Automatic MVA Evaluation

Thomas Keck
Moritz Gelb
Nils Braun

August 9, 2019

Abstract

Evaluation plots

Contents

1 Classifiers

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

Table 1: Abbreviations of identifiers

Identifier	Abbreviation
mva_cs_id_2.xml	mva_c

1.1 mva_cs_id_2.xml

```
<?xml version="1.0" encoding="utf-8"?>
<method>FastBDT</method>
<weightfile>mva_cs_id_2.xml</weightfile>
<treename>tree</treename>
<target_variable>isSignal</target_variable>
<weight_variable/>
<signal_class>1</signal_class>
<max_events>0</max_events>
<number_feature_variables>10</number_feature_variables>
<variable0>R2</variable0>
<variable1>thrustBm</variable1>
<variable2>thrustOm</variable2>
<variable3>cosTBTO</variable3>
<variable4>cosTBz</variable4>
<variable5>KSFWVariables(et)</variable5>
<variable6>KSFWVariables(hso02)</variable6>
<variable7>KSFWVariables(hso14)</variable7>
<variable8>KSFWVariables(hoo2)</variable8>
<variable9>CleoConeCS(3)</variable9>
<number_spectator_variables>1</number_spectator_variables>
<spectator0>mbc</spectator0>
<number_data_files>1</number_data_files>
<datafile0>~/../eganiev/analysis/phase3/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>
```

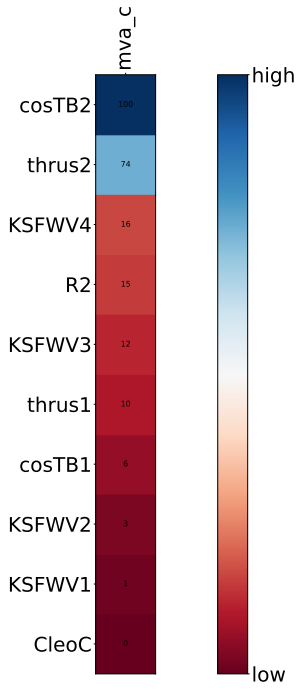
2 Variables

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.

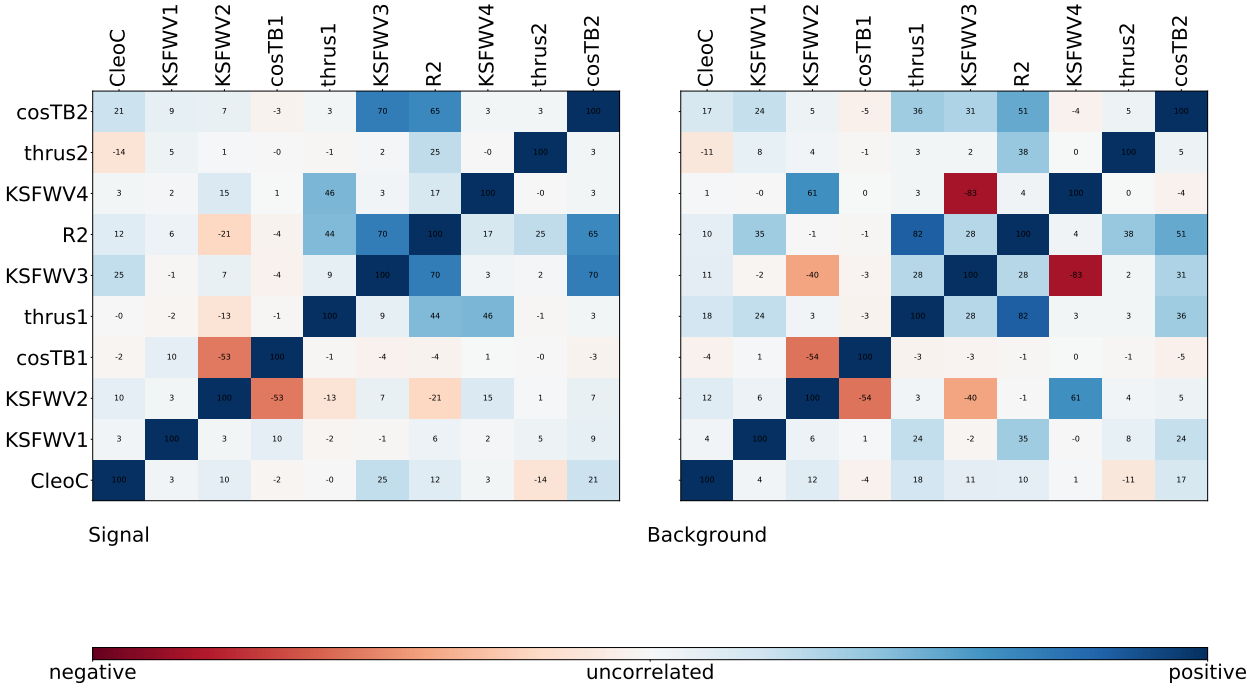
Table 2: Abbreviations of variables

Variable	Abbreviation
CleoConeCS(3)	CleoC
KSFWVariables(hso14)	KSFWV1
KSFWVariables(et)	KSFWV2
cosTBz	cosTB1
thrustOm	thrus1
KSFWVariables(hso02)	KSFWV3
R2	R2
KSFWVariables(hoo2)	KSFWV4
thrustBm	thrus2
cosTBTO	cosTB2

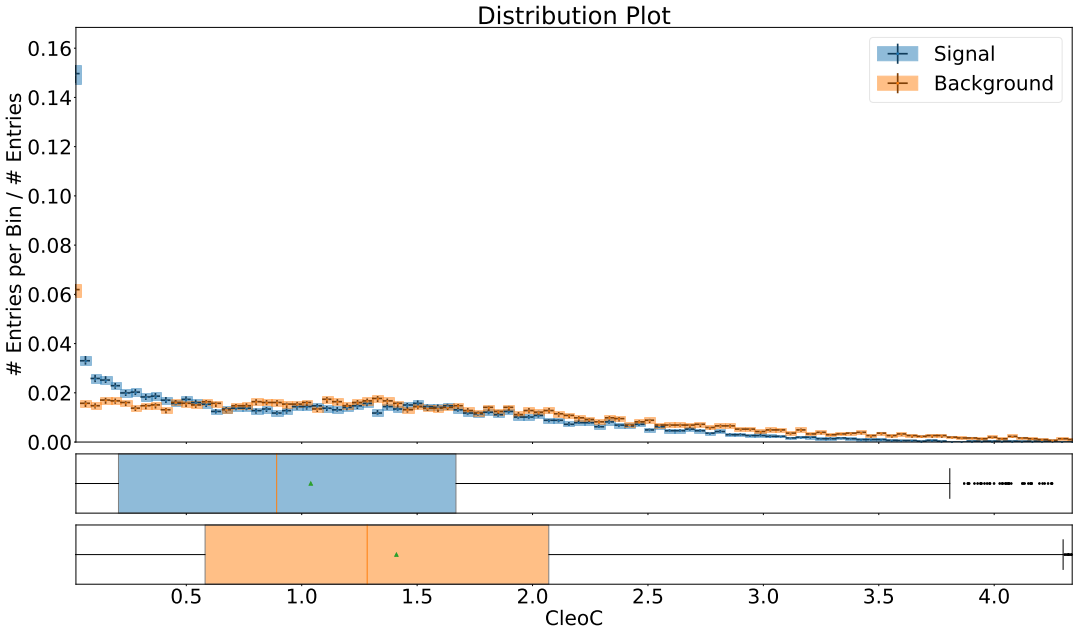
2.1 Importance



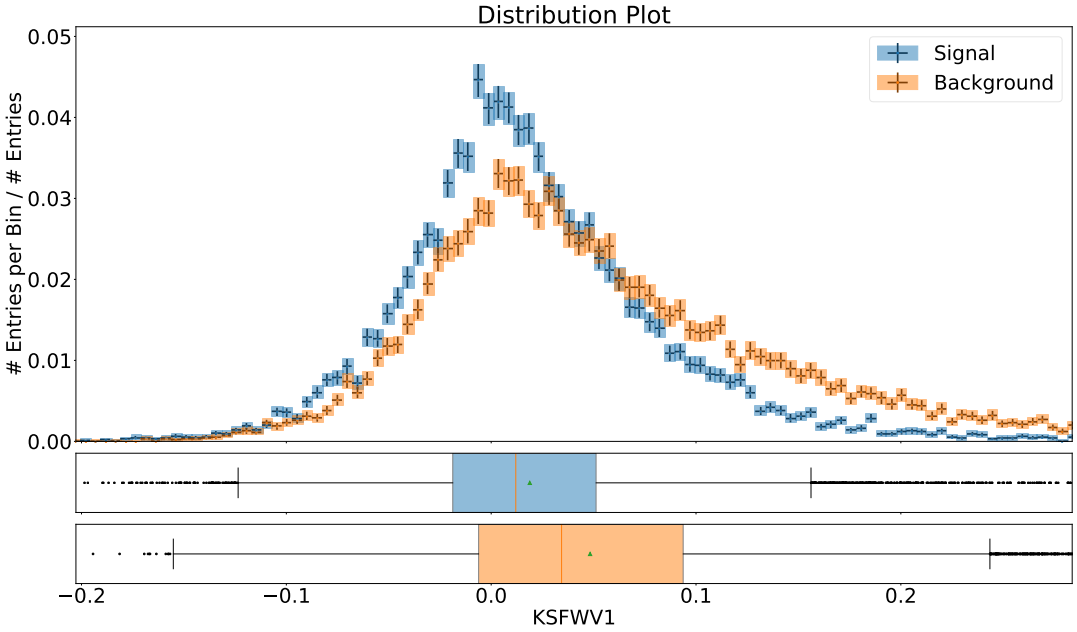
2.2 Correlation



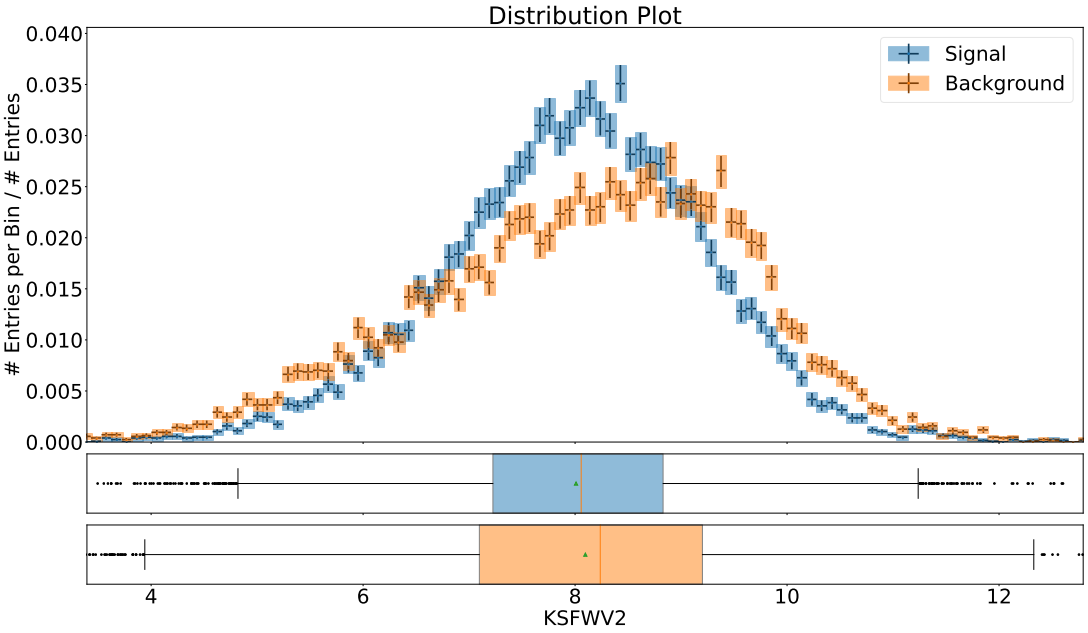
2.3 CleoConeCS(3)



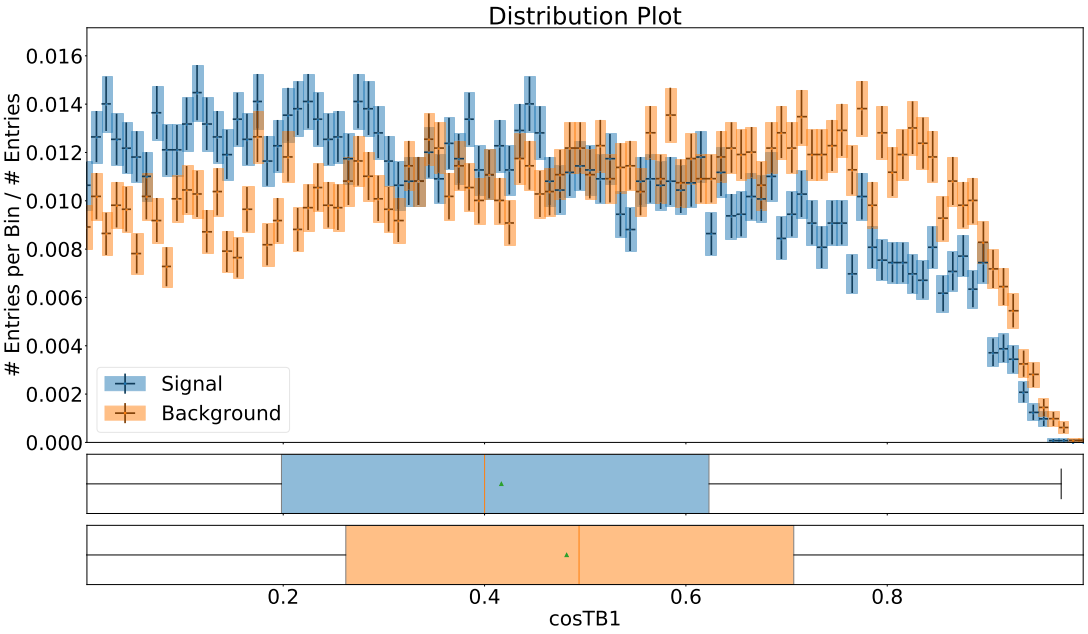
2.4 KSFWVariables(hso14)



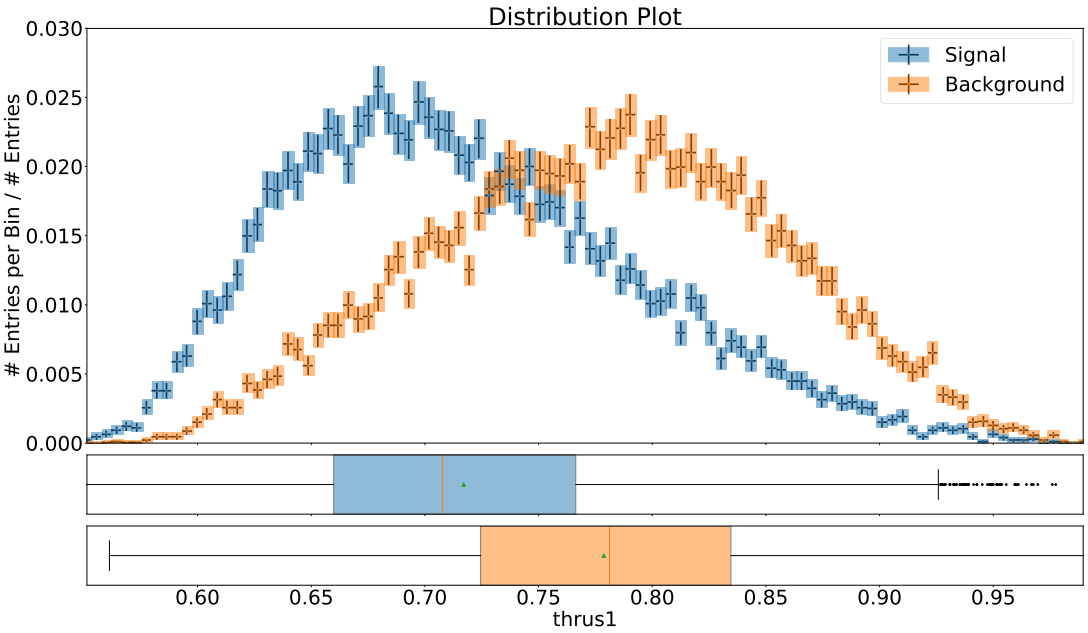
2.5 KSFWVariables(et)



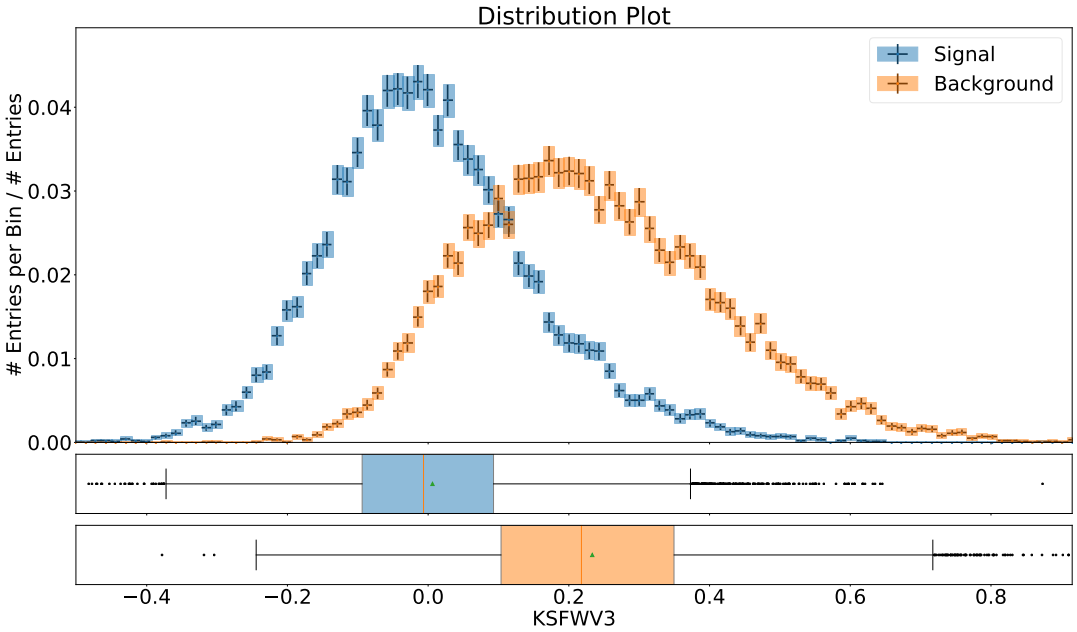
2.6 cosTBz



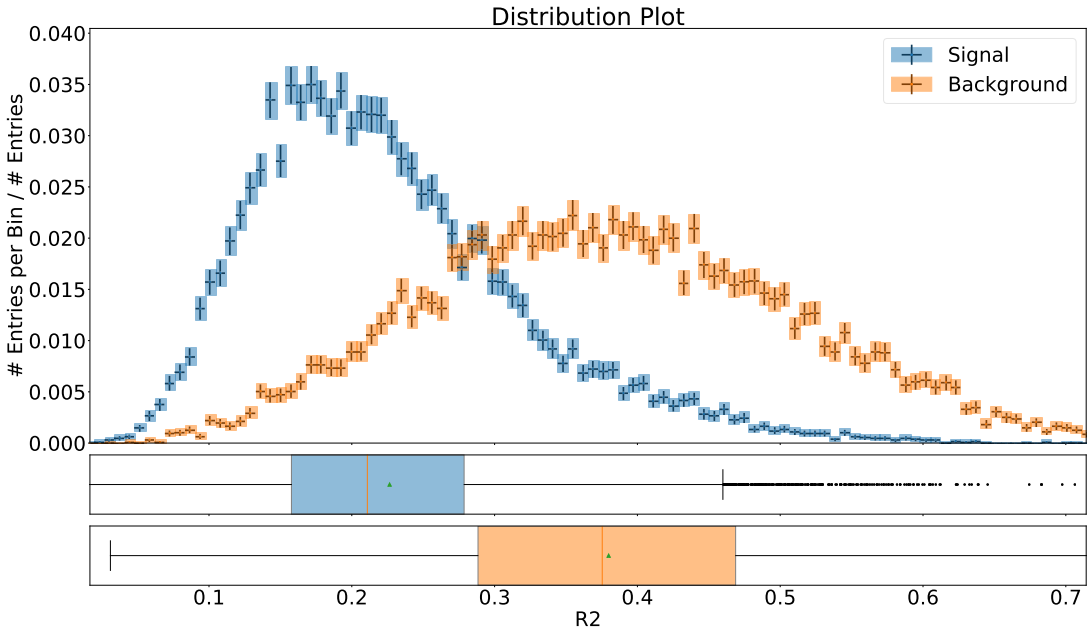
2.7 thrustOm



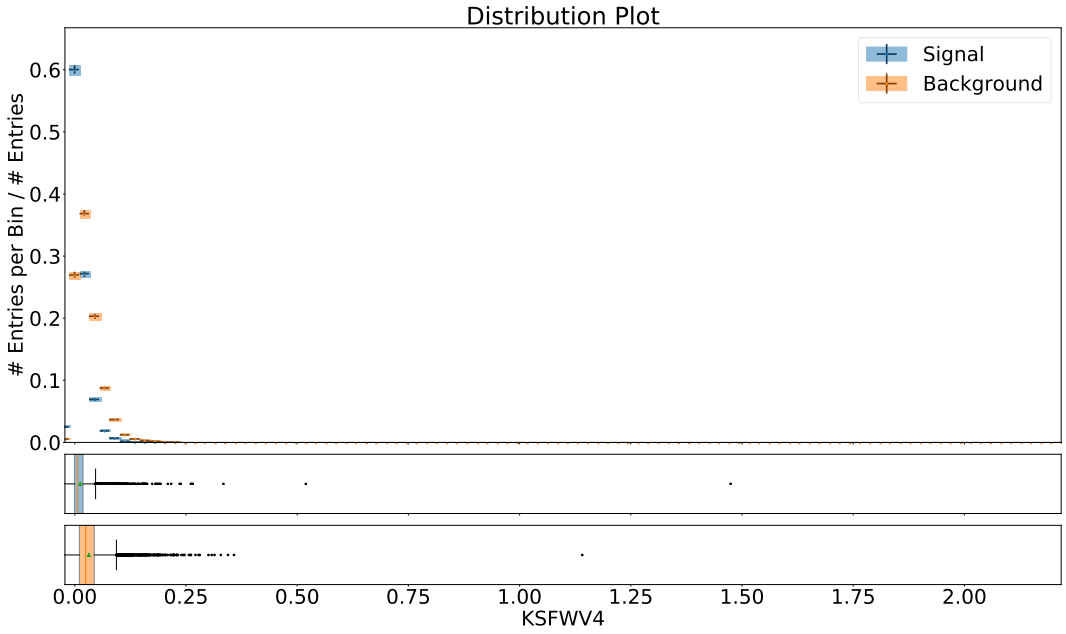
2.8 KSFWVariables(hso02)



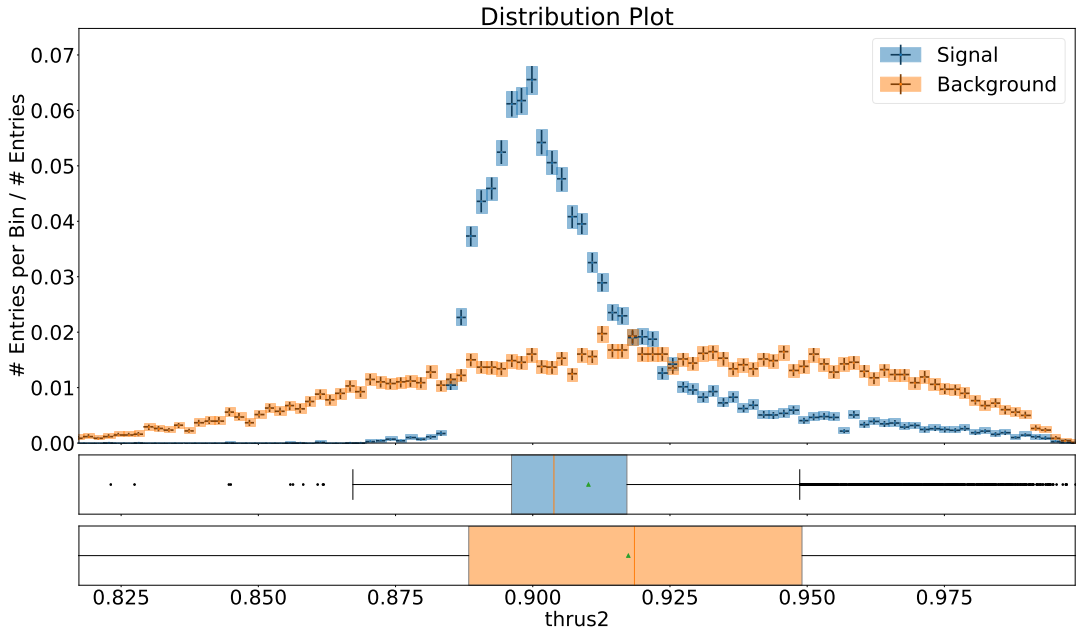
2.9 R2



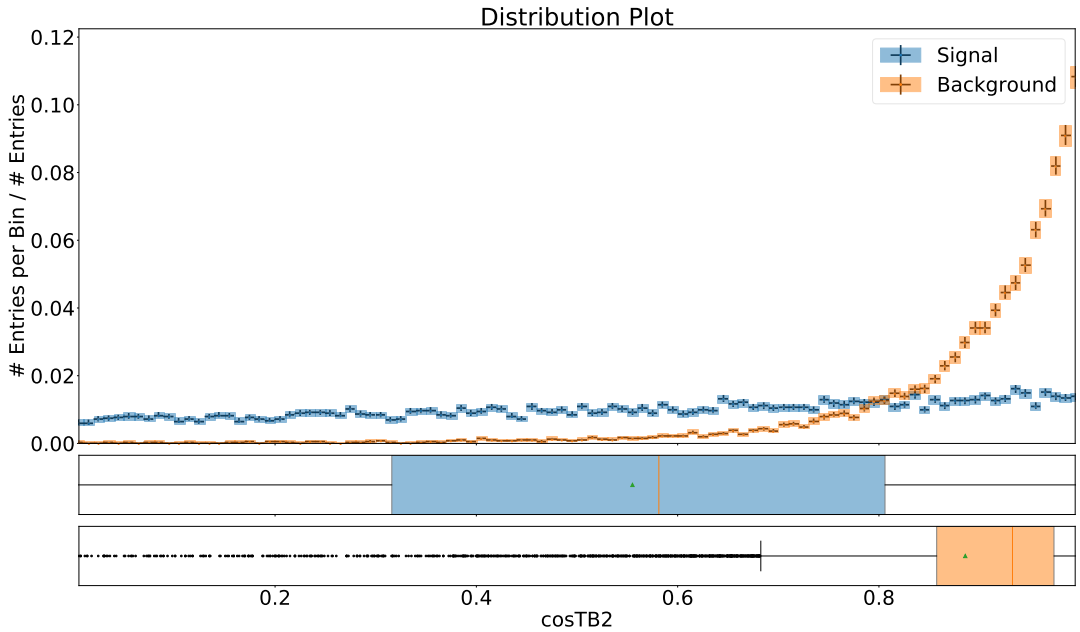
2.10 KSFWVariables(hoo2)



2.11 thrustBm



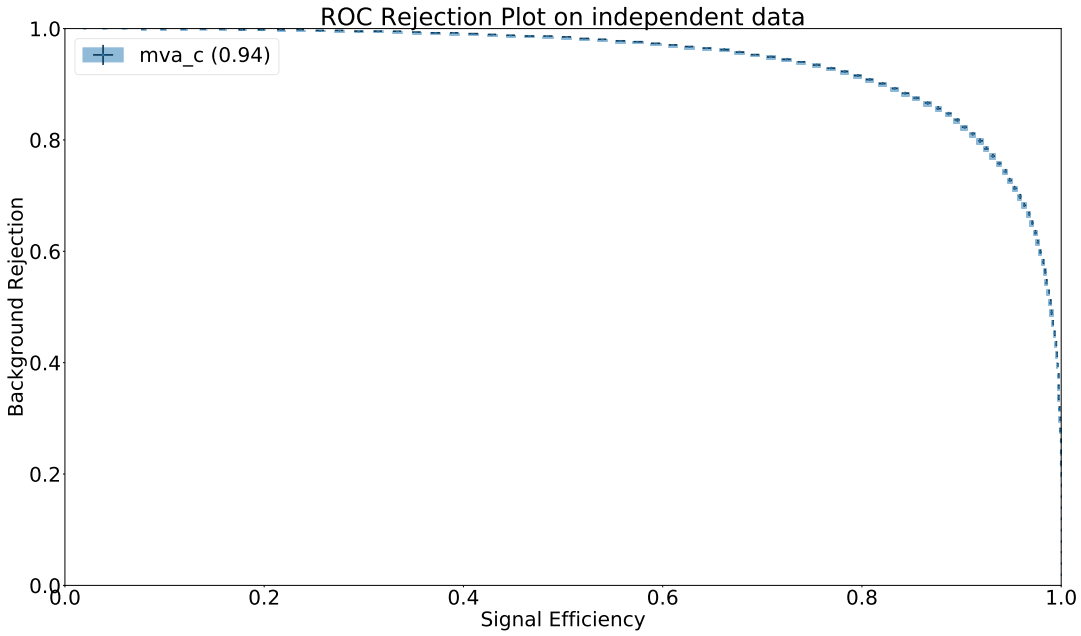
2.12 cosTBTO

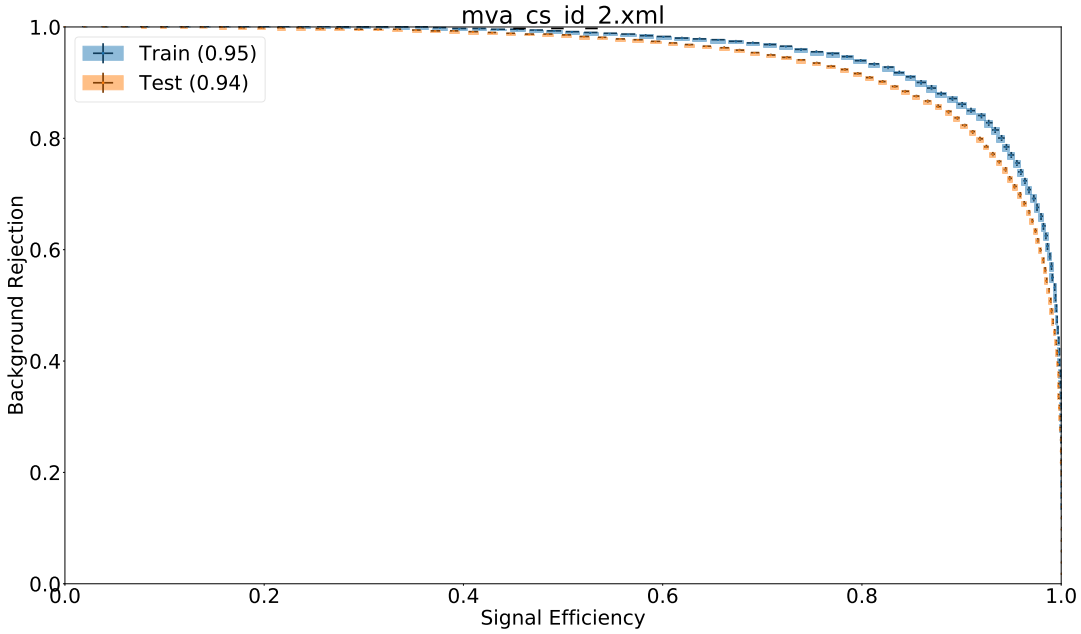


3 Classifier Plot

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.

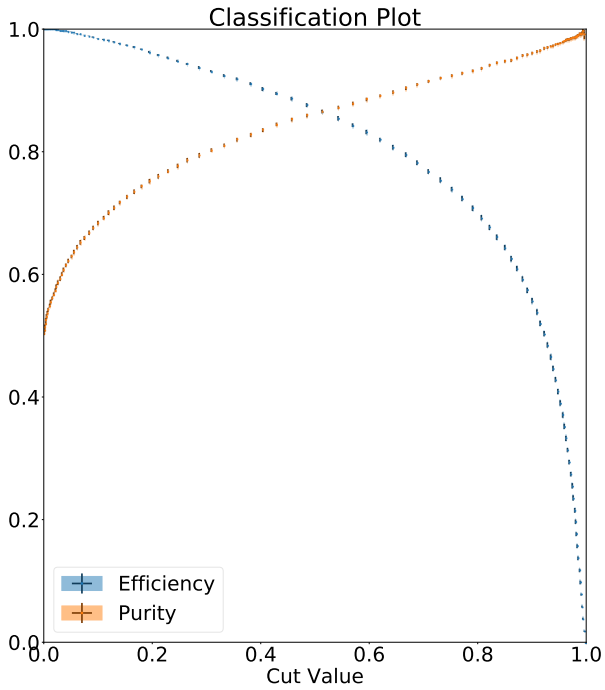
4 ROC Plot

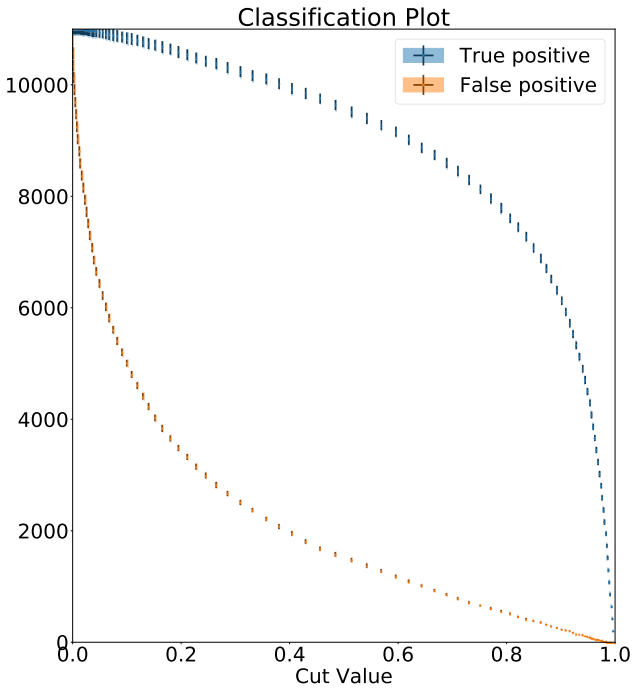

ROC Rejection Plot on independent data
mva_c (0.94)
Background Rejection
Signal Efficiency
0.0
0.2
0.4
0.6
0.8
1.0


mva_cs_id_2.xml
Train (0.95)
Test (0.94)
Background Rejection
Signal Efficiency
0.0
0.2
0.4
0.6
0.8
1.0

5 Classification Results

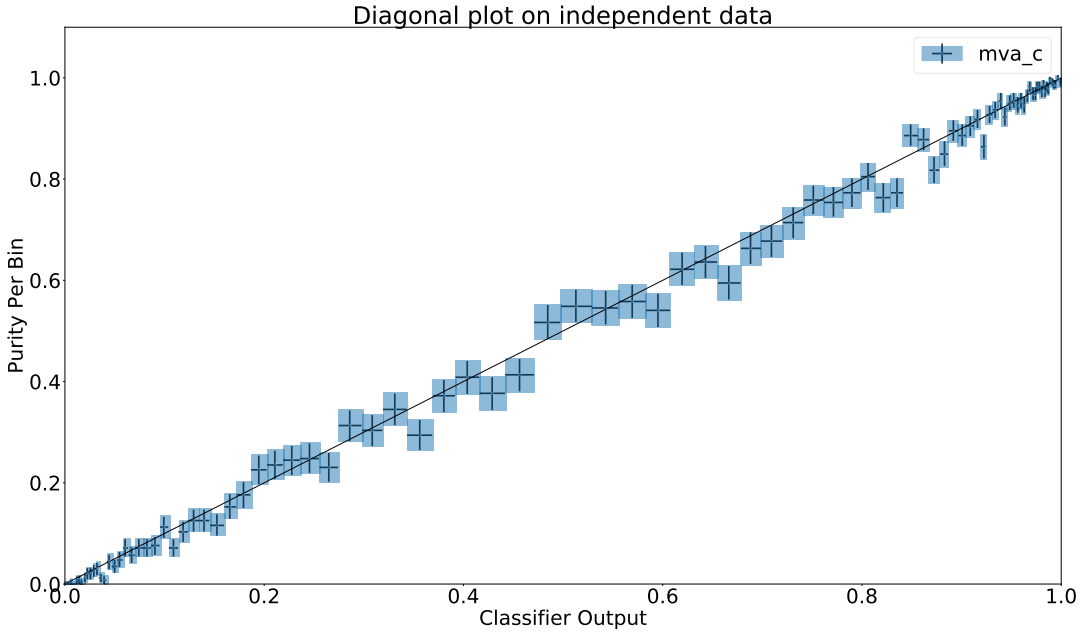
5.1 mva_c



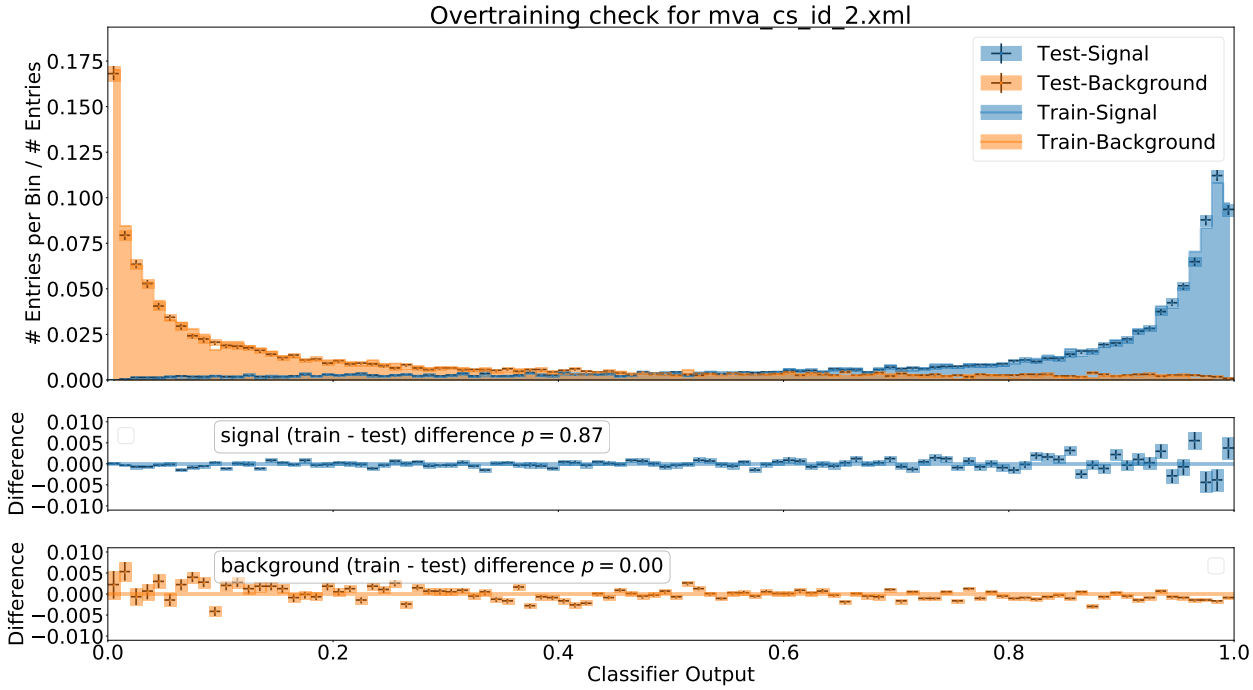


6 Diagonal Plot

6.1 mva_c



6.2 Overtraining Plot



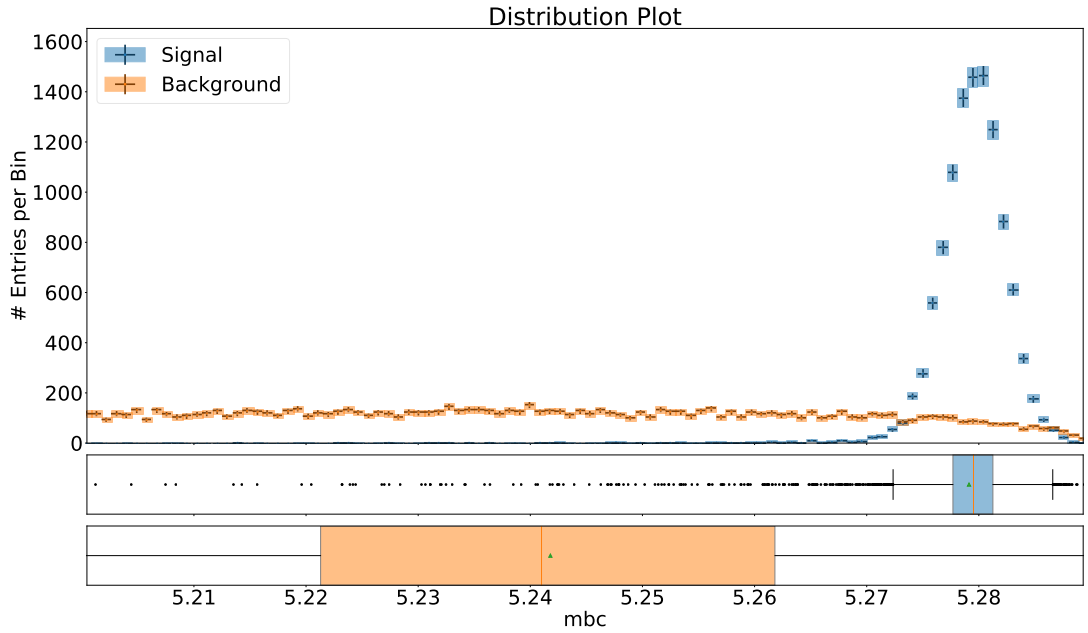
7 Spectators

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

Table 3: Abbreviations of spectators

Spectator	Abbreviation
mbc	mbc

7.1 mbc



7.1.1 mbc with classifier mva_cs_id_2.xml