Automatic MVA Evaluation

Thomas Keck
Moritz Gelb
Nils Braun

April 8, 2023

Abstract

Evaluation plots

Contents

1 Classifiers **3**
1.1 /home/belle2/ssana/MC15ri_cs1/cs/test/MVAFastBDT.root . . . 3

2 Variables **3**
2.1 Importance . . . 4
2.2 Correlation . . . 4
2.3 KSFWVariables(hoo0) . . . 5
2.4 CleoConeCS(1) . . . 5
2.5 KSFWVariables(et) . . . 6
2.6 thrustOm . . . 6
2.7 KSFWVariables(hso10) . . . 7
2.8 cosTBz . . . 7
2.9 KSFWVariables(hso02) . . . 8
2.10 KSFWVariables(hso12) . . . 8
2.11 CMS_cosTheta . . . 9
2.12 thrustBm . . . 9
2.13 cosTBTO . . . 10
2.14 abs_qr . . . 10
2.15 DeltaZ . . . 11
2.16 R2 . . . 11

3 Classifier Plot **11**

4 ROC Plot **12**

5 Classification Results **13**
5.1 /home . . . 13

6 Diagonal Plot **13**
6.1 /home . . . 13
6.2 Overtraining Plot . . . 14

7 Spectators **14**

1 Classifiers

This section contains the GeneralOptions and SpecificOptions of all classifiers represented by an XML tree. The same information can be retrieved using the basf2_mva_info tool.

Table 1: Abbreviations of identifiers

Identifier	Abbreviation
/home/belle2/ssana/MC15ri_cs1/cs/test/MVAFastBDT.root	/home

1.1 /home/belle2/ssana/MC15ri_cs1/cs/test/MVAFastBDT.root


```
<?xml version="1.0" encoding="utf-8"?>
<method>FastBDT</method>
<weightfile>/home/belle2/ssana/MC15ri_cs1/cs/test/MVAFastBDT.root</weightfile>
<treename>tree</treename>
<target_variable>isSignal</target_variable>
<weight_variable>__weight__</weight_variable>
<signal_class>1</signal_class>
<max_events>0</max_events>
<number_feature_variables>14</number_feature_variables>
<variable0>abs_qr</variable0>
<variable1>DeltaZ</variable1>
<variable2>R2</variable2>
<variable3>thrustBm</variable3>
<variable4>thrustOm</variable4>
<variable5>cosTBTO</variable5>
<variable6>cosTBz</variable6>
<variable7>CMS_cosTheta</variable7>
<variable8>KSFWVariables(et)</variable8>
<variable9>KSFWVariables(hso02)</variable9>
<variable10>KSFWVariables(hso10)</variable10>
<variable11>KSFWVariables(hso12)</variable11>
<variable12>KSFWVariables(hoo0)</variable12>
<variable13>CleoConeCS(1)</variable13>
<number_spectator_variables>0</number_spectator_variables>
<number_data_files>1</number_data_files>
<datafile0>/home/belle2/ssana/MC15ri_cs1/cs/train/signal_scaled/train.root</datafile0>
<FastBDT_version>2</FastBDT_version>
<FastBDT_nTrees>200</FastBDT_nTrees>
<FastBDT_nCuts>8</FastBDT_nCuts>
<FastBDT_nLevels>3</FastBDT_nLevels>
<FastBDT_shrinkage>0.10000000000000001</FastBDT_shrinkage>
<FastBDT_randRatio>0.5</FastBDT_randRatio>
<FastBDT_flatnessLoss>-1</FastBDT_flatnessLoss>
<FastBDT_sPlot>false</FastBDT_sPlot>
<FastBDT_number_individual_nCuts>0</FastBDT_number_individual_nCuts>
<FastBDT_purityTransformation>false</FastBDT_purityTransformation>
<FastBDT_number_individualPurityTransformation>0</FastBDT_number_individualPurityTransformation>
```


2 Variables

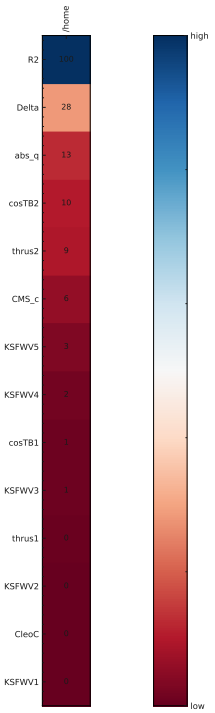
This section contains an overview of the importance and correlation of the variables used by the classifiers. And distribution plots of the variables on the independent dataset. The distributions are normed for signal and background separately, and only the region +- 3 sigma around the mean is shown.

Table 2: Abbreviations of variables

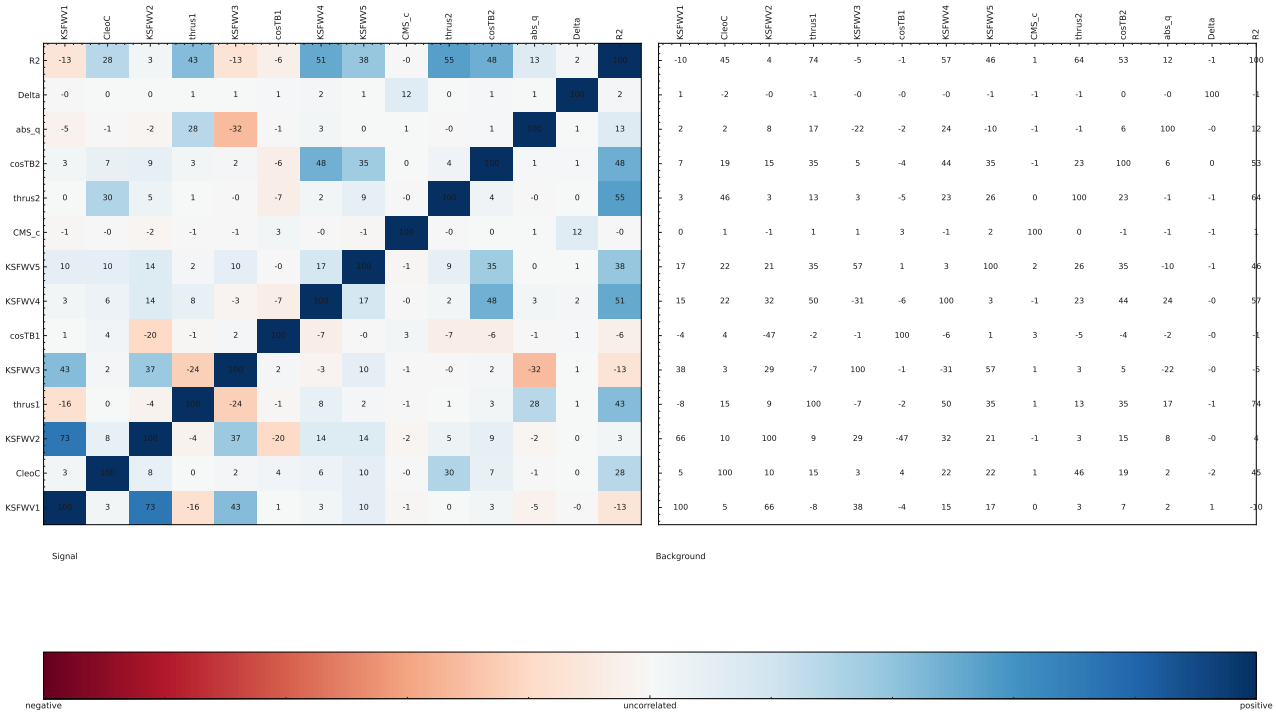
Variable	Abbreviation
KSFWVariables(hoo0)	KSFWV1
CleoConeCS(1)	CleoC
KSFWVariables(et)	KSFWV2
thrustOm	thrus1
KSFWVariables(hso10)	KSFWV3
cosTBz	cosTB1
KSFWVariables(hso02)	KSFWV4
KSFWVariables(hso12)	KSFWV5

CMS_cosTheta	CMS_c
thrustBm	thrus2
cosTBTO	cosTB2
abs_qr	abs_q
DeltaZ	Delta
R2	R2

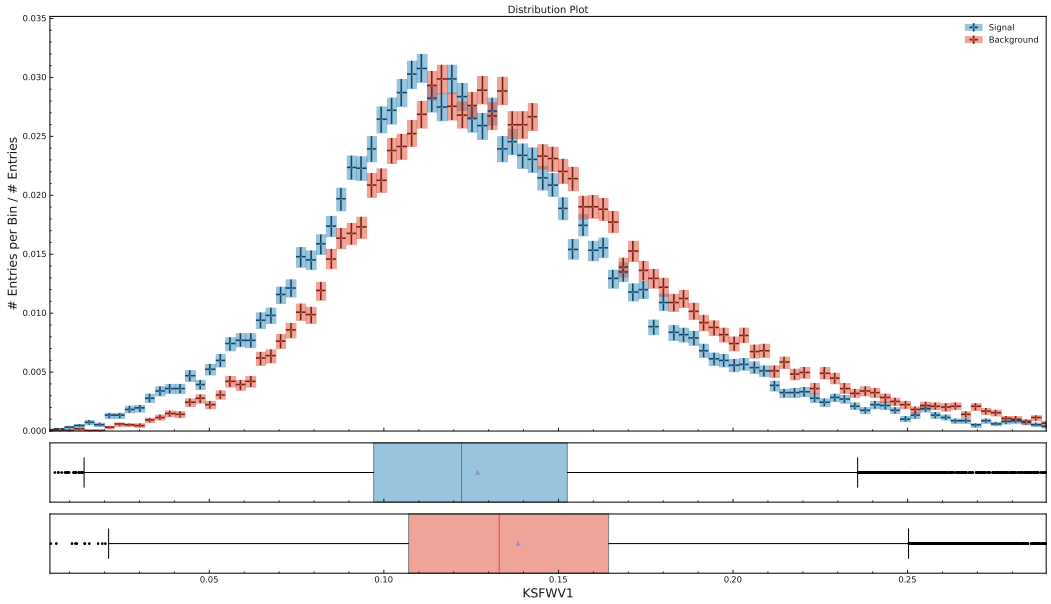
2.1 Importance



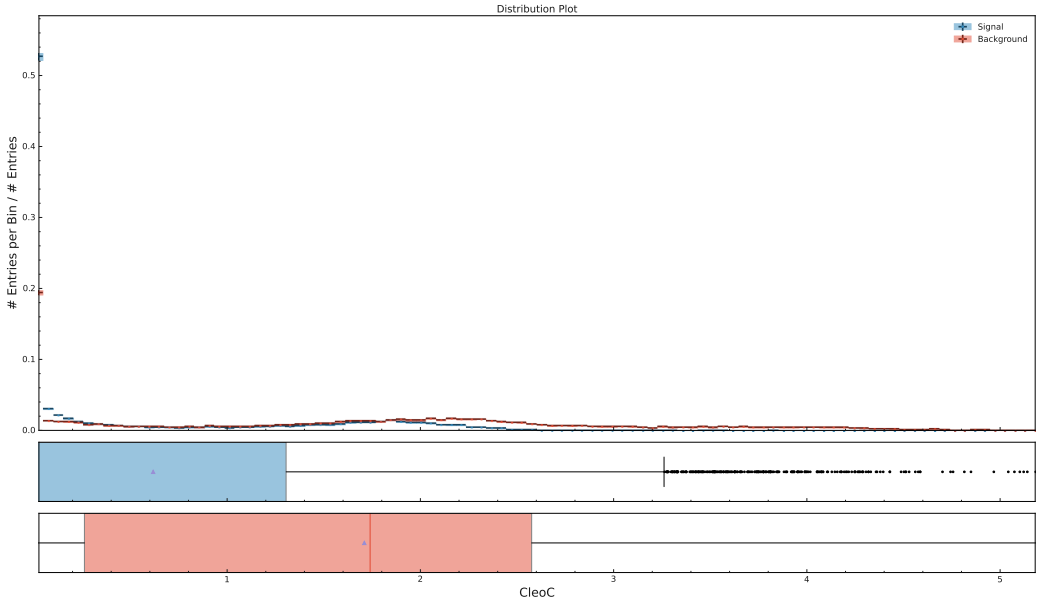
2.2 Correlation



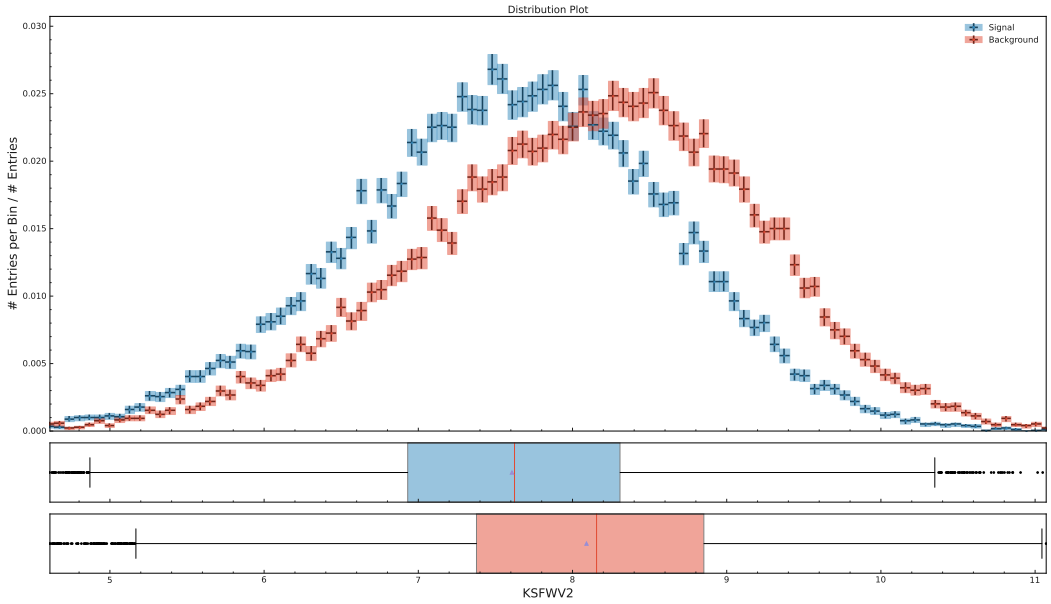
2.3 KSFWVariables(hoo0)



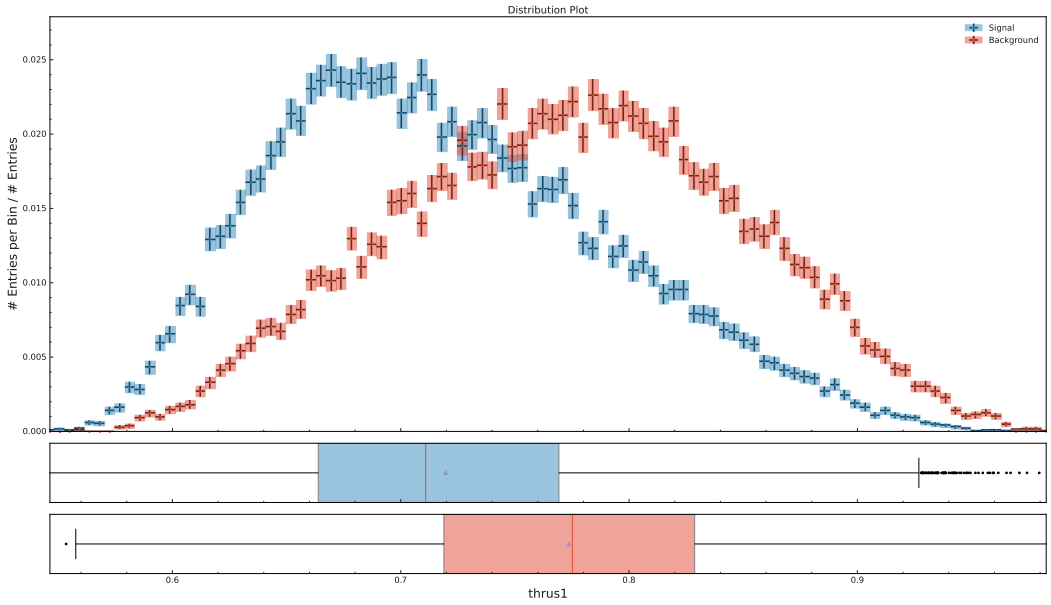
2.4 CleoConeCS(1)



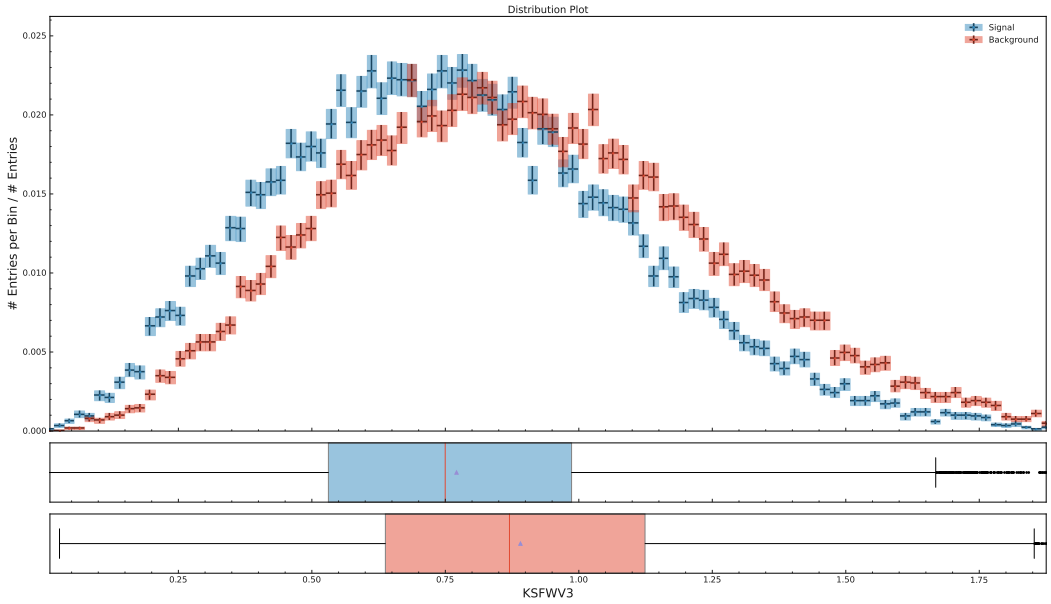
2.5 KSFWVariables(et)



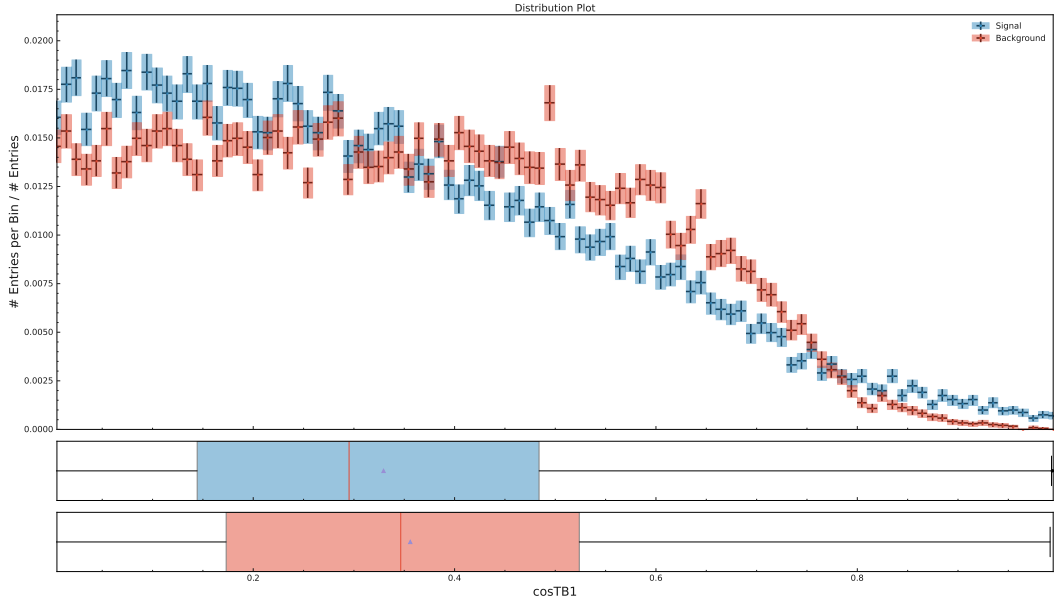
2.6 thrustOm



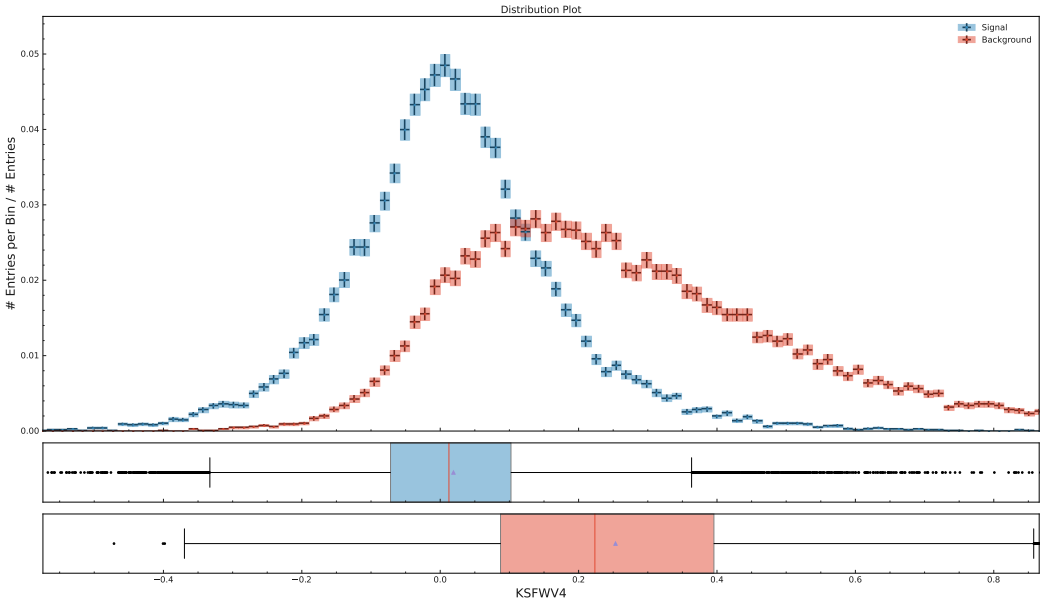
2.7 KSFWVariables(hso10)



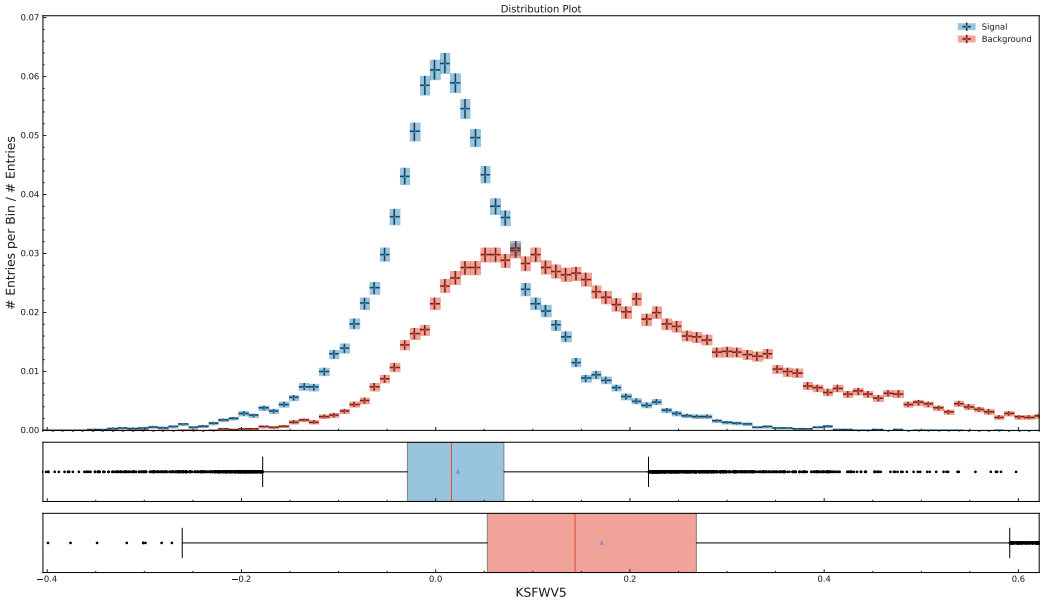
2.8 cosTBz



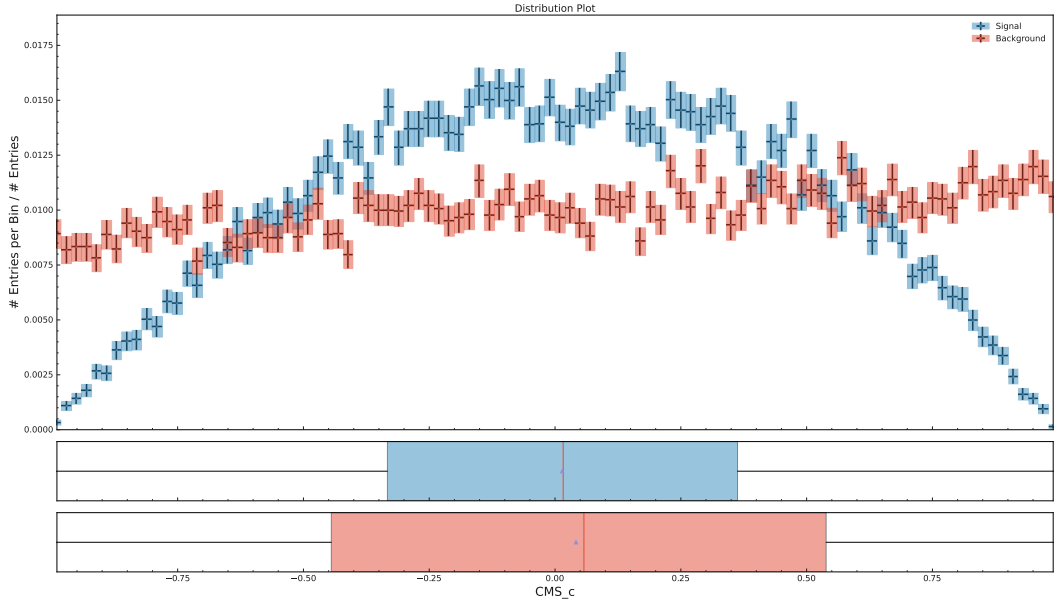
2.9 KSFWVariables(hso02)



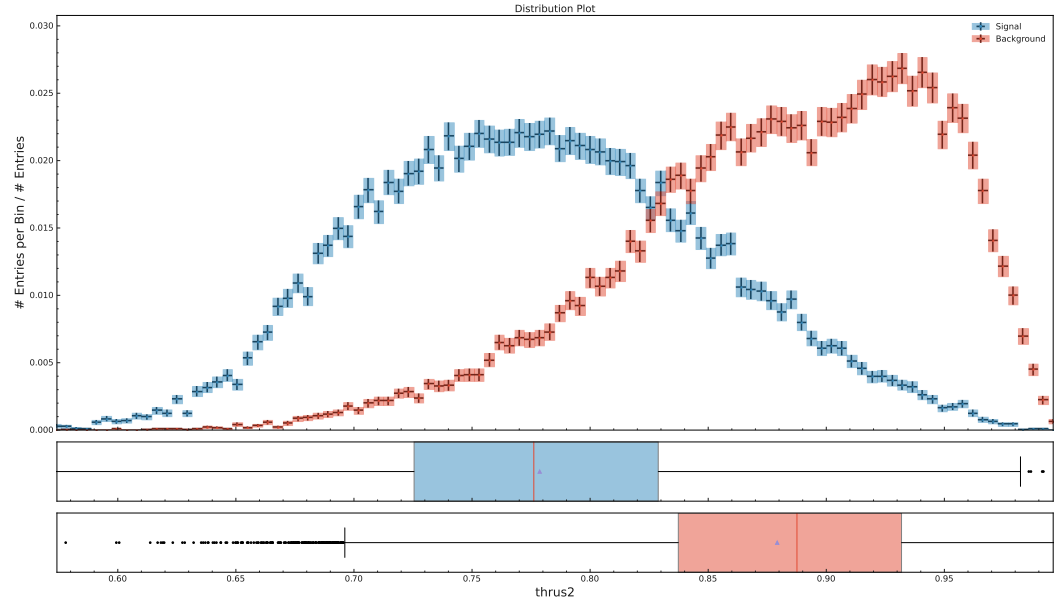
2.10 KSFWVariables(hso12)



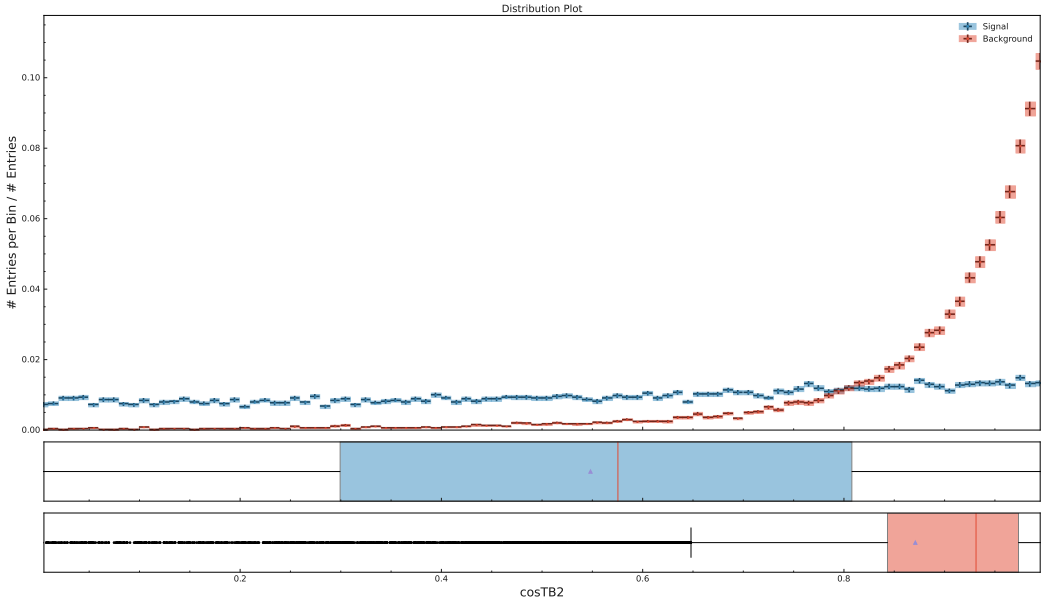
2.11 CMS_cosTheta



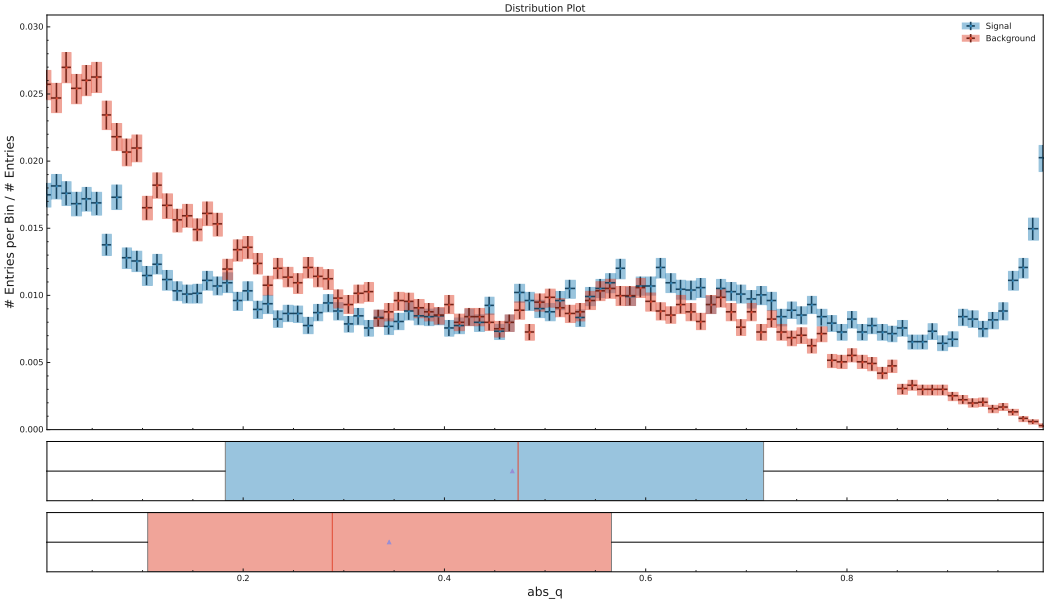
2.12 thrustBm



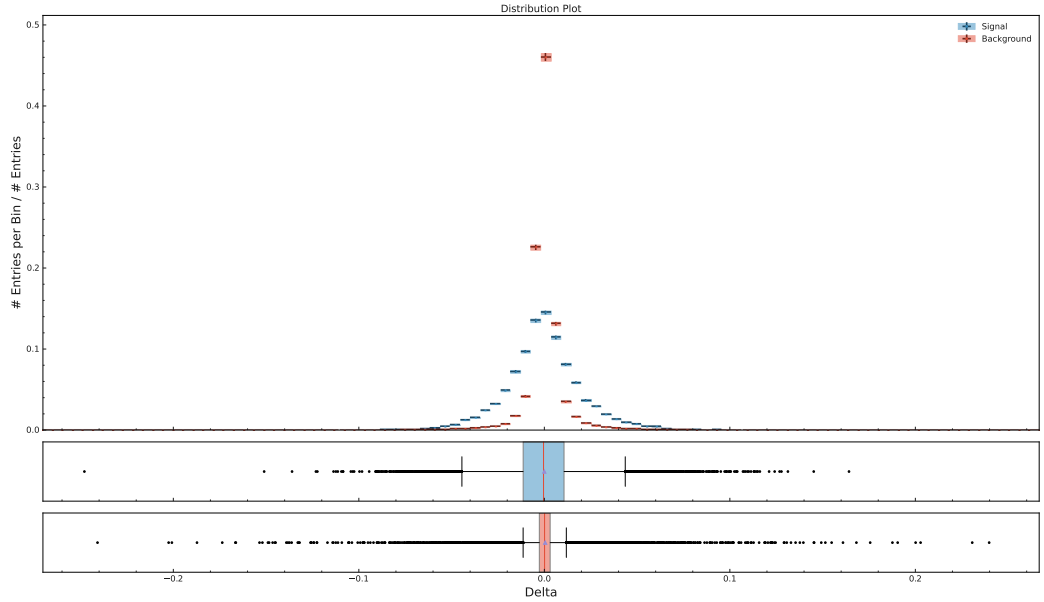
2.13 cosTBTO



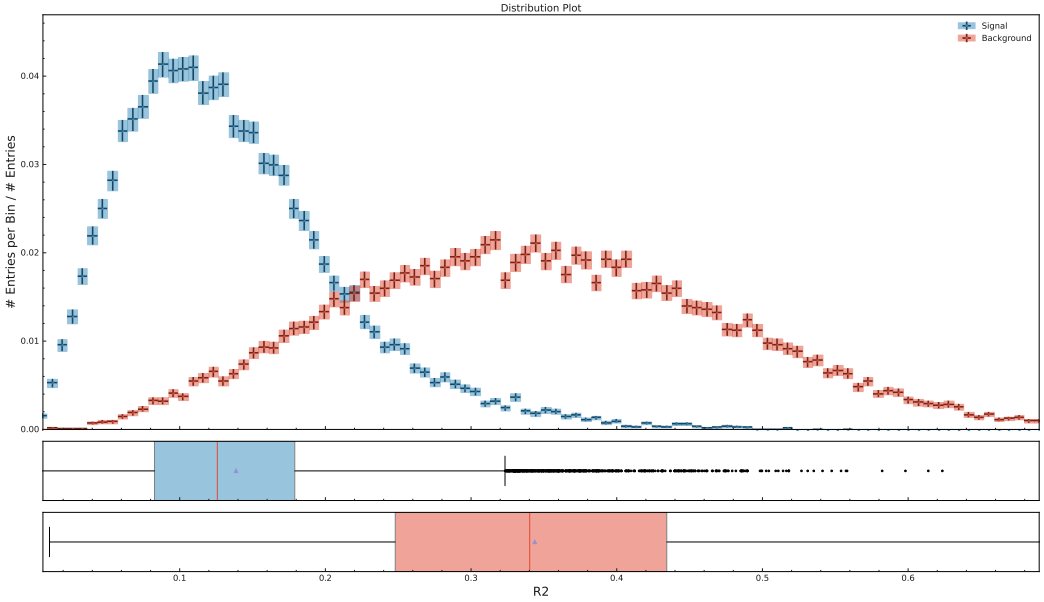
2.14 abs_qr



2.15 DeltaZ



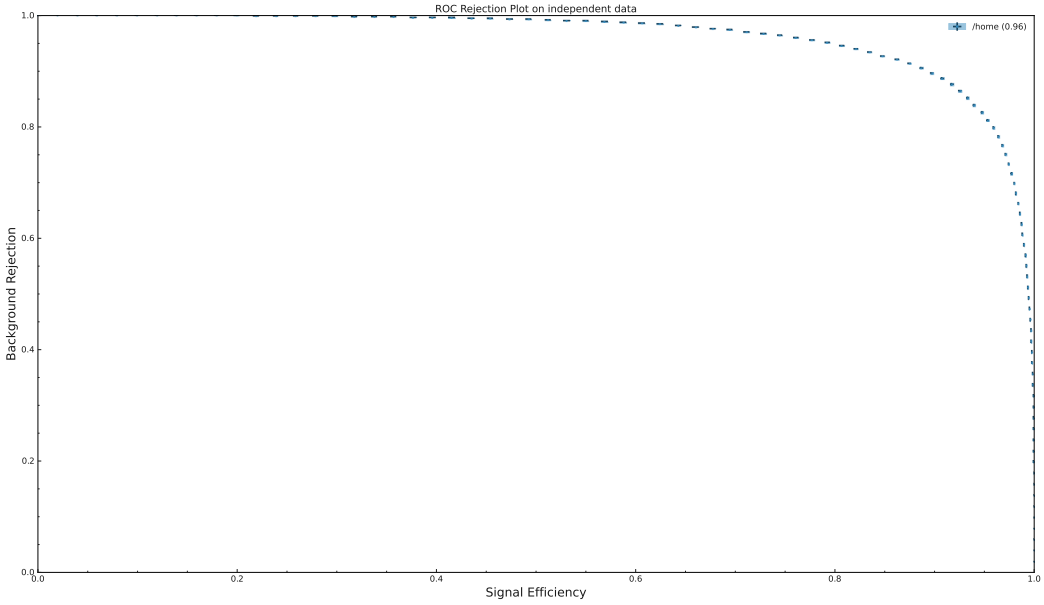
2.16 R2

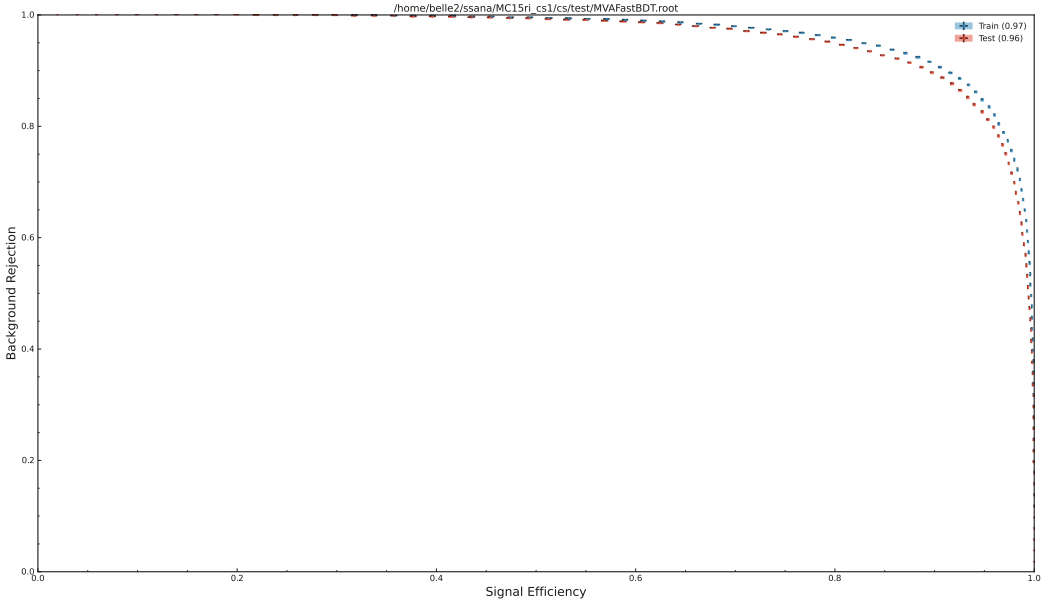


3 Classifier Plot

This section contains the receiver operating characteristics (ROC), purity projection, ...of the classifiers on training and independent data.The legend of each plot contains the shortened identifier and the area under the ROC curvein parenthesis.

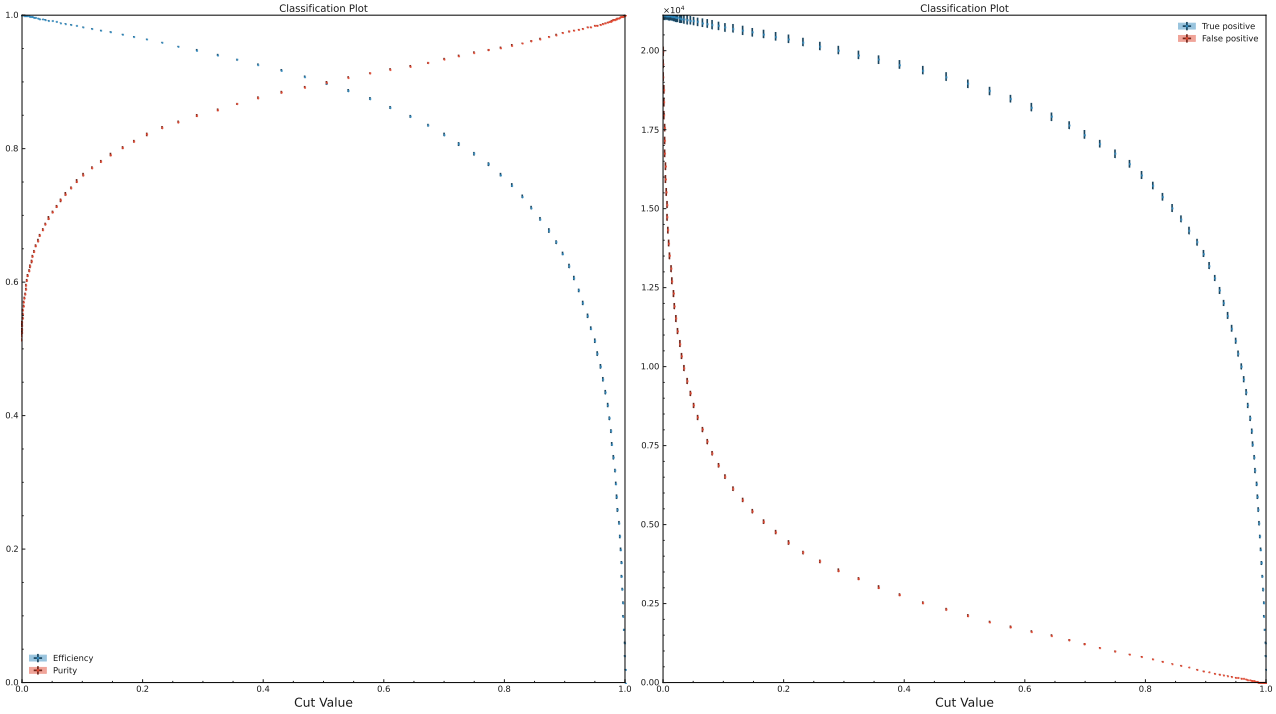
4 ROC Plot


ROC Rejection Plot on independent data
/home (0.96)
Background Rejection
Signal Efficiency
0.0
0.2
0.4
0.6
0.8
1.0


/home/belle2/ssana/MC15ri_cs1/cs/test/MVAFastBDT.root
Train (0.97)
Test (0.96)
Background Rejection
Signal Efficiency
0.0
0.2
0.4
0.6
0.8
1.0

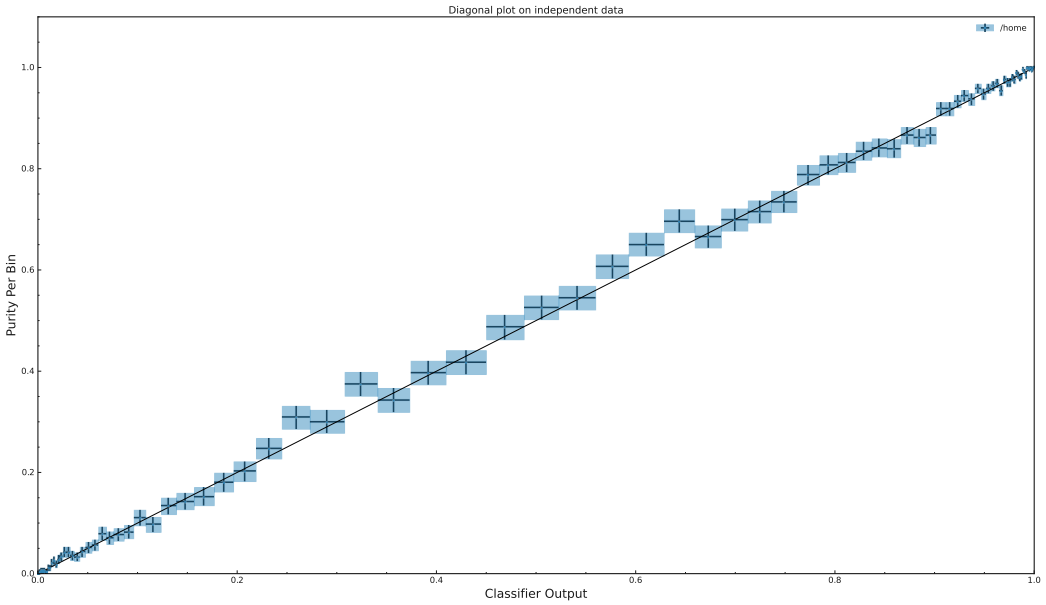
5 Classification Results

5.1 /home

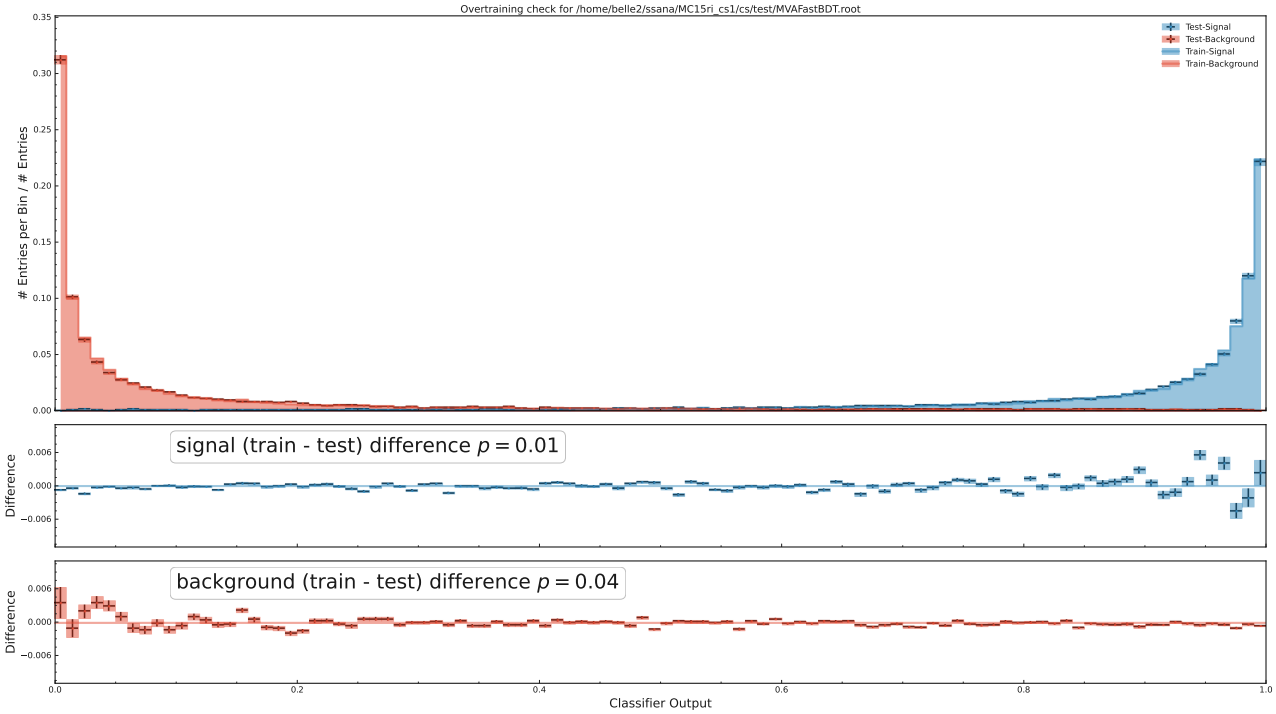


6 Diagonal Plot

6.1 /home



6.2 Overtraining Plot



7 Spectators

This section contains the distribution and dependence on theclassifier outputs of all spectator variables.

Table 3: Abbreviations of spectators

Spectator	Abbreviation