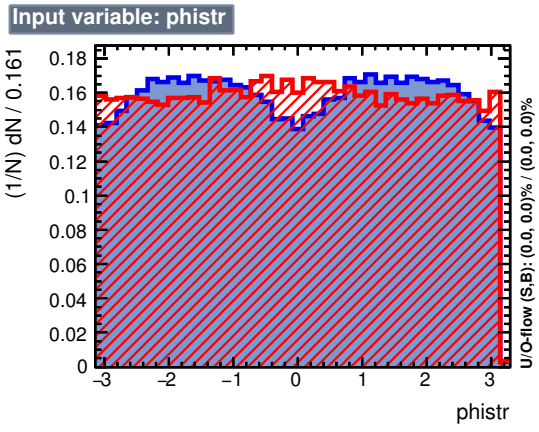
Input Variables Distribution - Forward



Input Variables Distribution - Forward







Correlation Matrix - Forward

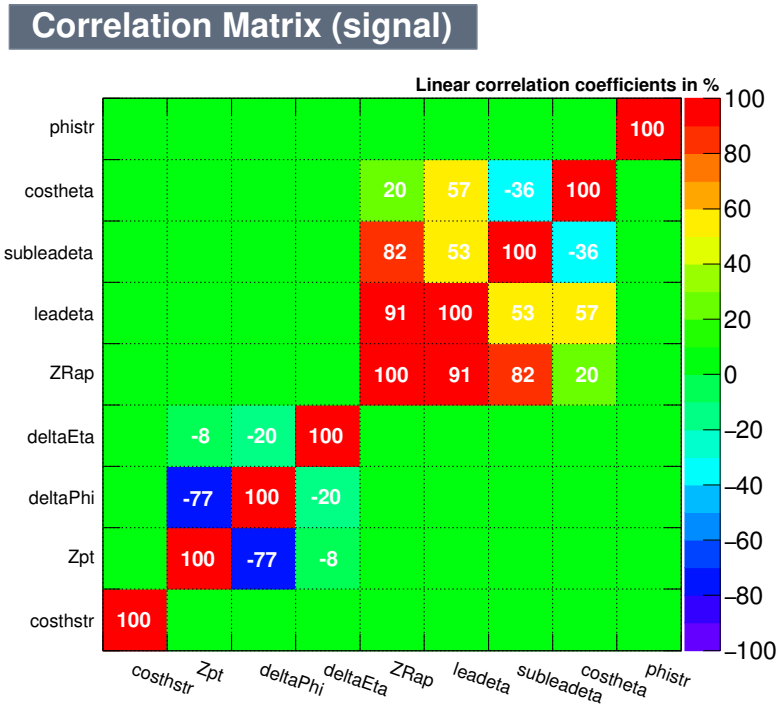


Figure: Signal

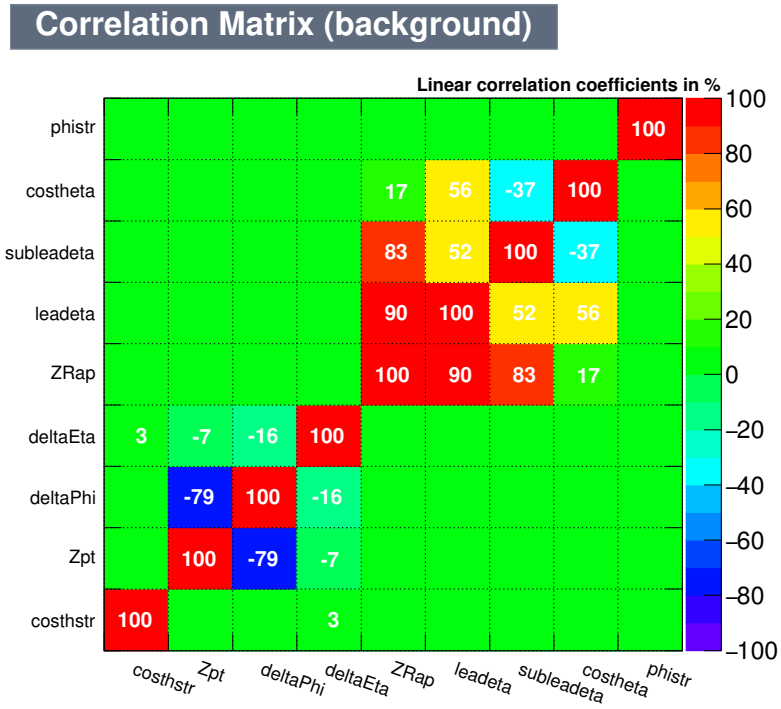
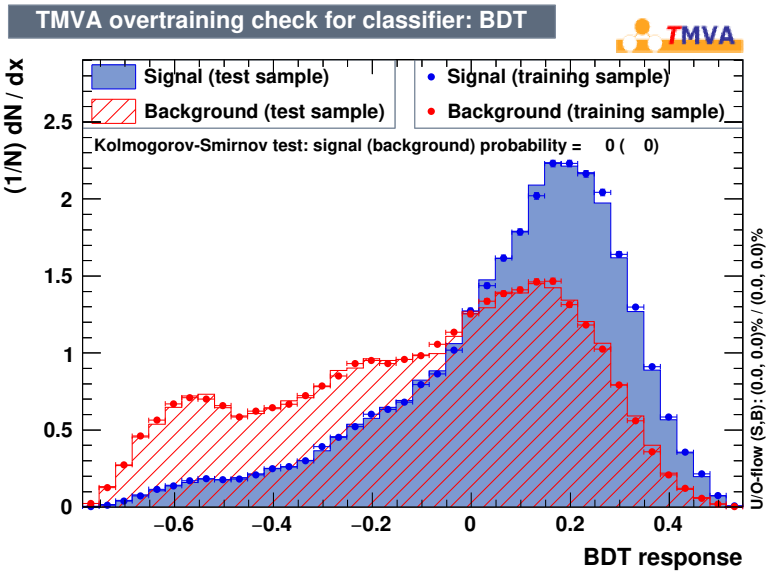
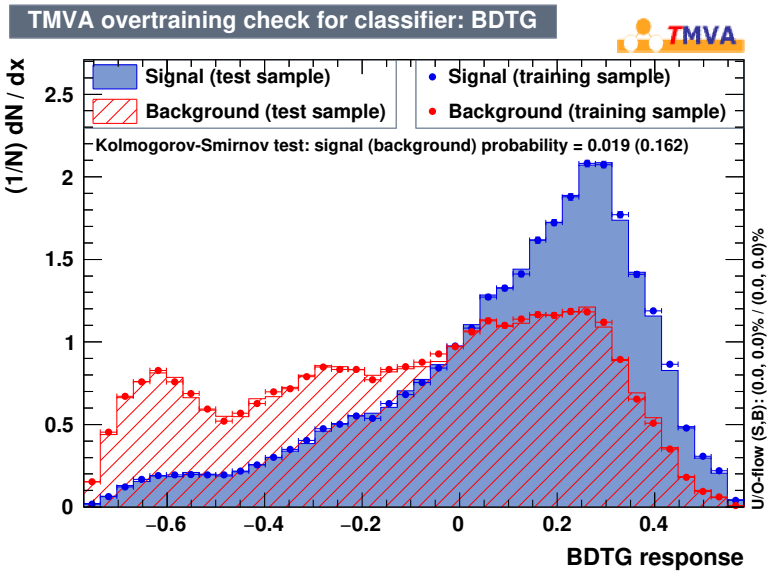
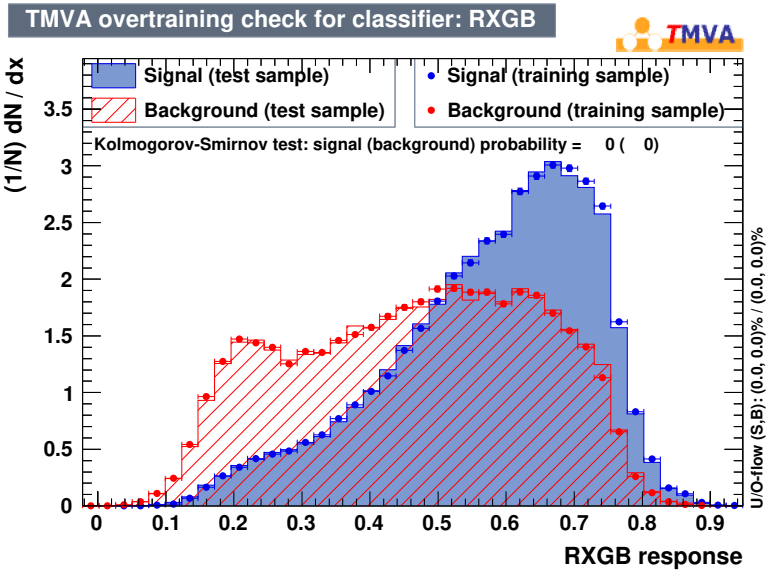


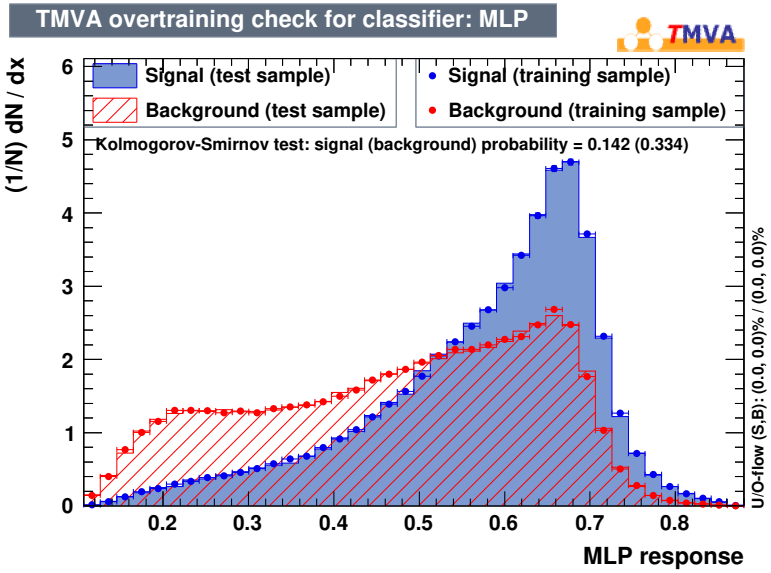
Figure: Background

MVA Output - Forward

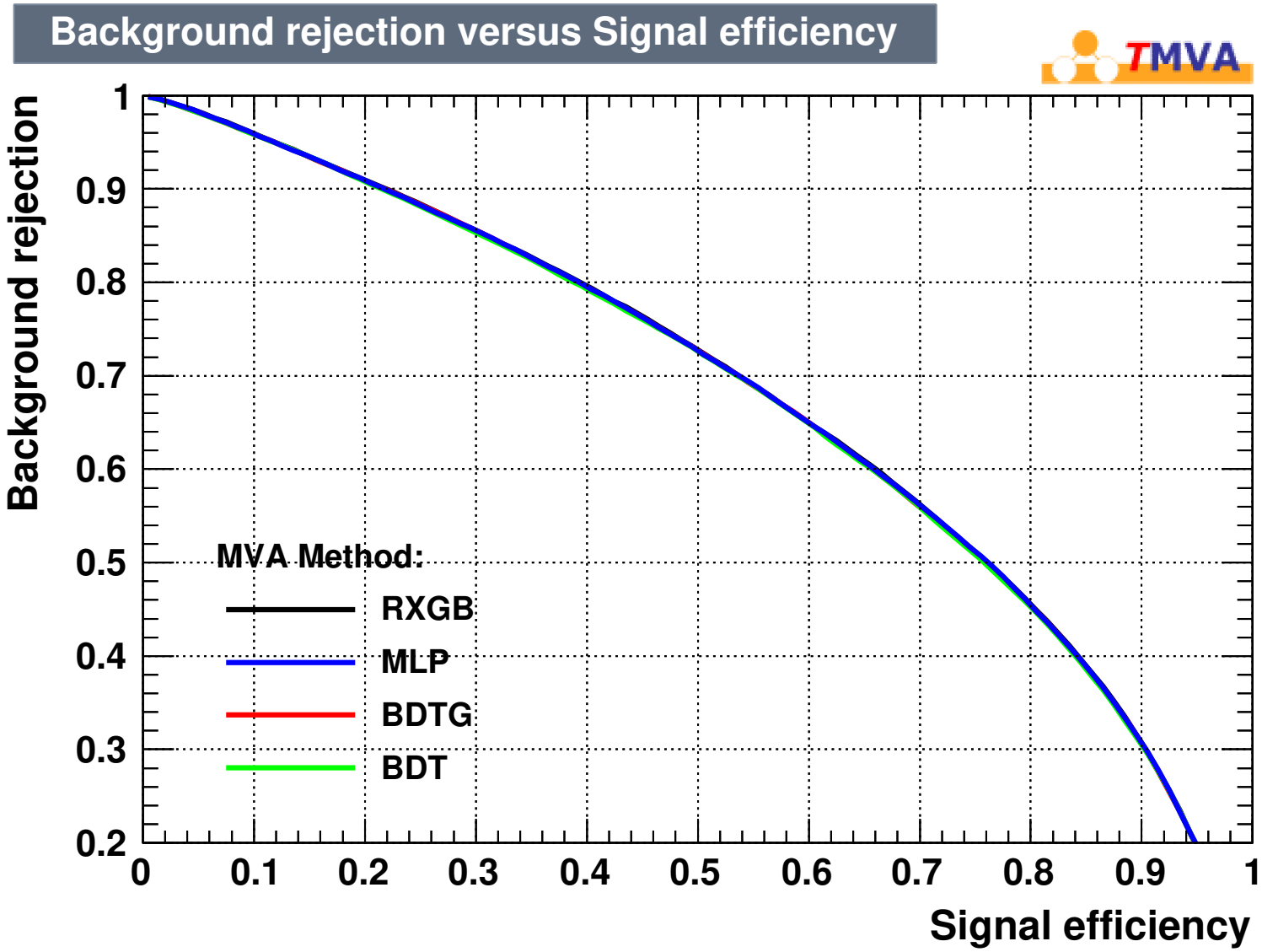








ROC Curve - Forward



MVA Method Used

- Scanned over different MVA methods and parameters roughly

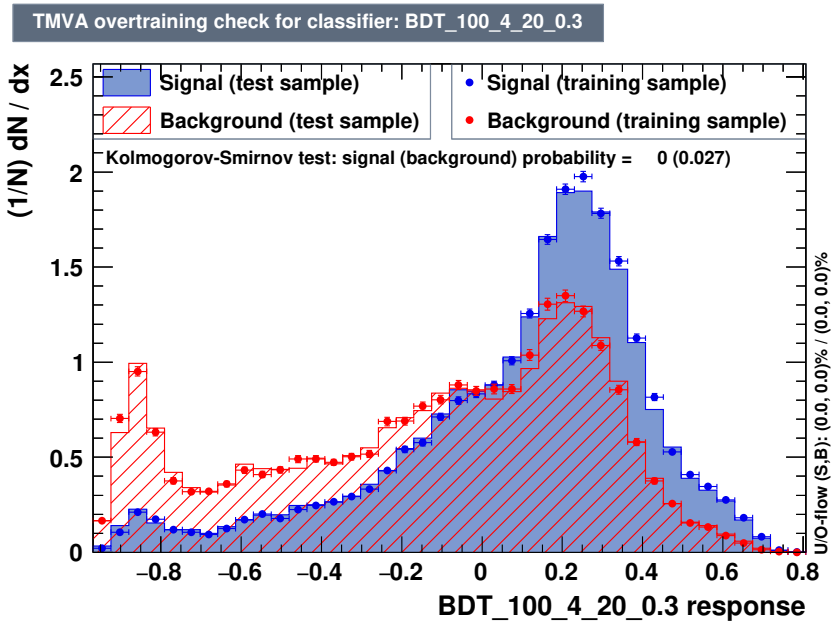


Figure: mva_central

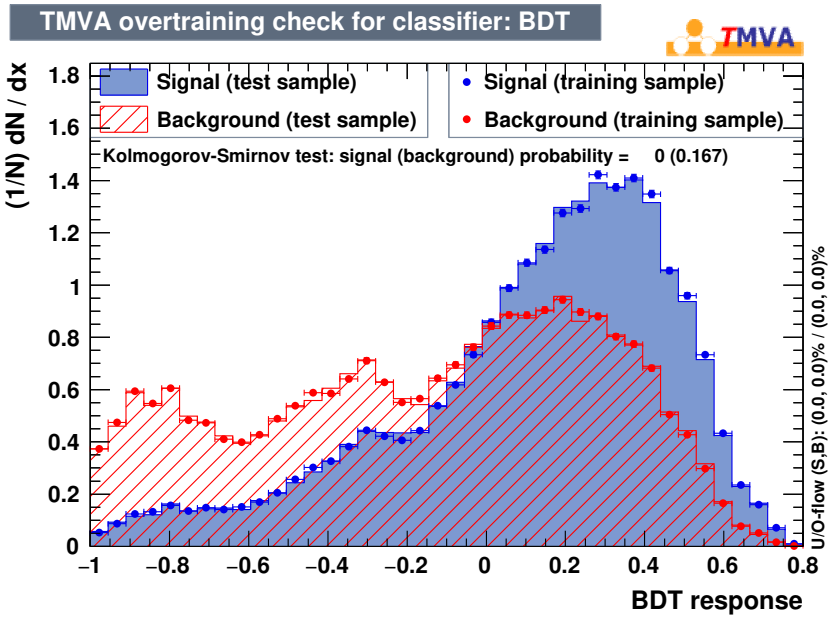


Figure: mva_forward

- Use Adaptive BDT method for both central and forward region
- ntrees = 100, Max depth = 4, nCuts = 20, Learning rate = 0.3

Application and New Categories

- Do application on signal and data
- Two output: mva_central and mva_forward
- Cut on: (-0.4, 0.2) for both regions
- 6 ggF Categories:
  - LowMVA Central: mva_central < -0.4
  - LowMVA Forward: mva_forward < -0.4
  - MidMVA Central: -0.4 < mva_central < 0.2
  - MidMVA Forward: -0.4 < mva_forward < 0.2
  - HighMVA Central: mva_central > 0.2
  - HighMVA Forward: mva_forward > 0.2
- 2 VBF Categories:
  - VBF Loose
  - VBF Tight

Event Yields in 120 -130 GeV

Region	S	B	$S/\sqrt{B}$
Central LowPtll	11.32	9714.0	0.11
Forward LowPtll	30.85	37941.0	0.16
Central MidPtll	23.81	8027.0	0.27
Forward MidPtll	63.84	29756.0	0.37
Central HighPtll	14.80	3416.0	0.25
Forward HighPtll	35.30	11255.0	0.33
VBF Loose	3.25	237.0	0.21
VBF Tight	3.30	69.0	0.40

Table: Cut based categories

Region	S	B	$S/\sqrt{B}$
Central LowMVA	4.663	5815.0	0.06
Forward LowMVA	12.18	22539.0	0.08
Central MidMVA	20.65	9211.0	0.22
Forward MidMVA	52.22	34389.0	0.29
Central HighMVA	24.62	6131.0	0.31
Forward HighMVA	65.60	22024.0	0.44
VBF Loose	3.25	237.0	0.21
VBF Tight	3.30	69.0	0.40

Table: MVA method categories

95% CL limits

Category	Obs.	Exp.	+2Sig	+1Sig	-1Sig	-2Sig
Cut based:	3.1118	3.1557	5.8852	4.3613	2.2739	1.6937
MVA Method:	2.9973	3.1227	5.8240	4.3151	2.2500	1.6760

3.1557 vs 3.1227, about 1% better

Other Settings

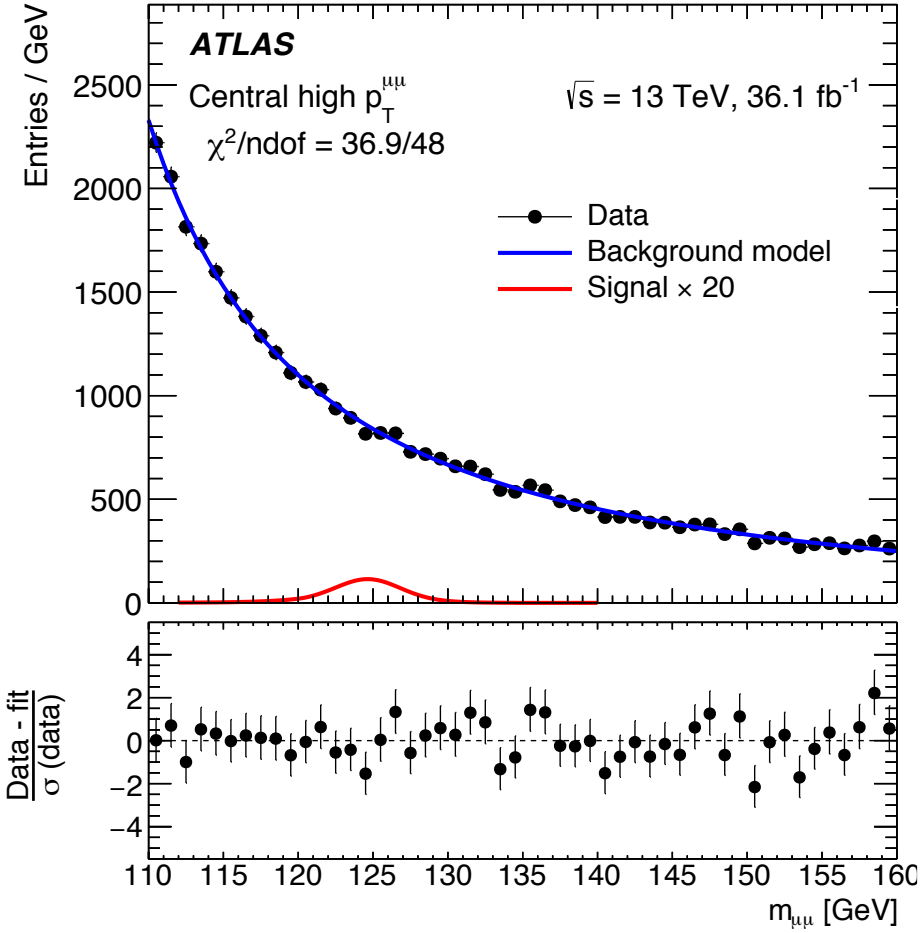
- Tried different cuts on MVA output
- Most of them give around 1% improvement
- Two special points give 5% improvement:
  - Central: (-0.3, 0.1) Forward: (-0.3, 0.1)
  - Central: (-0.3, 0.1) Forward: (-0.4, 0.1)
- However, these two points are extremly sensitive
- Change 0.1 in Central to 0.11 or 0.09 (or even 0.101), results go back to 1% (Comparision in back up)
- Most likely due to fluctuation
- Also tried to add one more category for each region (4 overall for each), no better results observed

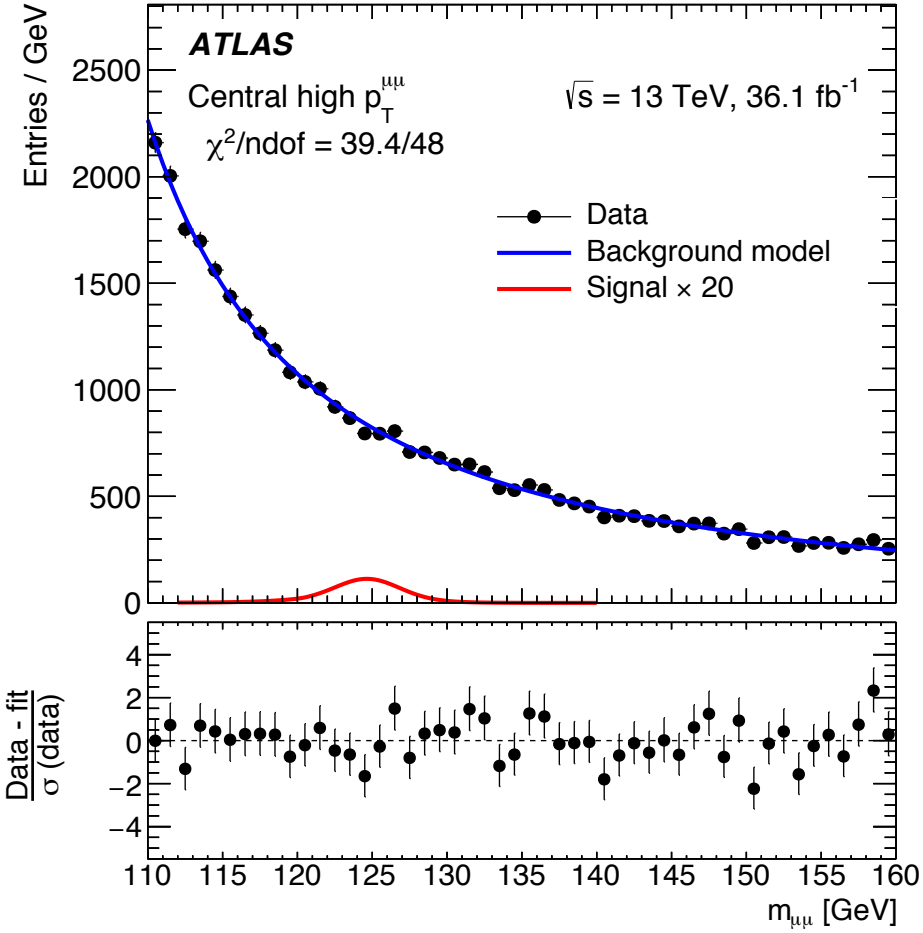
Conclusion

- Use MVA instead of Cut based method in ggF region
- Use BDT with 9 variables to separate signal and background
- Performance not very good, about 1% improvement on 95% CL limits
- Tried different cut methods, no better results observed

Back up

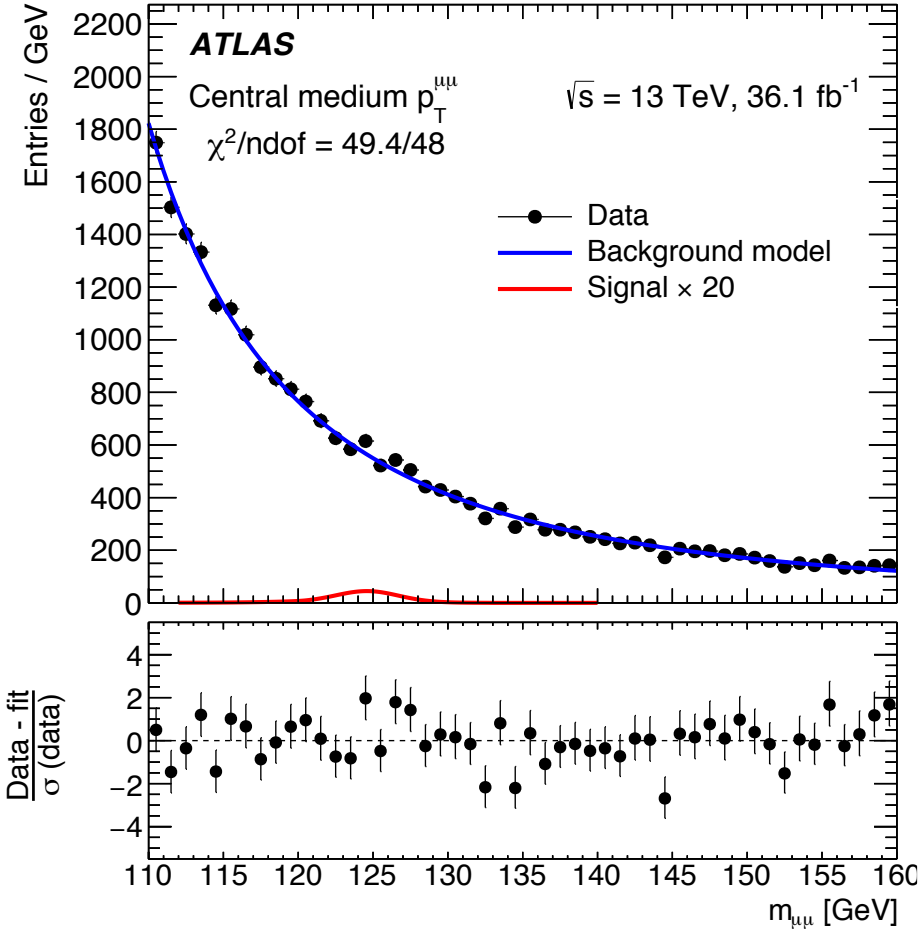
Fitting Compare in HighMVA Central, 0.1 vs 0.11

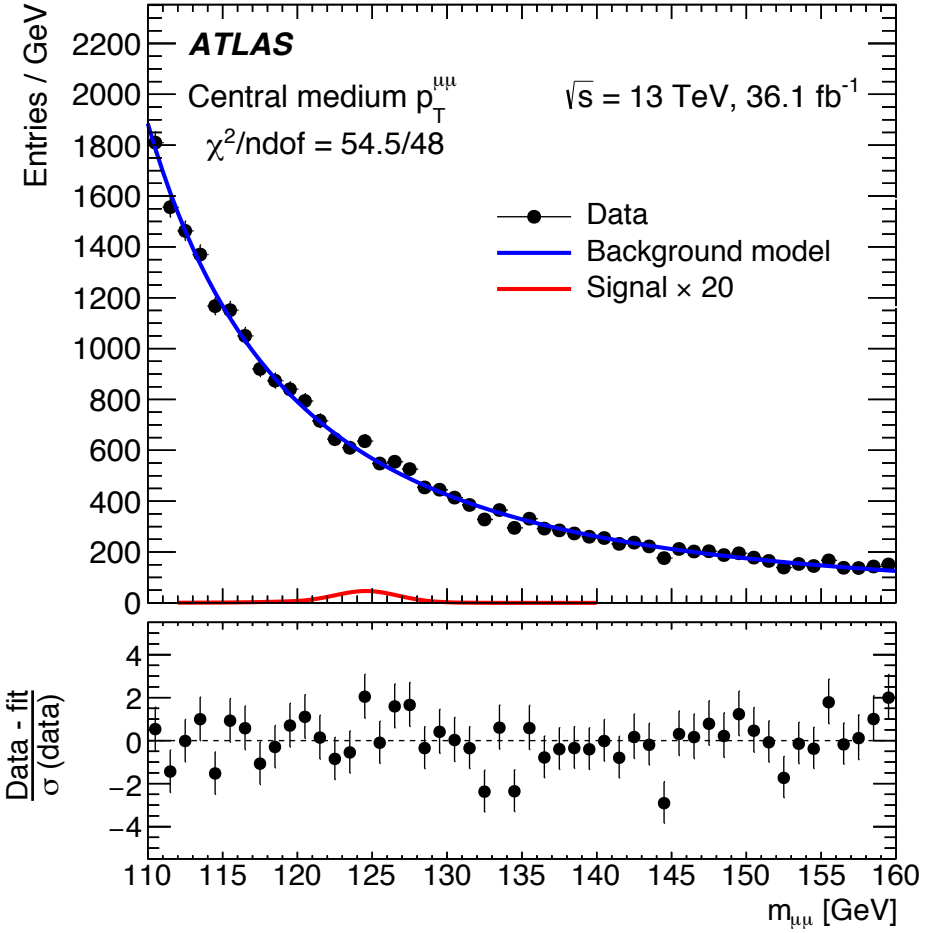




Fittings look similar

Fitting Compare in MidMVA Central , 0.1 vs 0.11





Fittings look similar

Cross Check for VBF BDT and Investigation on Centrality

Haifeng Li, Yidi Qi, Dmitri Tsybychev

Thank Yanlin for the help!

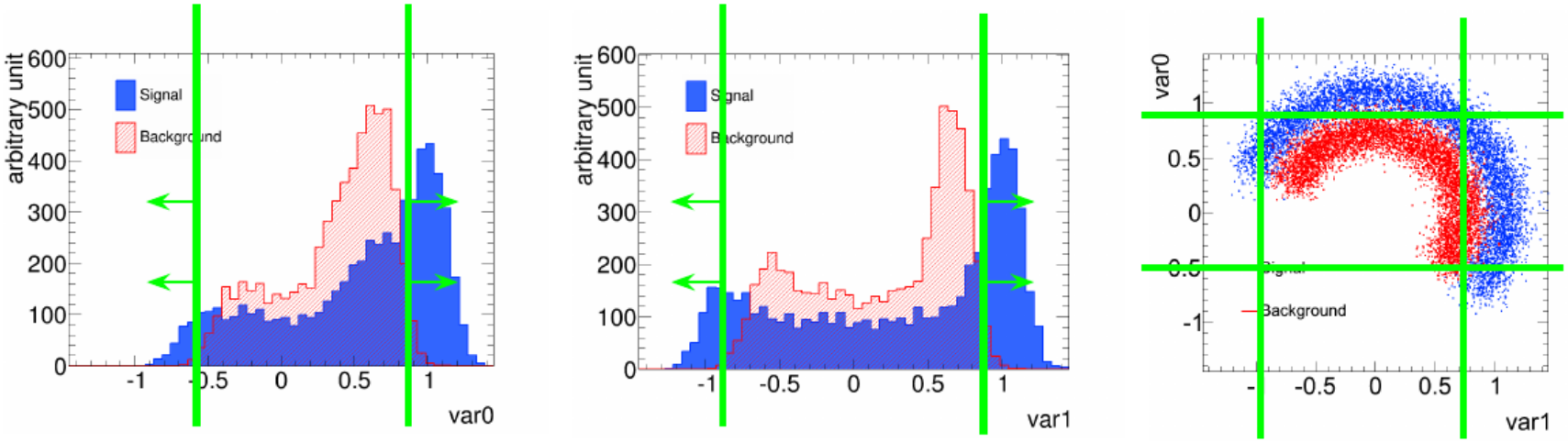
Stony Brook University

July 13, 2017

Introduction

Introduction to Multivariable Analysis

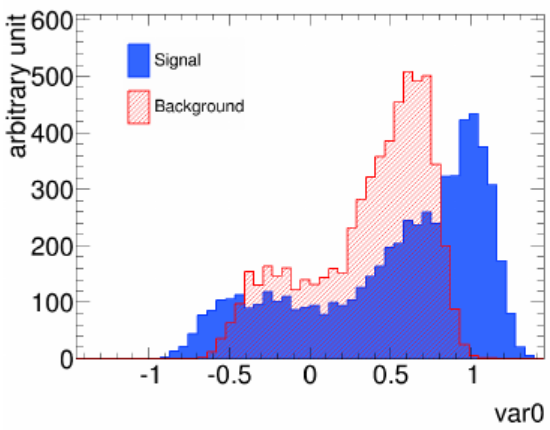
- What we want - Better Signal vs Background
- Why multivariable? Why not simply cut on variables?
  - Correlations unknown

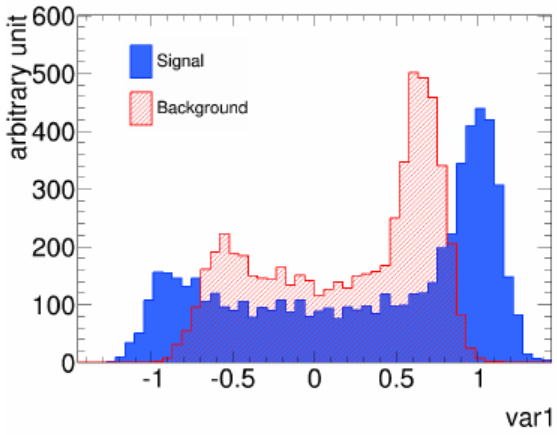


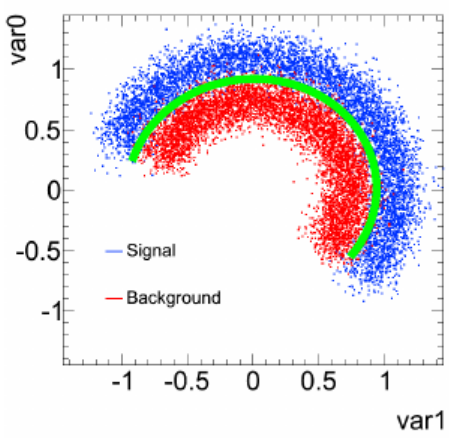
Ref: https://indico.scc.kit.edu/indico/event/48/session/4/contribution/35/material/slides/0.pdf

Introduction to Multivariable Analysis

- What we want - Better Signal vs Background
- Why multivariable? Why not simply cut on variables?
  - Correlations unknown

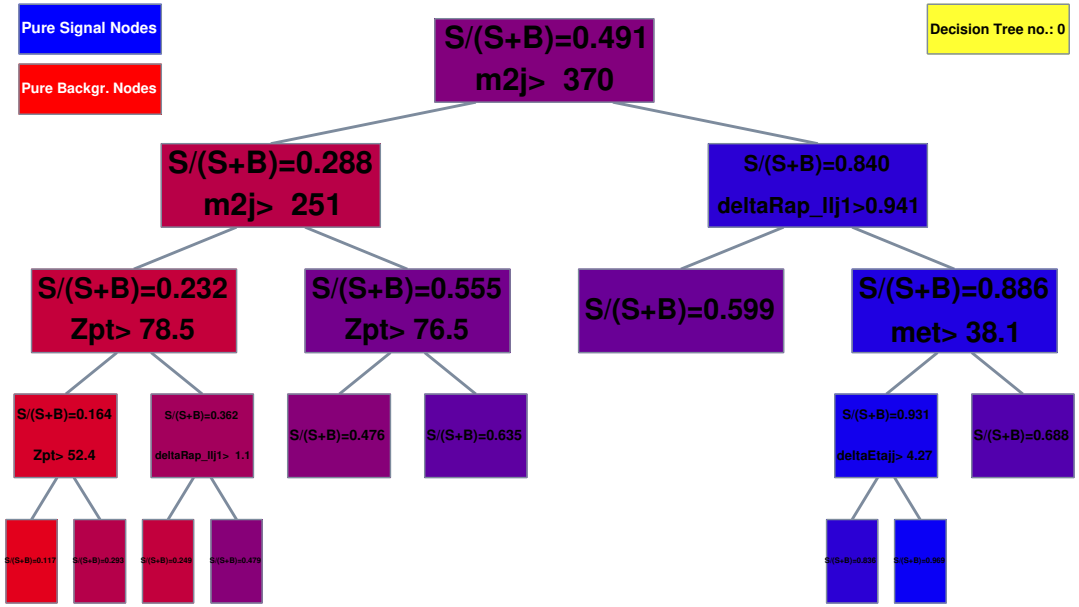






Ref: https://indico.scc.kit.edu/indico/event/48/session/4/contribution/35/material/slides/0.pdf

Boosted Desicion Tree



- Desicion Tree: Classifier with a set of questions (nodes)
- Boosting: A single tree is weak, use a bunch of trees, combine weak learners into a strong learner
  - Boosting Method: AdaBoost, Gradient Boosting, Random forest...
- BDT is like a function: Event $\Rightarrow f\,(\mathrm{Event}) \Rightarrow$ Value
  Then use the values to seperate signal and background

Cross Check for VBF BDT

- Cross check to Yanlin's VBF BDT results in the supporting note ATL-COM-PHYS-2016-1194,
- Input Ntuple was provided by Yanlin, slightly different from the one used in the note.
- Use a different macro from Yanlin's

MVA Implementation

- Choose following variables (14 overall):
  - $\Delta\eta_{jj}$
  - ΔR_{jj}
  - $centrality = \frac{y_{\mu\mu} - \frac{y_{j_1} + y_{j_2}}{2}}{|y_{j_1} - y_{j_2}|}$
  - m_{jj}
  - p_T^{jj}
  - $p_T^{\mu\mu}$
  - E_T^{miss}
  - Ht (Scalar sum of muons and jets p_T)
  - Rapity and p_T between dimuon system with leading/subleading/both jet(s):
  - $y_{\mu\mu j1}$
  - $y_{\mu\mu j2}$
  - $y_{\mu\mu jj}$
  - $p_T^{\mu\mu j1}$
  - $p_T^{\mu\mu j2}$
  - $p_T^{\mu\mu jj}$

MVA Implementation

- Use Z+jets, Top and Diboson as background for training
- Split the events in two, first half for training and second for testing
- $110 < m_{\mu\mu} < 140$
- Use Gradient boosting type, bagged boost is adopted
- ntrees = 200, Max depth = 4, nCuts = 20, shrinkage = 0.1

Input Variables Distribution

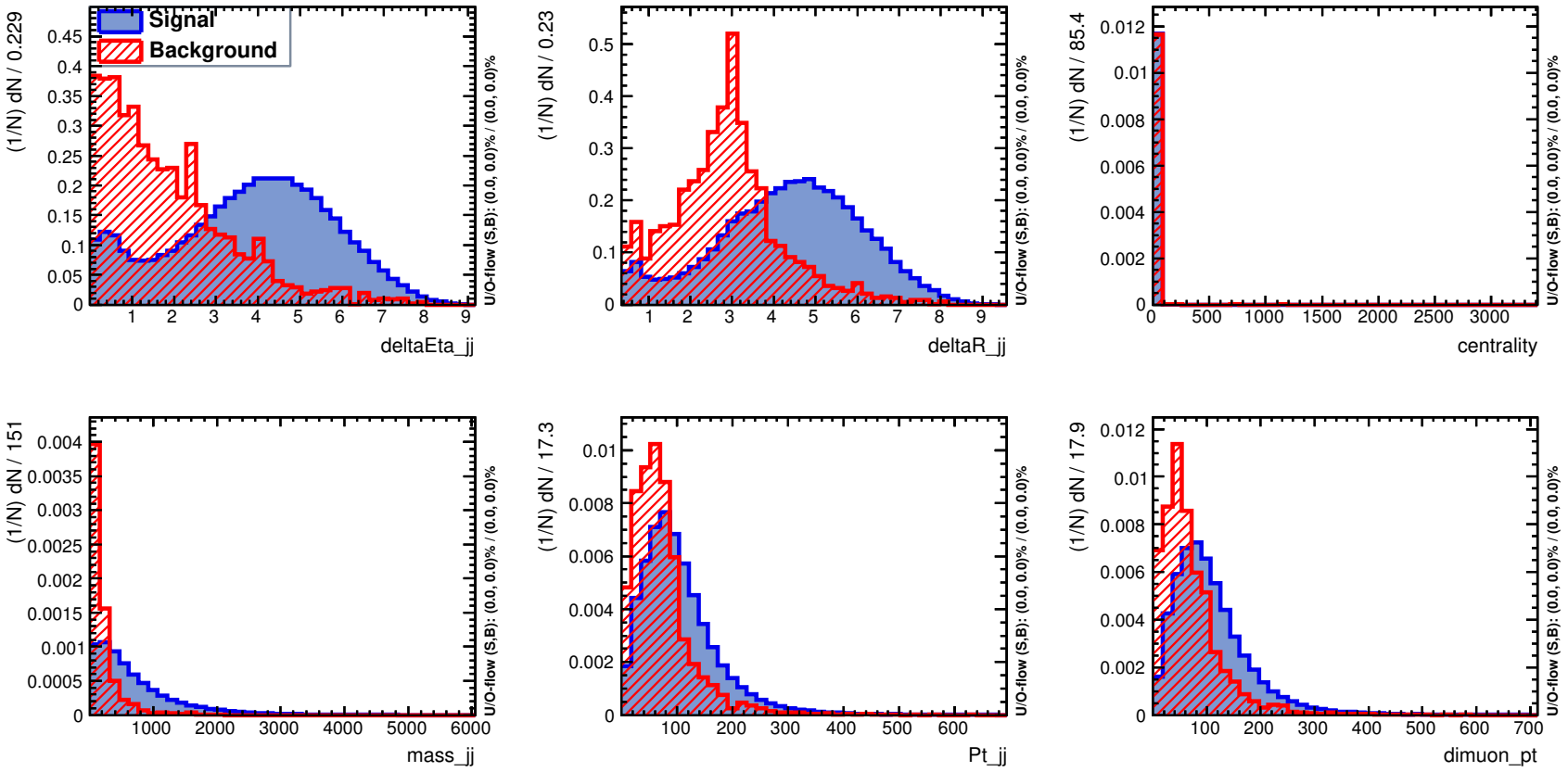


Figure: Input variables distribution

Input Variables Distribution

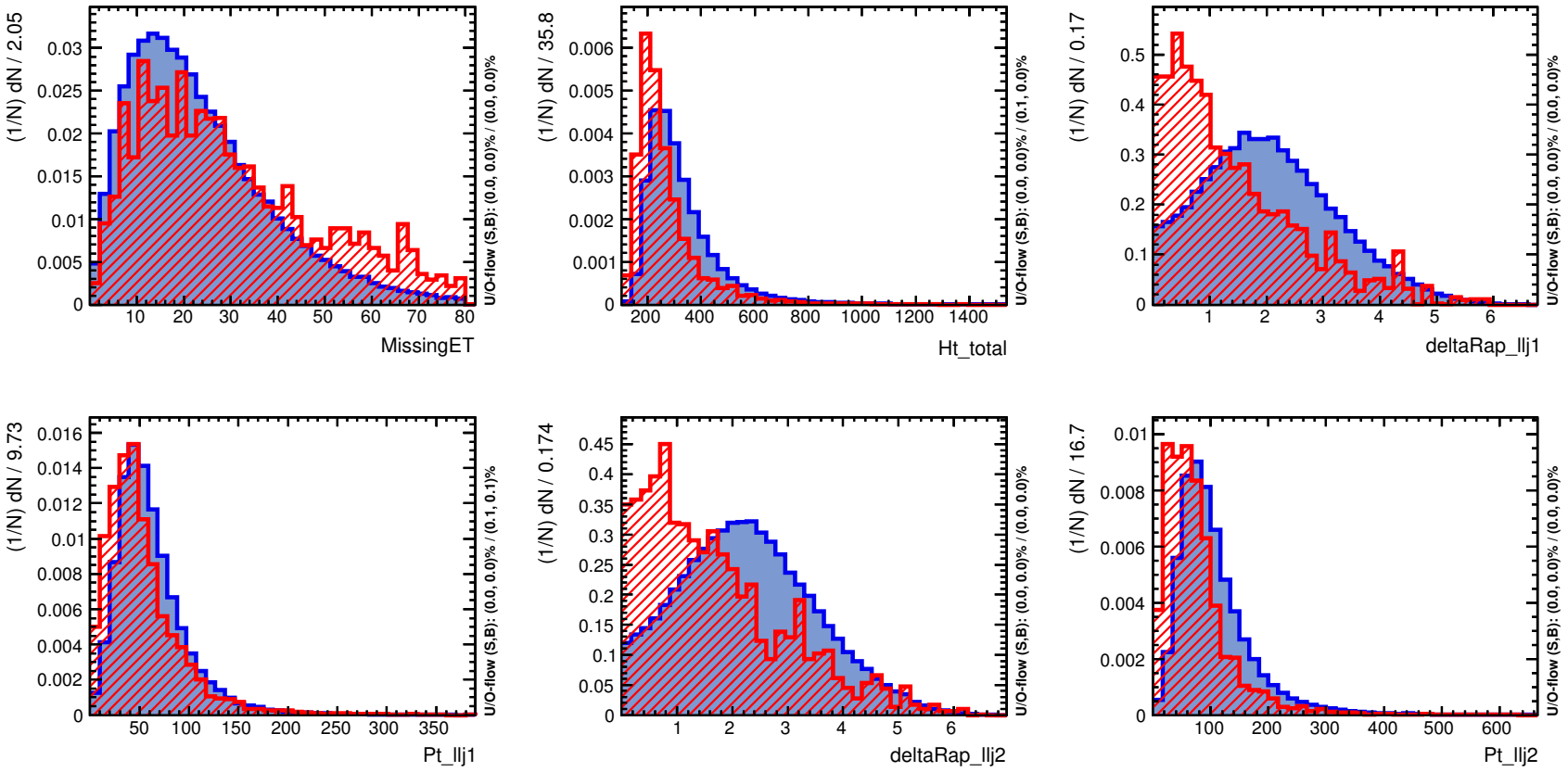


Figure: Input variables distribution

Input Variables Distribution

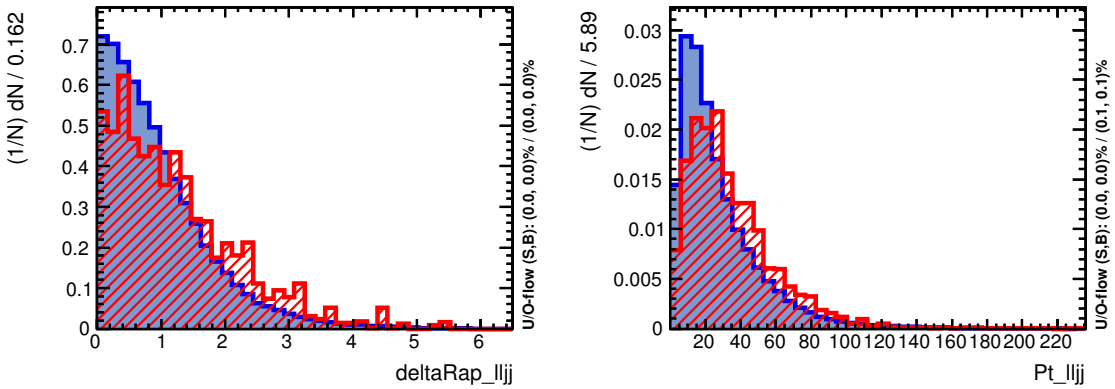


Figure: Input variables distribution

Correlation Matrix (Signal)

Yidi:

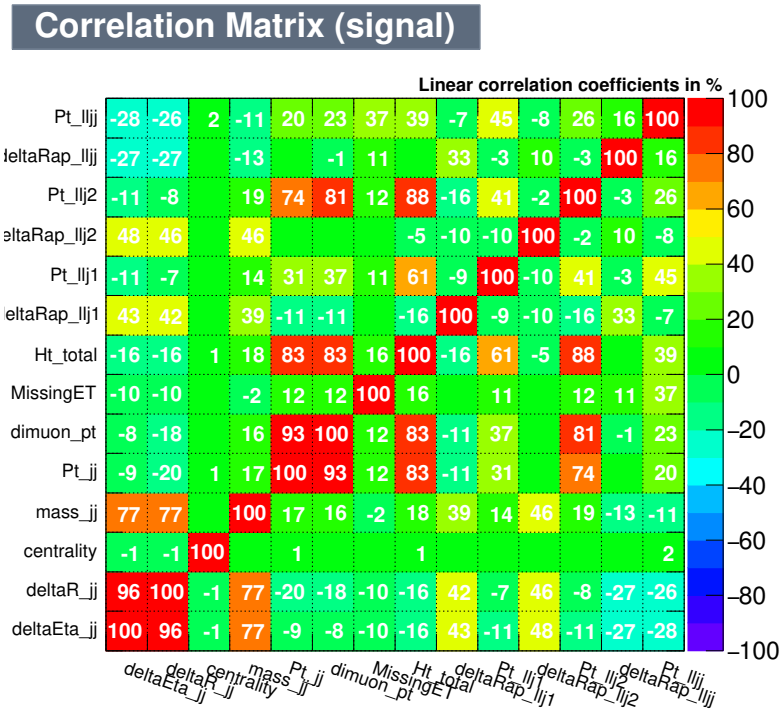


Figure: Correlation in cross check

Yanlin:

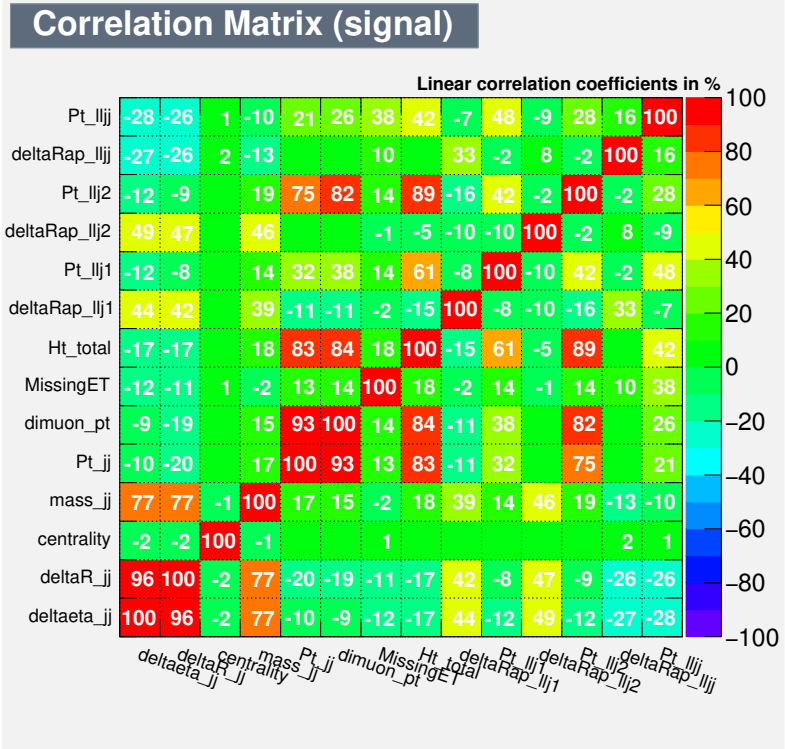


Figure: Correlation in the note

Correlation Matrix (Background)

Yidi:

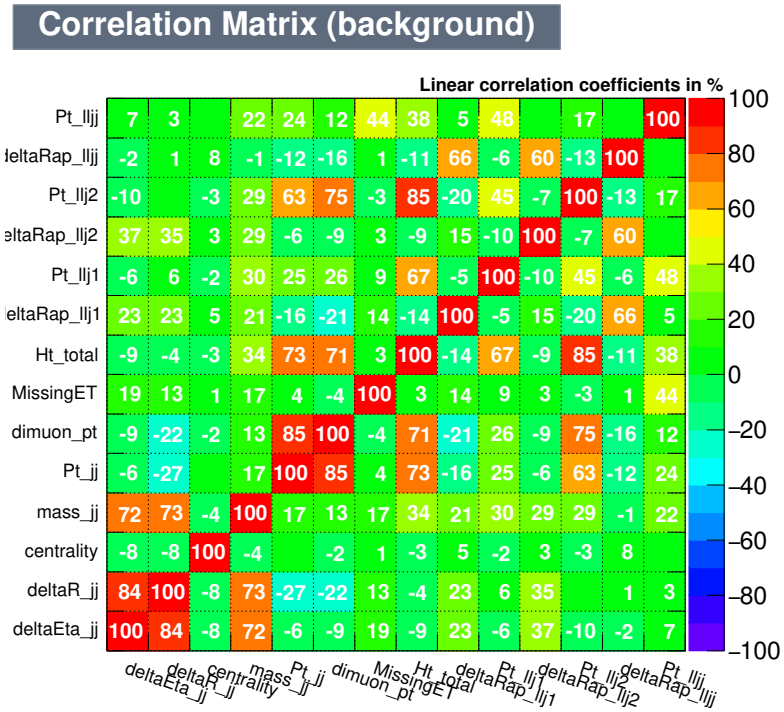


Figure: Correlation in cross check

Yanlin:

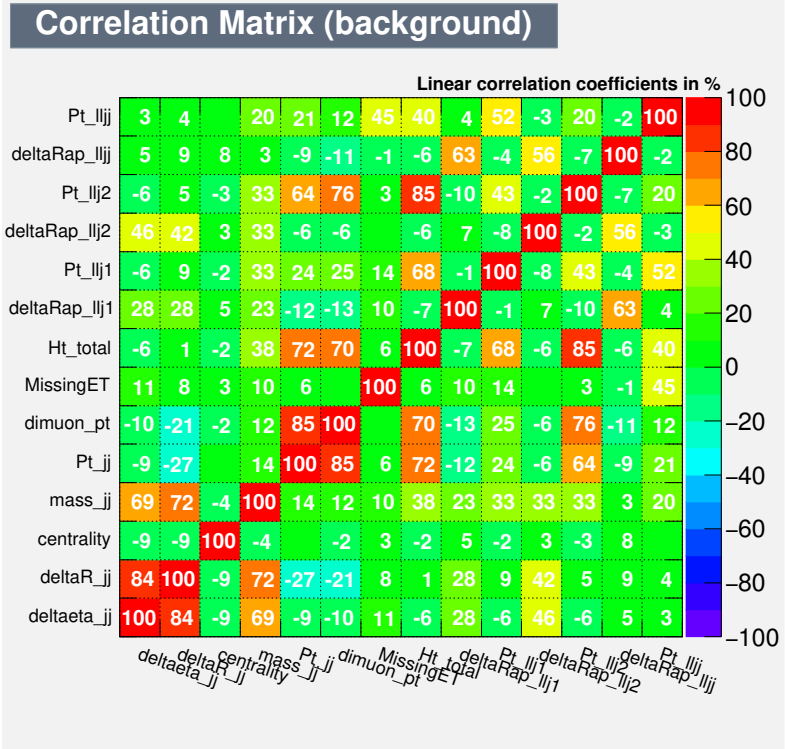


Figure: Correlation in the note

Ranking

Yidi:

Rank	Variable	Variable Importance
1	$\Delta\eta_{jj}$	2.061e-01
2	m_{jj}	1.722e-01
3	$p_T^{\mu\mu}$	1.373e-01
4	Ht	7.660e-02
5	ΔR_{jj}	6.848e-02
6	p_T^{llj2}	6.117e-02
7	$y_{\mu\mu j1}$	4.533e-02
8	p_T^{jj}	4.256e-02
9	centrality	4.205e-02
10	E_T^{miss}	3.730e-02
11	p_T^{llj1}	3.604e-02
12	p_T^{lljj}	3.032e-02
13	$y_{\mu\mu jj}$	2.890e-02
14	$y_{\mu\mu j2}$	1.573e-02

Table: Ranking in cross check

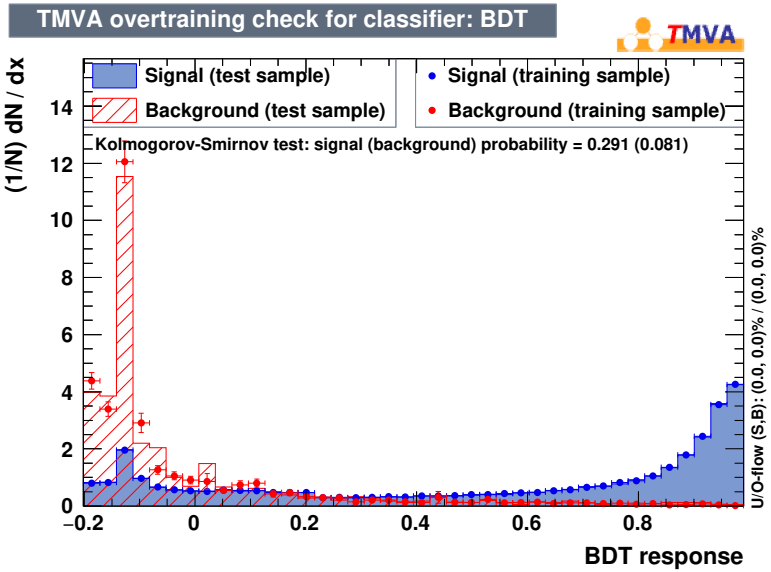
Yanlin:

Rank	Variable	Variable Importance
1	$\Delta\eta_{jj}$	2.440e-01
2	m_{jj}	1.731e-01
3	$p_T^{\mu\mu}$	1.338e-01
4	ΔR_{jj}	9.217e-02
5	p_T^{jj}	5.416e-02
6	$y_{\mu\mu j1}$	5.209e-02
7	centrality	4.729e-02
8	Ht	4.381e-02
9	p_T^{lljj}	4.040e-02
10	E_T^{miss}	3.308e-02
11	$y_{\mu\mu jj}$	2.719e-02
12	p_T^{llj1}	2.361e-02
13	p_T^{llj2}	1.923e-02
14	$y_{\mu\mu j2}$	1.602e-02

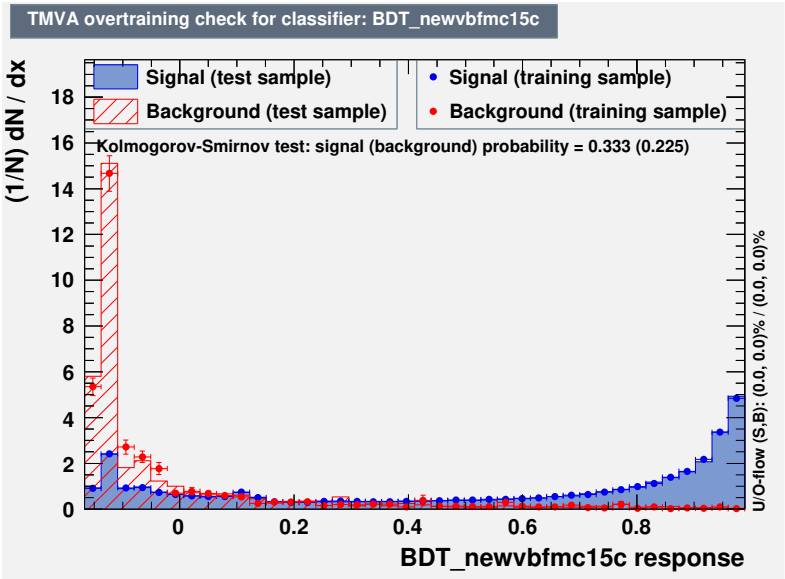
Table: Ranking in the note

MVA Output

Yidi:



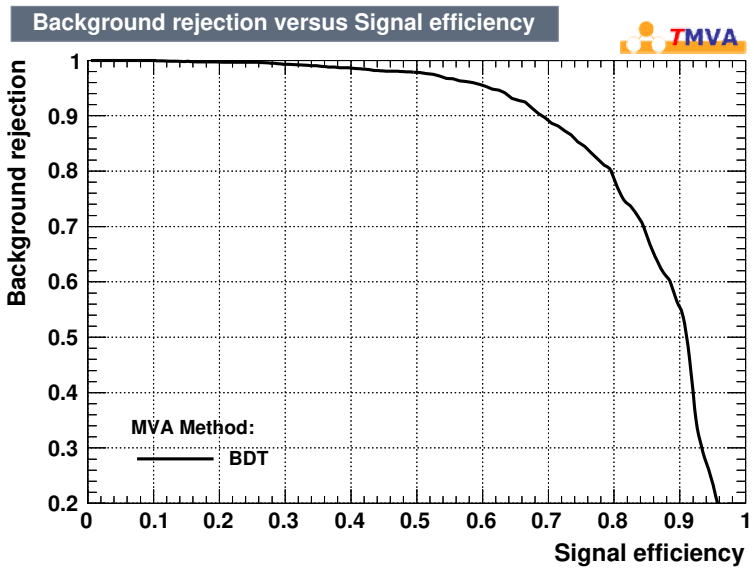
Yanlin:



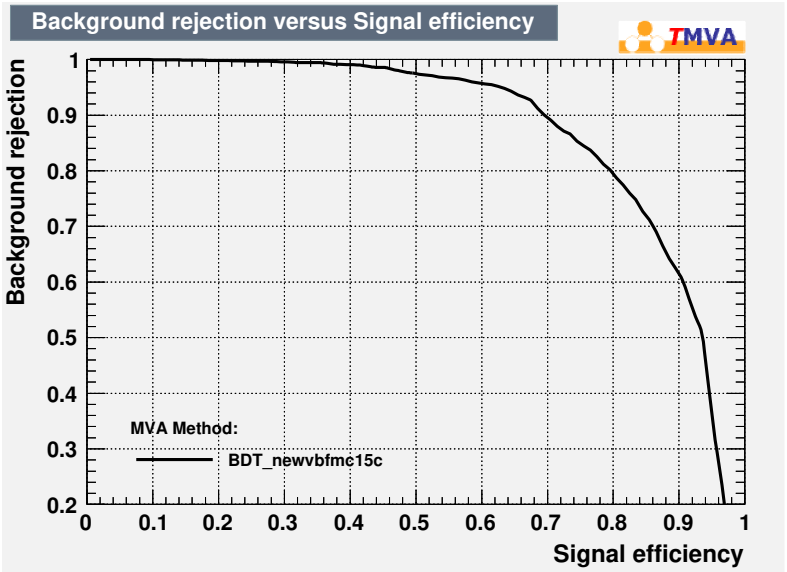
- Similar to the results in the note
- Slight differences due to the different input file

MVA Output

Yidi:



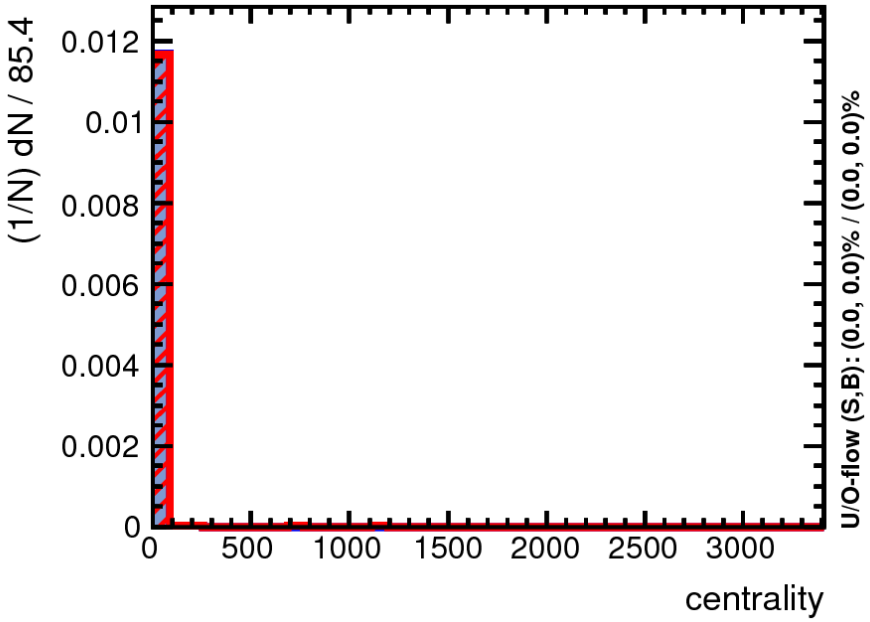
Yanlin:



- Similar to the results in the note
- Slight differences due to the different input file

Investigation on Centrality

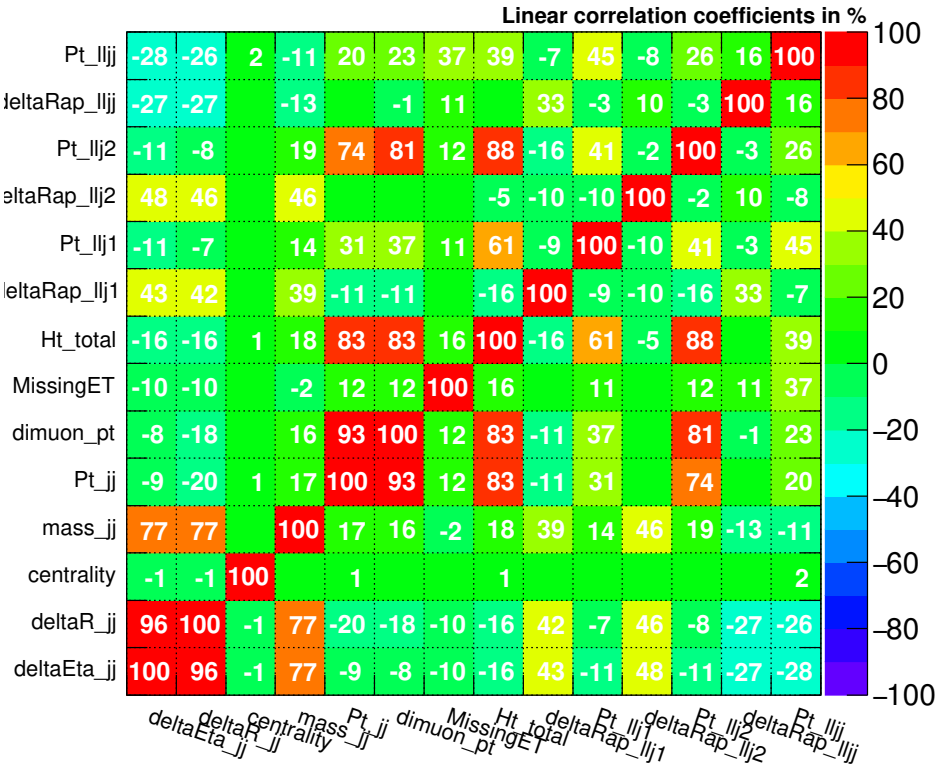
Centrality Distribution



- The distribution of centrality looks weird (supposed to be small)
- $centrality = \frac{y_{\mu\mu} - \frac{y_{j_1} + y_{j_2}}{2}}{|y_{j_1} - y_{j_2}|}$

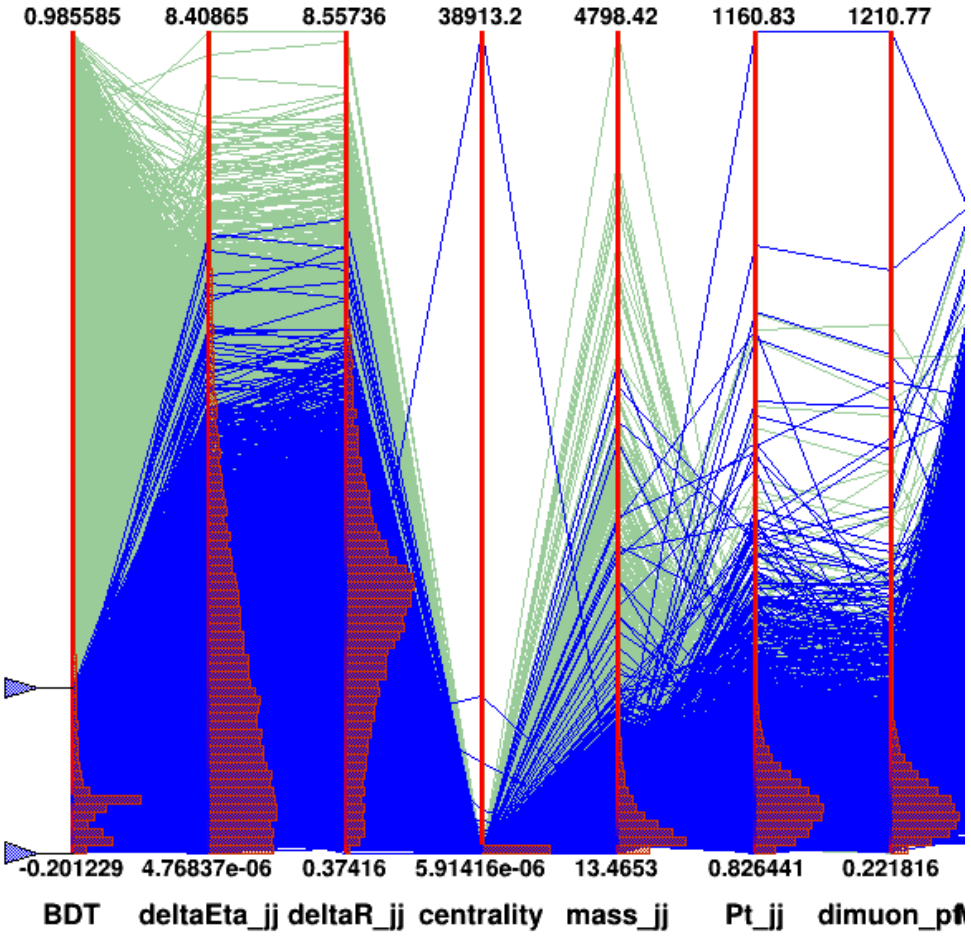
Another Look at the Correlations

Correlation Matrix (signal)



- Have another look at the correlations between centrality and other variables
- Centrality seems to be independent
- $centrality = \frac{y_{\mu\mu} - \frac{y_{j_1} + y_{j_2}}{2}}{|y_{j_1} - y_{j_2}|}$

Parallel coordinate representation



- Visualize the correlations among the classifier and the input variables
- Only a few events have really large centrality, try cuts on these events to see if there will be any effect
- Try cuts:
  centrality < 200
  centrality < 20

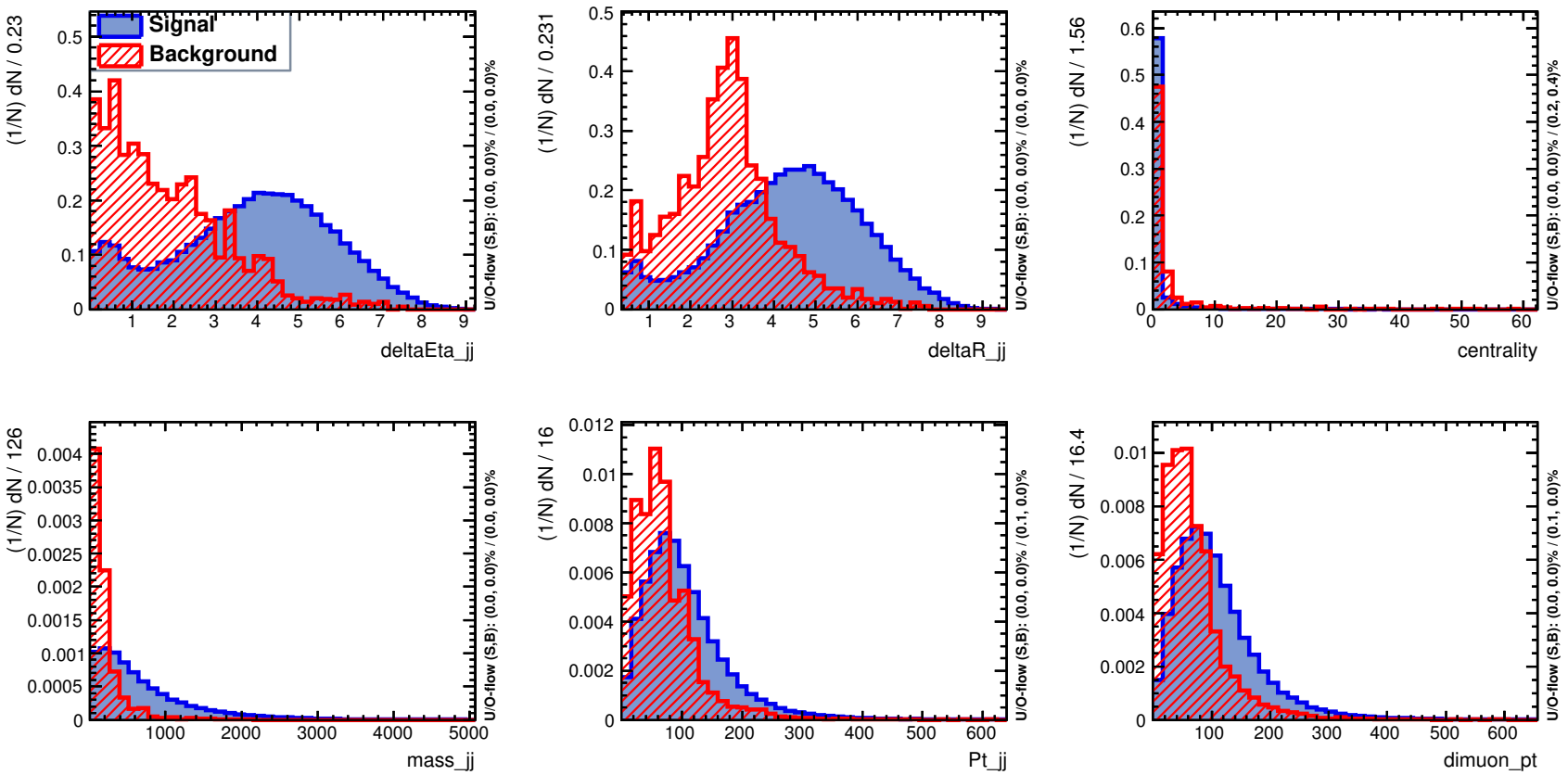


Figure: Input variables distribution with centrality < 200

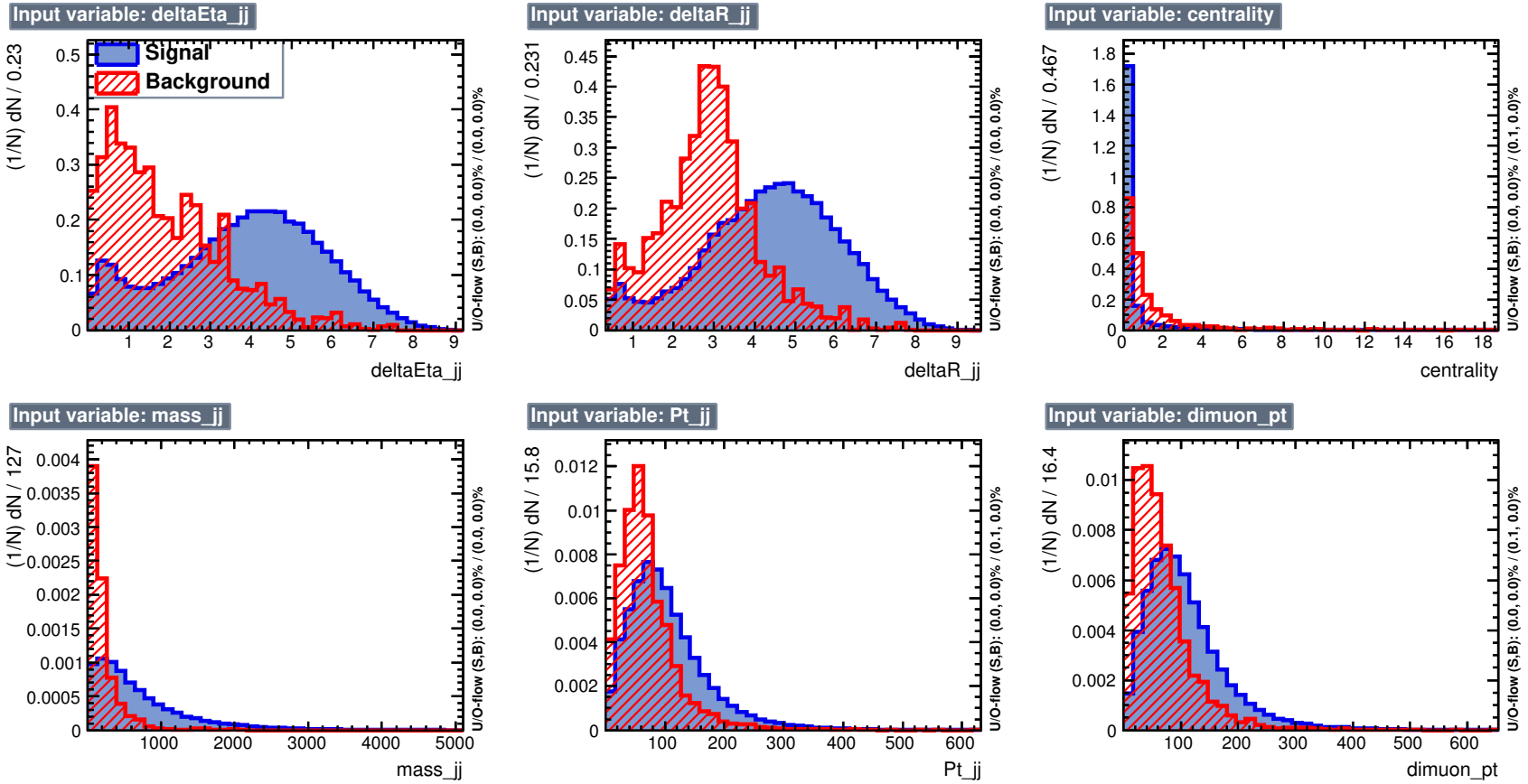


Figure: Input variables distribution with centrality < 20

Correlation Matrix

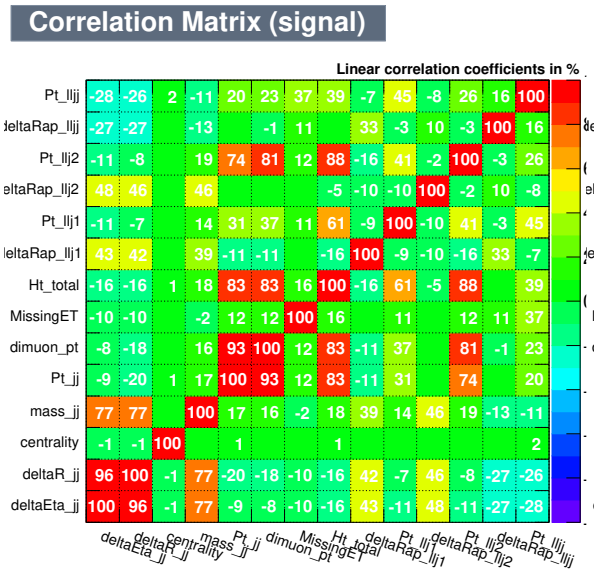


Figure: no cut

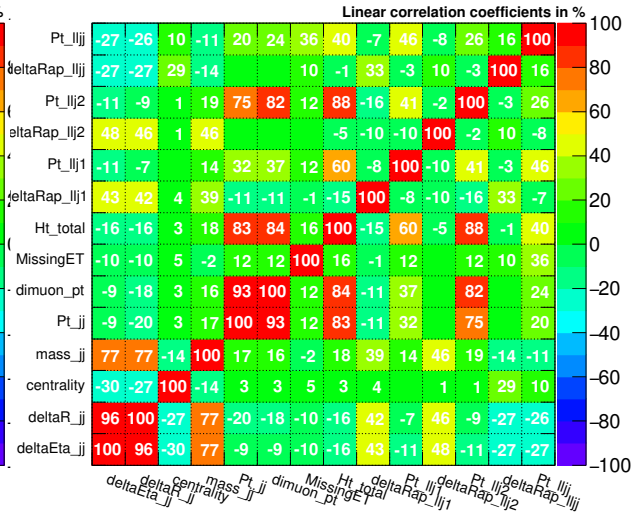


Figure: cut 200

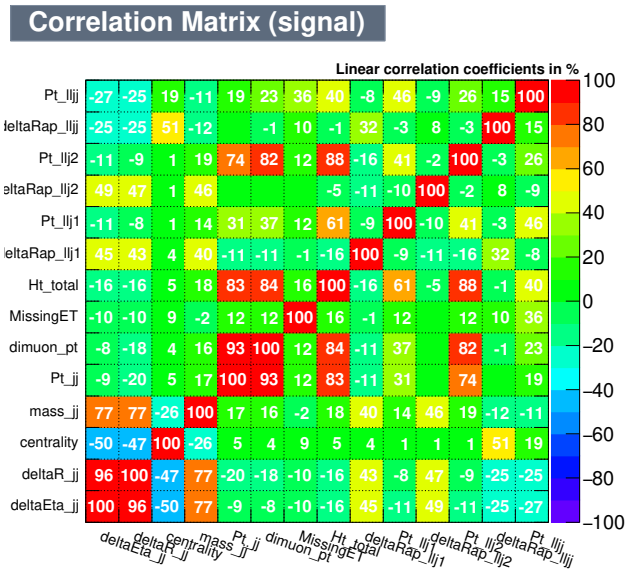
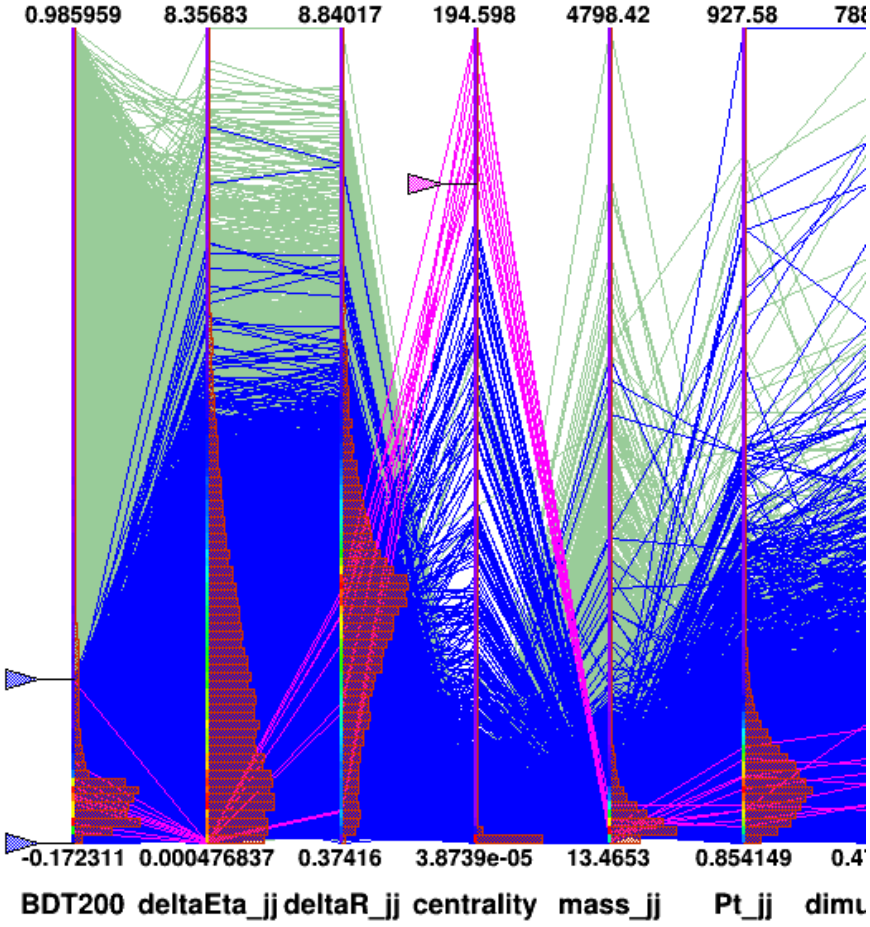


Figure: cut 20

The correlations become visible (Bottom left)

Parallel coordinate representation with Centrality < 200



- The correlations between centrality and $\Delta\eta_{jj}$, ΔR_{jj} become visible

BDT Output

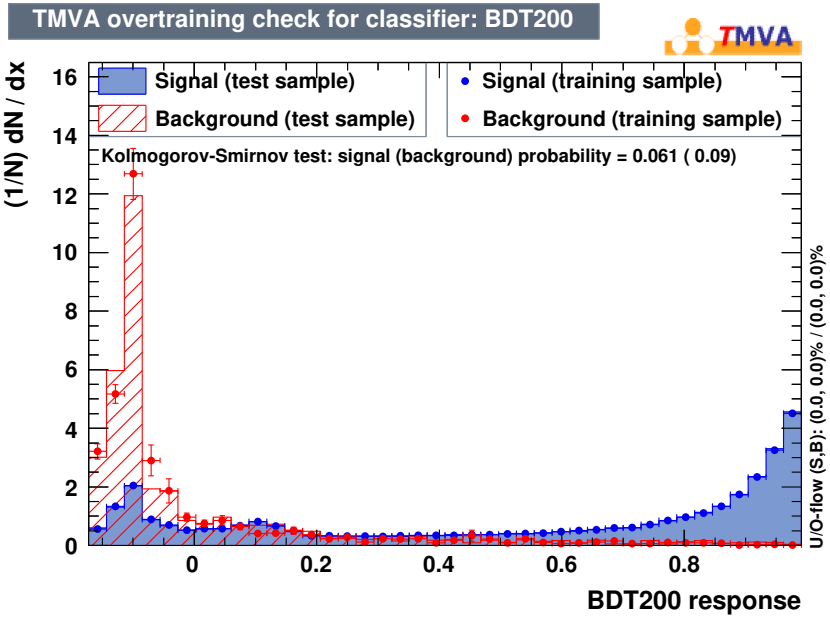


Figure: cut 200

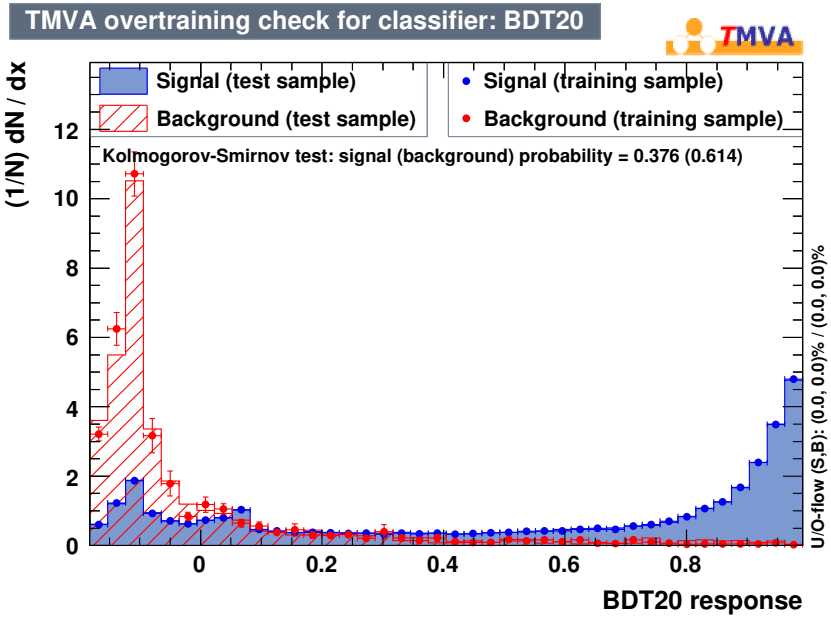
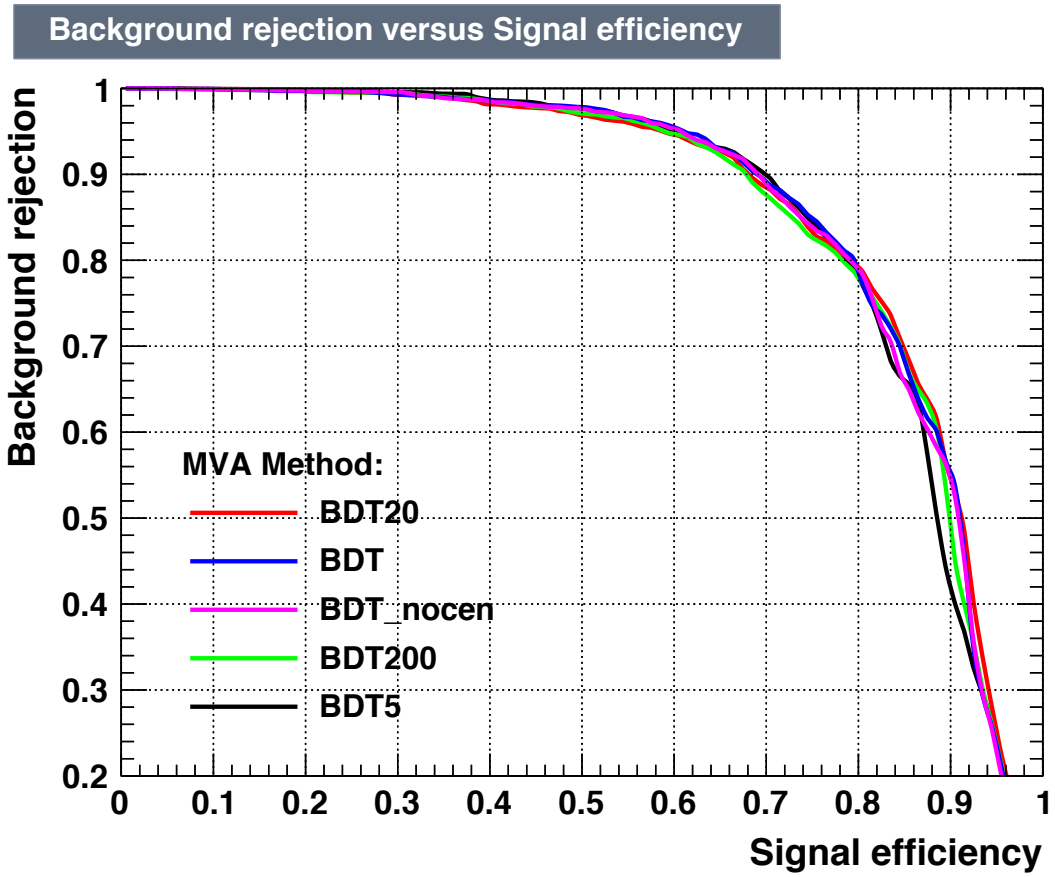


Figure: cut 20

ROC Curve



- Tried no cut on centrality, centrality < 200, 20, 5 and without centrality
- No obivous effect on the results observed

Summary

- A cross check to VBF BDT in the note has been done. The results are consistent with Yanlin's results in the note.
- The input distribution of centrality looks weird, I did some investigation and tried to give some cuts on it, but it seems to have no obvious effect on the results.

$S/\sqrt{B}$, 0.1 vs 0.11

- HighMVA Central: 0.3443 vs 0.3429
- MidMVA Central: 0.1583 vs 0.1621