

# Deleting variables

Giulia Maineri

Università degli studi di Milano

April 2022

# Table of Contents

- 1 BookMethod
- 2 All variables
- 3 No  $\vec{E}_T^{miss}$
- 4 No  $\vec{E}_T^{miss}$  significance
- 5 No  $m_{ll}$
- 6 No  $m_T$
- 7 No  $\Delta\Phi(\vec{E}_T^{miss}, \vec{p}_T^{ll\gamma})$
- 8 No  $\Delta\Phi(\vec{E}_T^{miss}, \vec{p}_T^{closest})$
- 9 No  $\Delta\Phi(\vec{E}_T^{miss}, \vec{p}_T^{closestjet})$
- 10 No  $\vec{p}_T^\gamma$
- 11 No  $\vec{p}_T^{balance}$
- 12 ROC comparisons
- 13 Variable ranking

# BookMethod

Trees training was done with the following options:

- !H; in order not to print a method-specific help message
- !V; necessary to deactivate "verbose mode", which prints explanations of what's going on
- NTrees=850; number of trees in the forest
- MinNodeSize=5%; minimum percentage of training events required in a leaf node
- MaxDepth=3; maximum depth of the decision tree allowed

- BoostType=AdaBoost; boosting type
- AdaBoostBeta=0.2; learning rate for AdaBoost algorithm
- SeparationType=GiniIndex; separation criterion for node splitting
- nCuts=20; number of grid points in variable range used in finding optimal cut in node splitting

All variables

# Classifier output distribution-Test and training sample

Figure 1: Output distribution

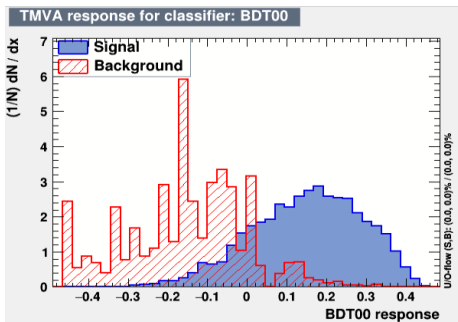


Figure 2: Overtraining control

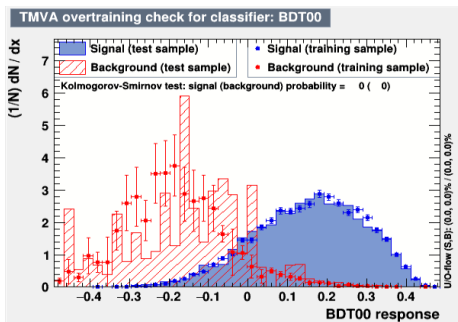


Figure 3: ROC curve

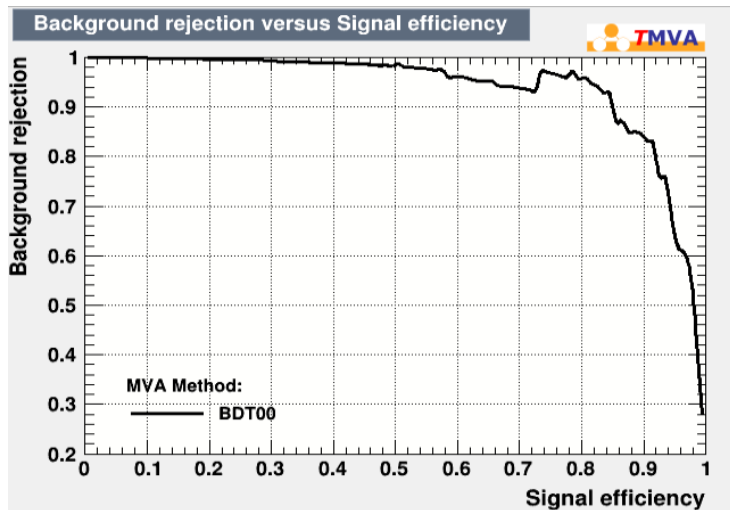
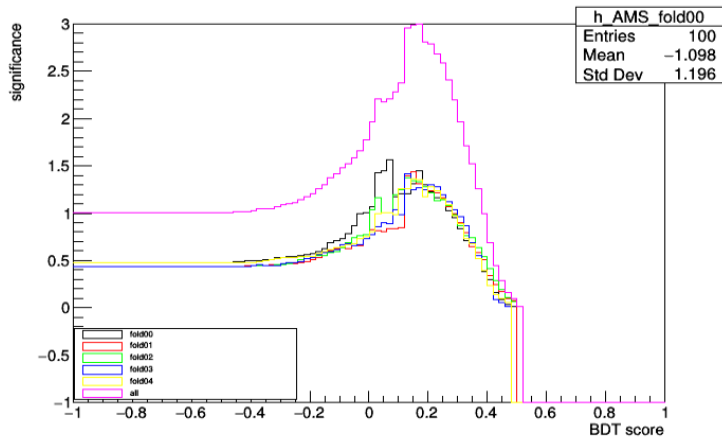




Figure 4: AMS



No  $\vec{E}_T^{miss}$

# Classifier output distribution-Test and training sample

Figure 5: Output distribution

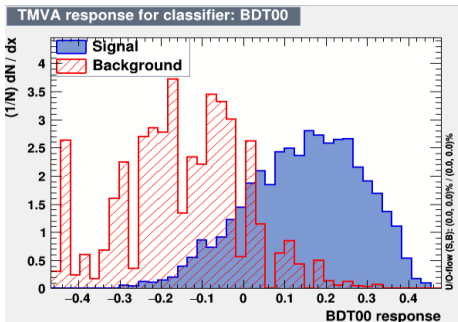


Figure 6: Overtraining control

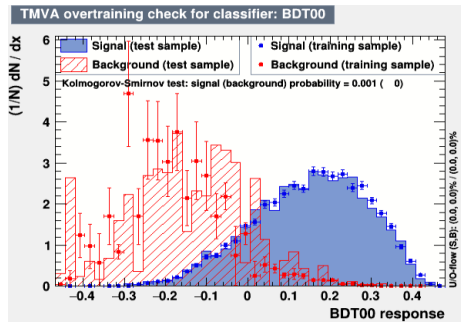


Figure 7: ROC curve

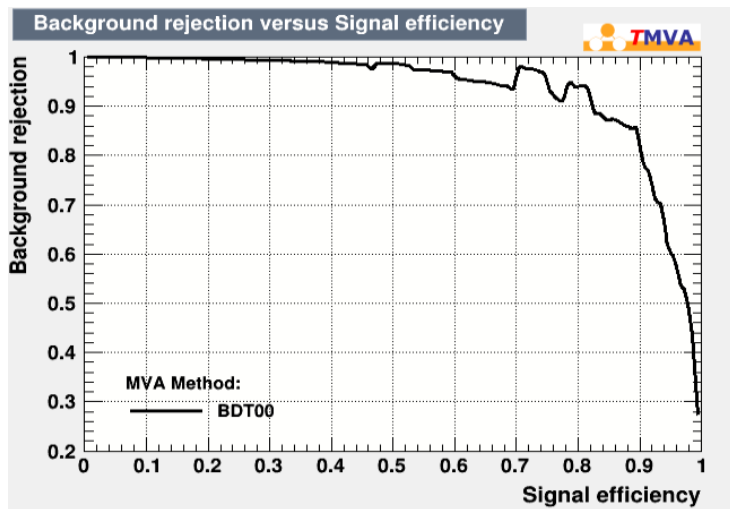
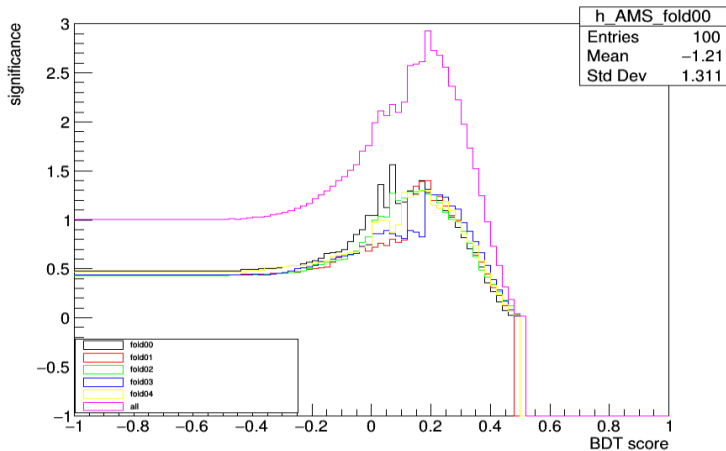


Figure 8: AMS



No  $\vec{E}_T^{miss}$  significance

# Classifier output distribution-Test and training sample

Figure 9: Output distribution

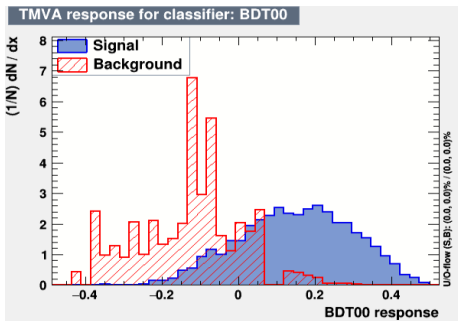


Figure 10: Overtraining control

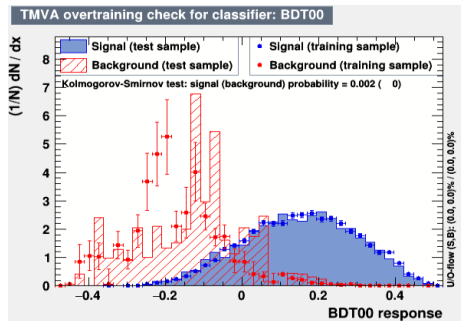


Figure 11: ROC curve

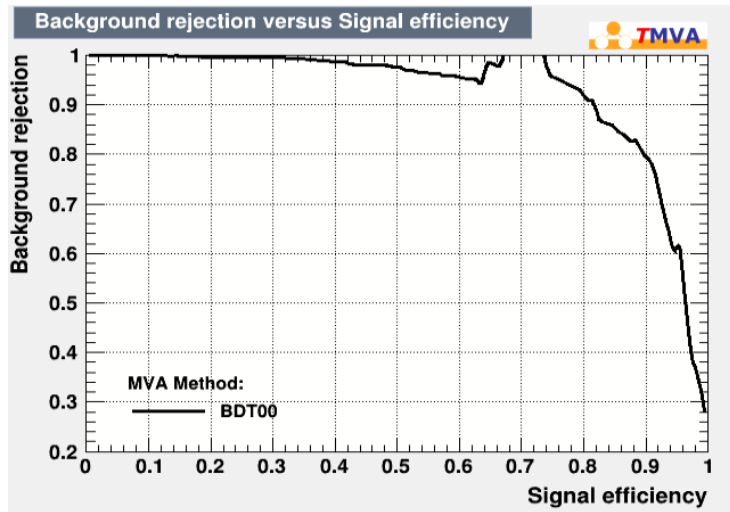
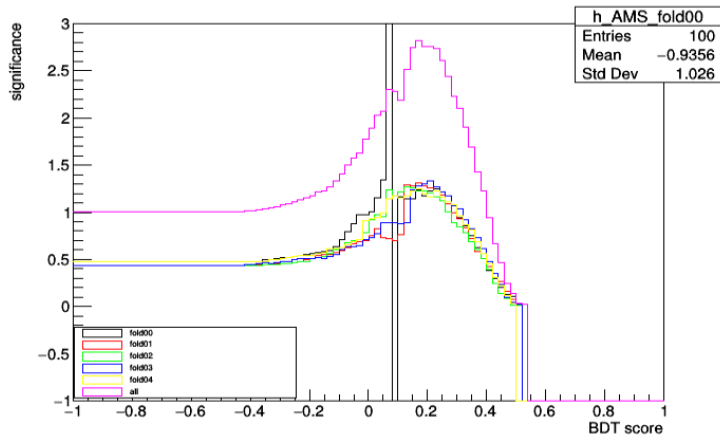




Figure 12: AMS



No  $m_{II}$

# Classifier output distribution-Test and training sample

Figure 13: Output distribution

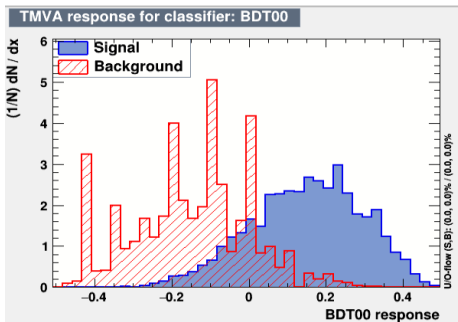


Figure 14: Overtraining control

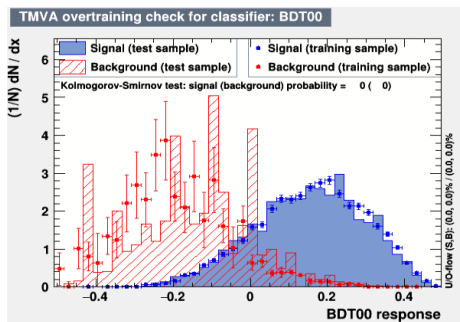


Figure 15: ROC curve

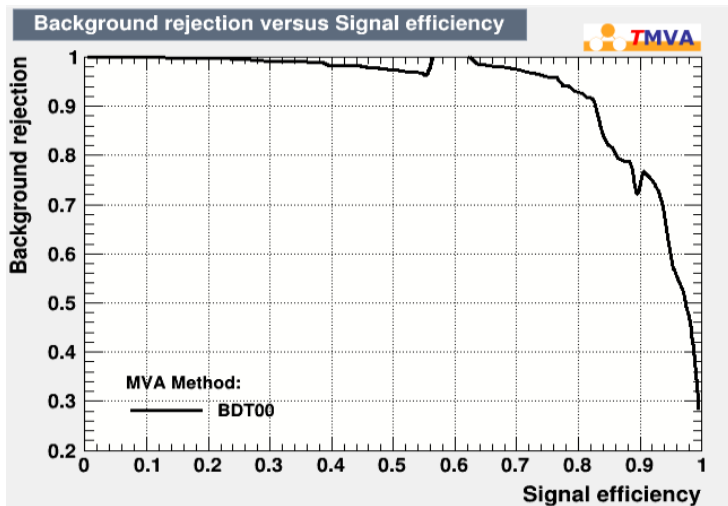
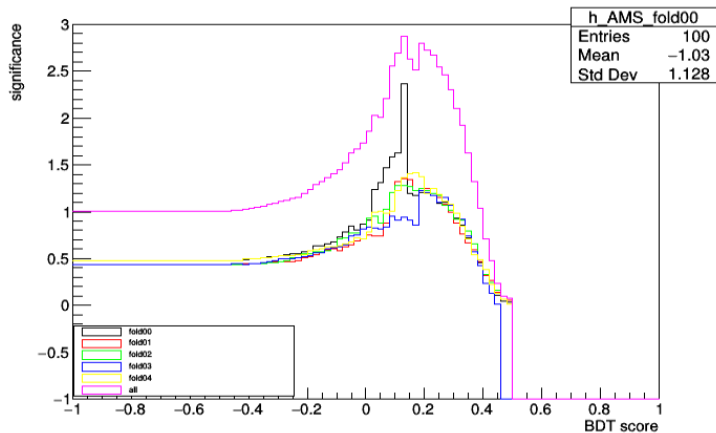


Figure 16: AMS



No  $m_T$

# Classifier output distribution-Test and training sample

Figure 17: Output distribution

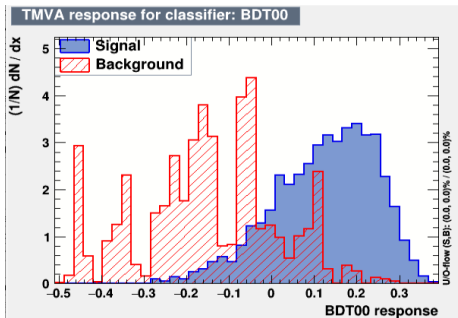


Figure 18: Overtraining control

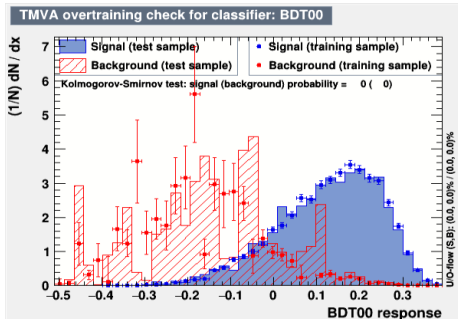


Figure 19: ROC curve

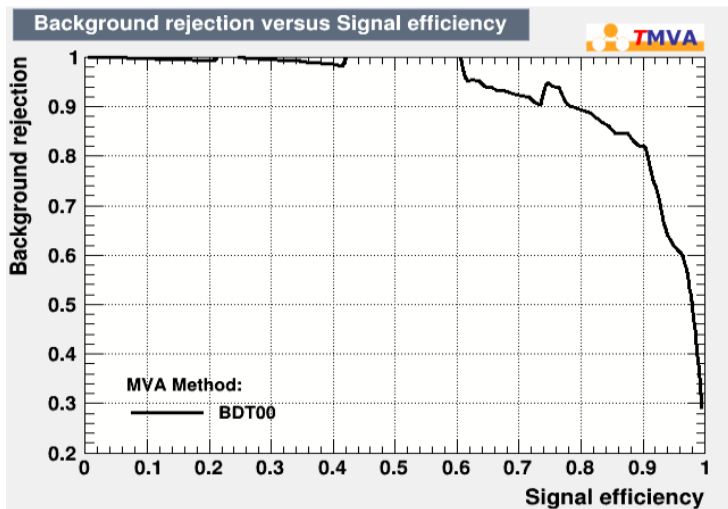
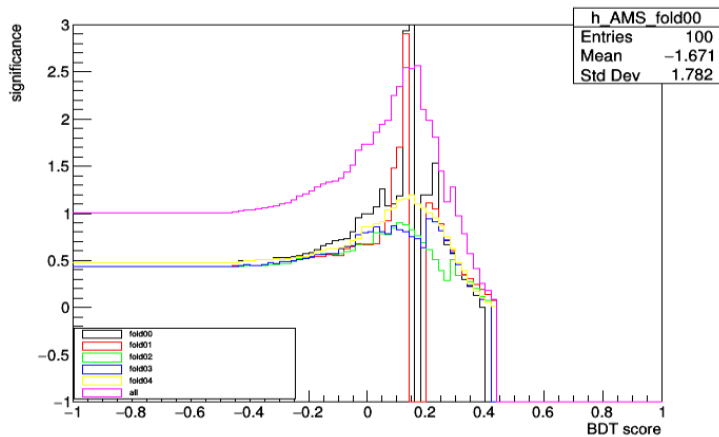




Figure 20: AMS



No  $\Delta\Phi(\vec{E}_T^{miss}, \vec{p}_T^{ll\gamma})$

# Classifier output distribution-Test and training sample

Figure 21: Output distribution

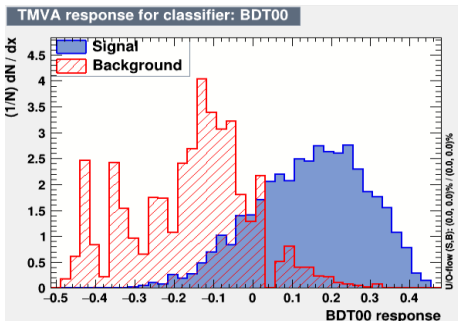


Figure 22: Overtraining control

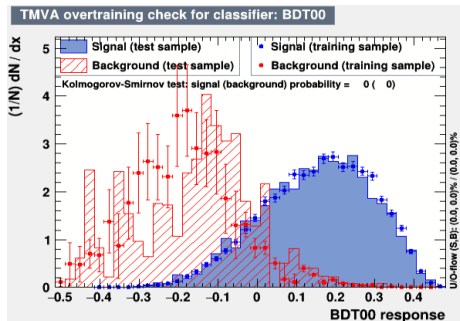


Figure 23: ROC curve

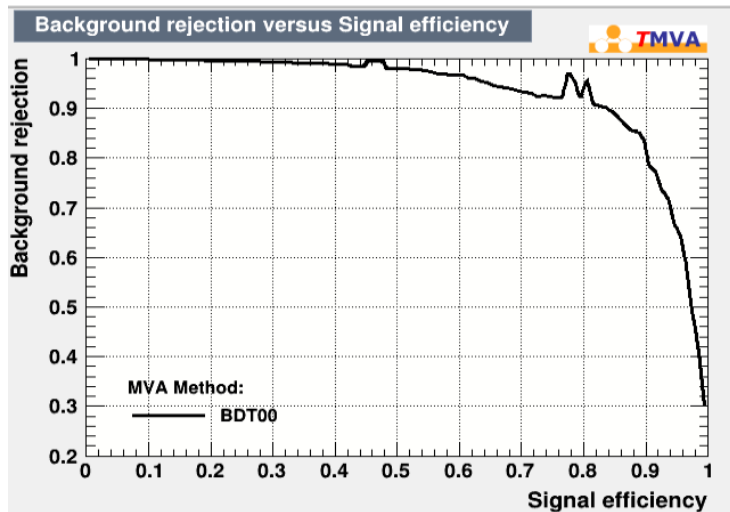
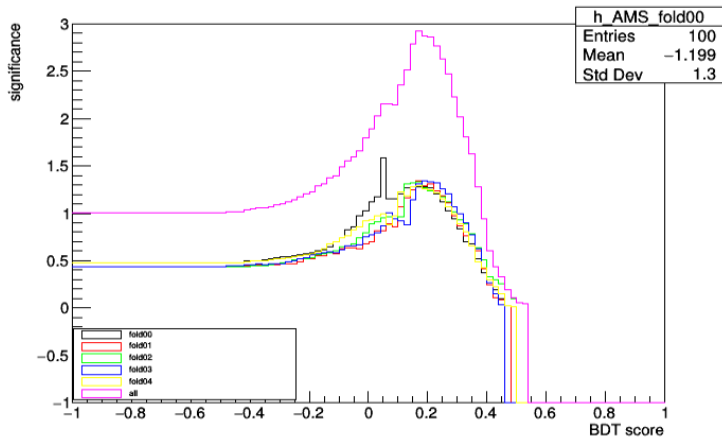


Figure 24: AMS



No  $\Delta\Phi(\vec{E}_T^{miss}, \vec{p}_T^{closest})$

# Classifier output distribution-Test and training sample

Figure 25: Output distribution

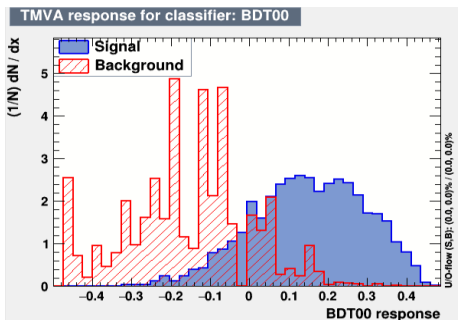


Figure 26: Overtraining control

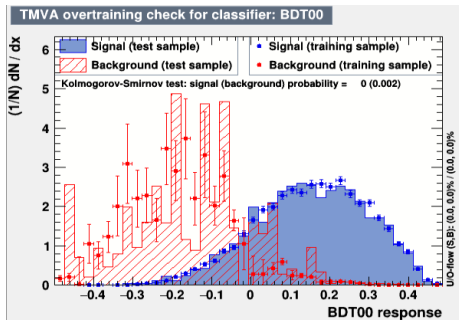


Figure 27: ROC curve

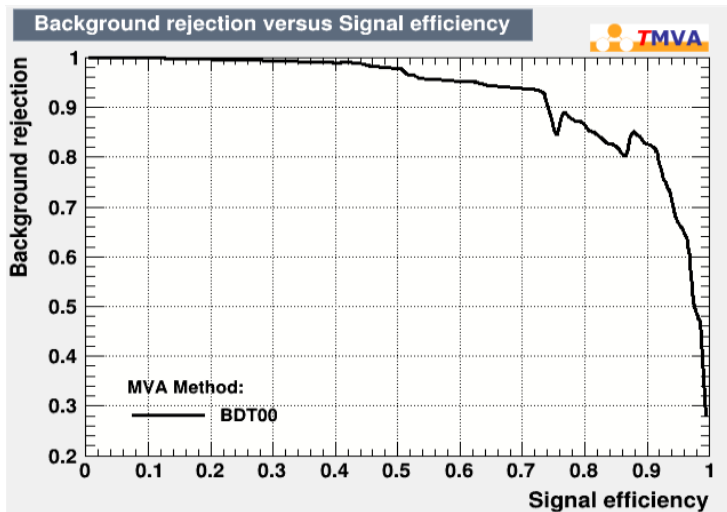
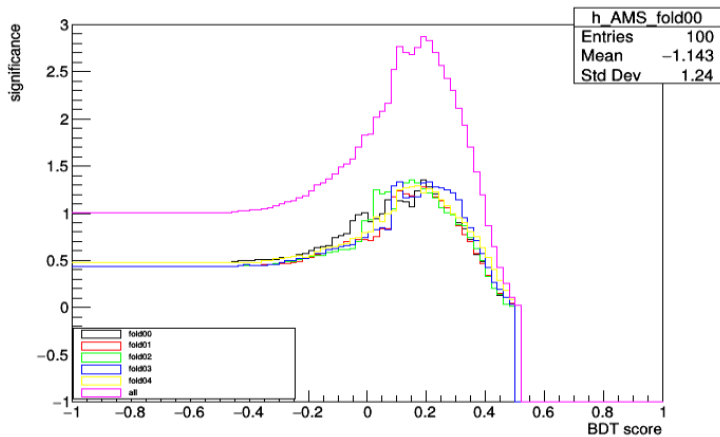




Figure 28: AMS



No  $\Delta\Phi(\vec{E}_T^{miss}, \vec{p}_T^{closestjet})$

# Classifier output distribution-Test and training sample

Figure 29: Output distribution

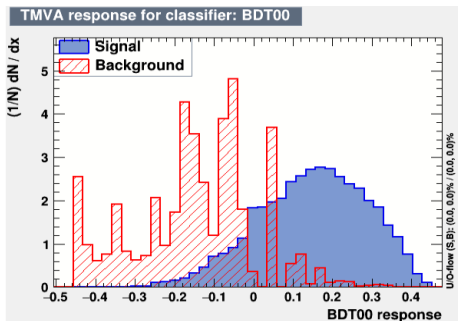


Figure 30: Overtraining control

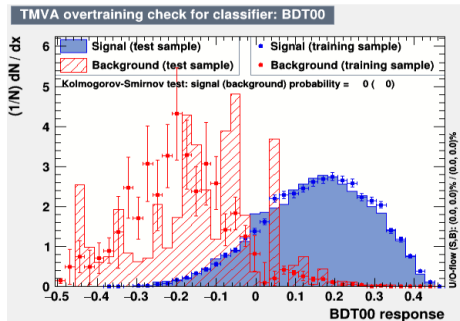


Figure 31: ROC curve

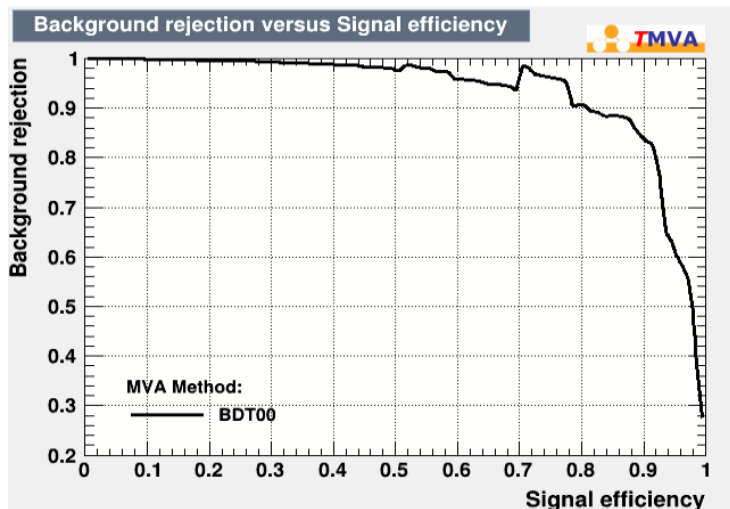
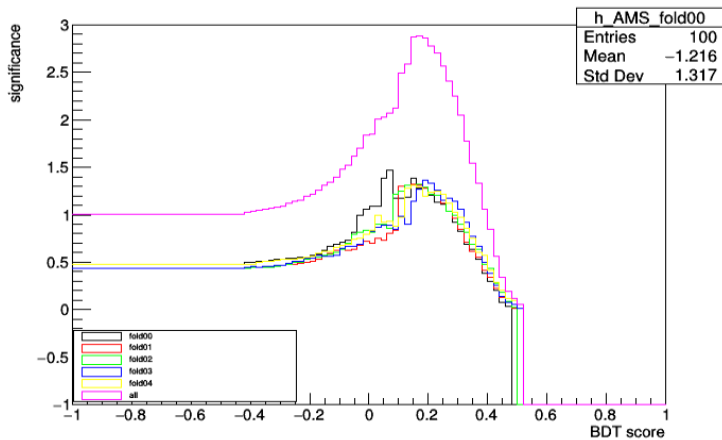


Figure 32: AMS



No  $\vec{p}_T^\gamma$

# Classifier output distribution-Test and training sample

Figure 33: Output distribution

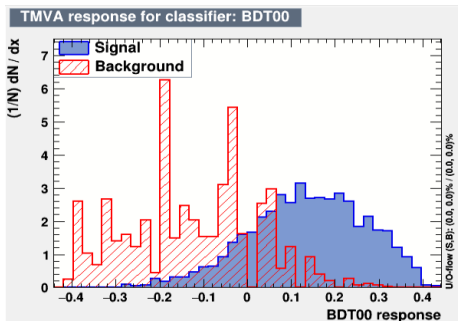


Figure 34: Overtraining control

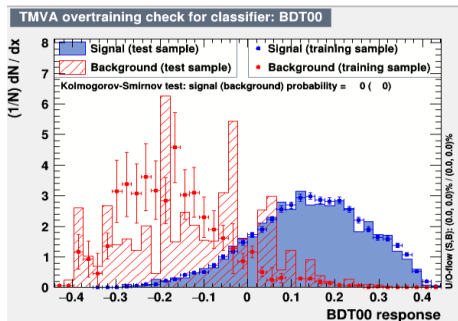


Figure 35: ROC curve

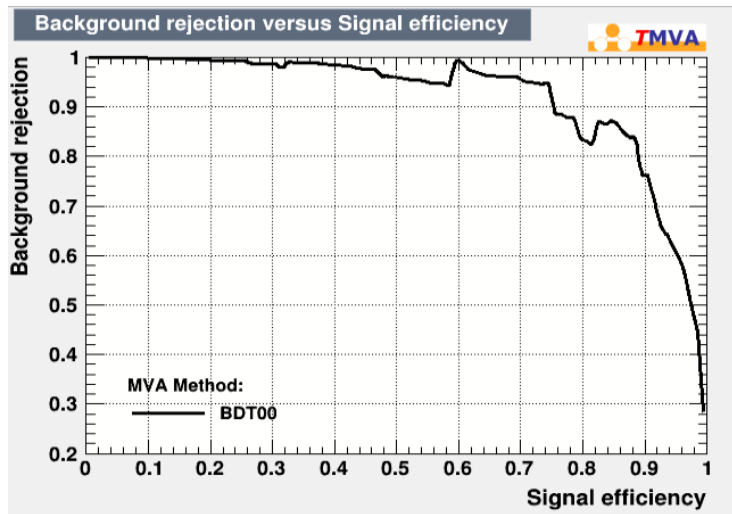
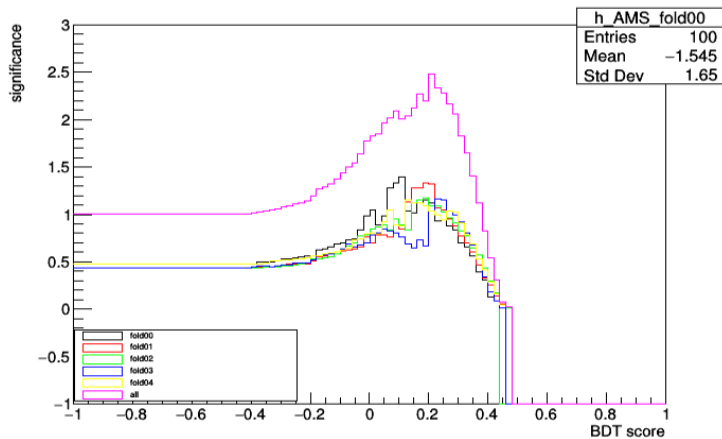




Figure 36: AMS



No  $\vec{p}_T^{balance}$

# Classifier output distribution-Test and training sample

Figure 37: Output distribution

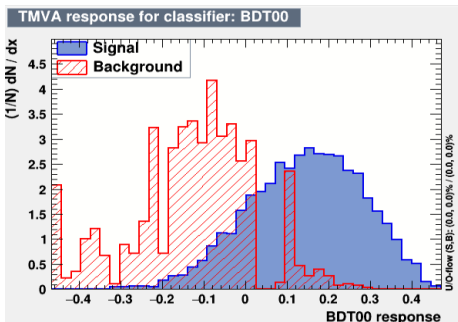


Figure 38: Overtraining control

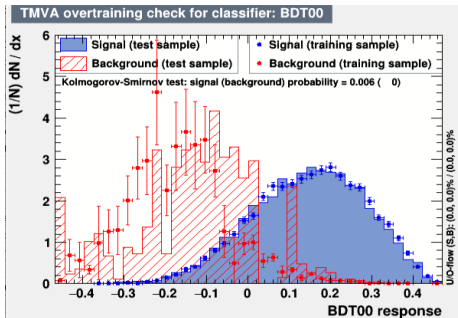


Figure 39: ROC curve

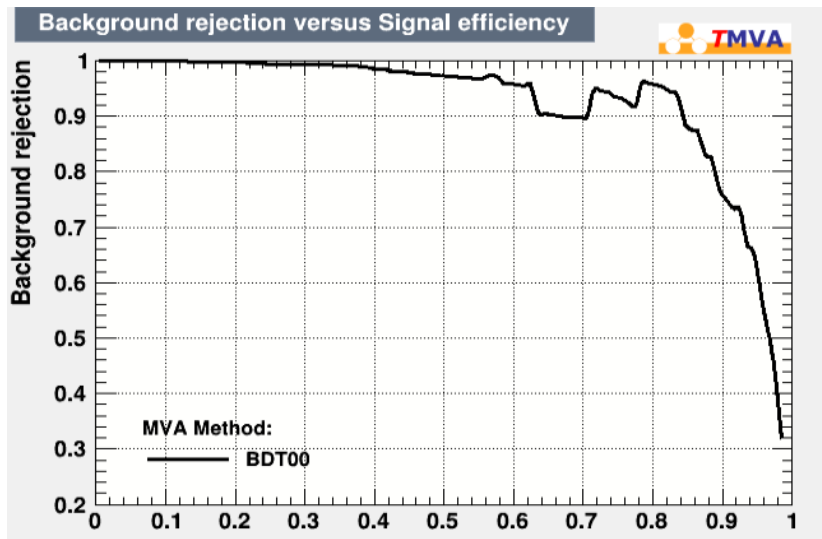
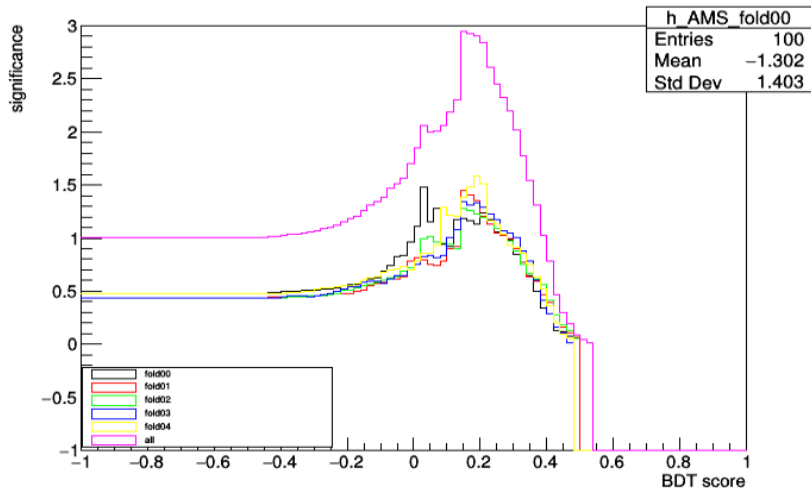


Figure 40: AMS



## ROC comparisons

Figure 41: ROC Comparison

Comparison

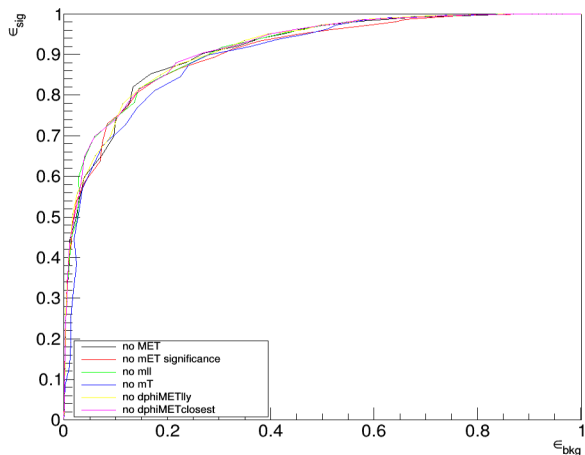


Figure 42: ROC Comparison

Comparison

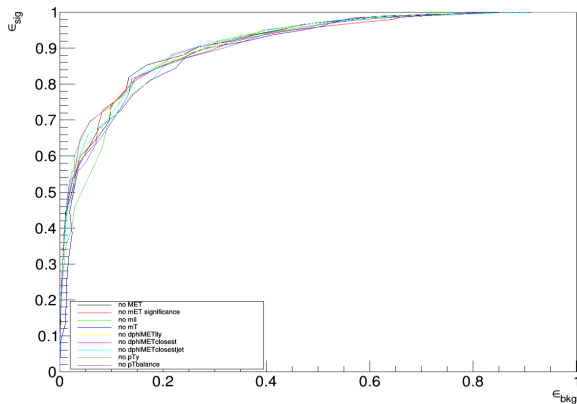
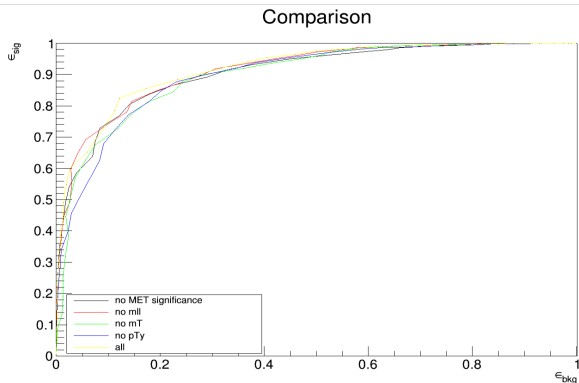




Figure 43: ROC Comparison - Important variables



## Variable ranking

<i>deleted variable</i>	<i>AMS</i>	$\Delta$ <i>AMS</i>
<i>none</i>	2.99522	0
$E_T^{miss}$	2.93307	0.06215
$sig E_T^{miss}$	2.8162	0.17902
$m_{ll}$	2.87348	0.12174
$m_T$	2.5651	0.39912
$\Delta\Phi(\vec{E}_T^{miss}, \vec{p}_T^{ll\gamma})$	2.92392	0.0713
$\Delta\Phi(\vec{E}_T^{miss}, \vec{p}_T^{closest})$	2.87644	0.11878
$\Delta\Phi(\vec{E}_T^{miss}, \vec{p}_T^{closestjet})$	2.88721	0.10801
$\vec{p}_T^\gamma$	2.48571	0.50951
$\vec{p}_T^{balance}$	2.94389	0.05133

Table 1: AMS values deleting variables one a time

# Variable ranking

<i>deleted variable</i>	<i>AMS</i>	$\Delta$ <i>AMS</i>
<i>none</i>	2.99522	0
$\vec{p}_T^\gamma$	2.48571	0.50951
$m_T$	2.56510	0.39912
$\text{sig}E_T^{\text{miss}}$	2.81620	0.17902
$m_{ll}$	2.87348	0.12174
$\Delta\Phi(\vec{E}_T^{\text{miss}}, \vec{p}_T^{\text{closest}})$	2.87644	0.11878
$\Delta\Phi(\vec{E}_T^{\text{miss}}, \vec{p}_T^{\text{closestjet}})$	2.88721	0.10801
$\Delta\Phi(\vec{E}_T^{\text{miss}}, \vec{p}_T^{\text{ll}\gamma})$	2.92392	0.07130
$E_T^{\text{miss}}$	2.93307	0.06215
$\vec{p}_T^{\text{balance}}$	2.94389	0.05133

Table 2: Variable ranking

# Counts comparison

This is a list showing how many times the algorithm chose to use one variable as discriminating variable during the training of the trees.

<i>variable</i>	<i>num</i>
$\vec{p}_T^\gamma$	360
$m_T$	364
$\text{sig}E_T^{\text{miss}}$	434
$m_{ll}$	531
$\Delta\Phi(\vec{E}_T^{\text{miss}}, \vec{p}_T^{\text{closest}})$	868
$\Delta\Phi(\vec{E}_T^{\text{miss}}, \vec{p}_T^{\text{closestjet}})$	232
$\Delta\Phi(\vec{E}_T^{\text{miss}}, \vec{p}_T^{ll\gamma})$	747
$E_T^{\text{miss}}$	134
$\vec{p}_T^{\text{balance}}$	187

Table 3: Variable ranking

<i>variable</i>	<i>num</i>
$\Delta\Phi(\vec{E}_T^{miss}, \vec{p}_T^{closest})$	868
$\Delta\Phi(\vec{E}_T^{miss}, \vec{p}_T^{l\gamma})$	747
$m_{ll}$	531
$sigE_T^{miss}$	434
$m_T$	364
$\vec{p}_T^\gamma$	360
$\Delta\Phi(\vec{E}_T^{miss}, \vec{p}_T^{closestjet})$	232
$\vec{p}_T^{balance}$	187
$E_T^{miss}$	134

Table 4: Variable ranking

Figure 44: Variable ranking - all variables

Rank	Variable	Variable Importance
1	METsig	1.349e-01
2	pTy	1.332e-01
3	dphi_MET_lly	1.277e-01
4	mT	1.273e-01
5	dphi_MET_closest	1.227e-01
6	pTbalance	1.139e-01
7	mll	1.065e-01
8	MET	8.021e-02
9	dphi_MET_closestjet	5.353e-02

Figure 45: Variable ranking - all variables

Rank	Variable	Variable Importance
1	mT	1.414e-01
2	METsig	1.301e-01
3	pTbalance	1.171e-01
4	dphi_MET_ll	1.143e-01
5	dphi_MET_closest	1.131e-01
6	pTy	1.130e-01
7	mll	1.056e-01
8	MET	8.890e-02
9	dphi_MET_closestjet	7.656e-02



Figure 46: Variable ranking - all variables

```
-----  
Rank : Variable           : Variable Importance  
-----  
 1 : mT                   : 1.477e-01  
 2 : pTy                  : 1.346e-01  
 3 : METsig               : 1.345e-01  
 4 : dphi_MET_closest    : 1.125e-01  
 5 : pTbalance           : 1.087e-01  
 6 : dphi_MET_lly        : 1.035e-01  
 7 : mll                  : 9.560e-02  
 8 : MET                  : 9.116e-02  
 9 : dphi_MET_closestjet : 7.178e-02  
-----
```

Figure 47: Variable ranking - all variables

```
-----  
Rank : Variable           : Variable Importance  
-----  
  1 : METsig              : 1.376e-01  
  2 : mT                  : 1.344e-01  
  3 : pTy                 : 1.267e-01  
  4 : pTbalance           : 1.261e-01  
  5 : dphi_MET_closest   : 1.072e-01  
  6 : dphi_MET_lly       : 1.053e-01  
  7 : mll                 : 1.015e-01  
  8 : MET                 : 8.384e-02  
  9 : dphi_MET_closestjet : 7.743e-02  
-----
```

Figure 48: Variable ranking - all variables

```
-----  
Rank : Variable                : Variable Importance  
-----  
1 : mT                          : 1.378e-01  
2 : METsig                       : 1.344e-01  
3 : pTbalance                    : 1.243e-01  
4 : dphi_MET_closest            : 1.146e-01  
5 : dphi_MET_lly                : 1.115e-01  
6 : pTy                          : 1.096e-01  
7 : mll                          : 1.007e-01  
8 : dphi_MET_closestjet         : 8.592e-02  
9 : MET                          : 8.126e-02  
-----
```

- Some variables, such as  $\vec{E}_T^{miss}$  and  $\Delta\Phi(\vec{E}_T^{miss}, \vec{p}_T^{closestjet})$ , appear at the bottom of the ranking, independently by the method used to do it.
- In the same way,  $m_T$  and  $\vec{E}_T^{miss}$  significance always appear in the first places.
- In the two first variable rankings, made deleting variables one a time and counting how many times each variable was used during the training,  $m_T$  and  $p_T^\gamma$  appear near in the ranking. These variables have a significant correlation in the background.