Automatic MVA Evaluation

Thomas Keck
Moritz Gelb
Nils Braun

April 9, 2023

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 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
<?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>11</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(hso02)</variable8>
<variable9>KSFWVariables(hso12)</variable9>
<variable10>KSFWVariables(hoo0)</variable10>
<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>
```
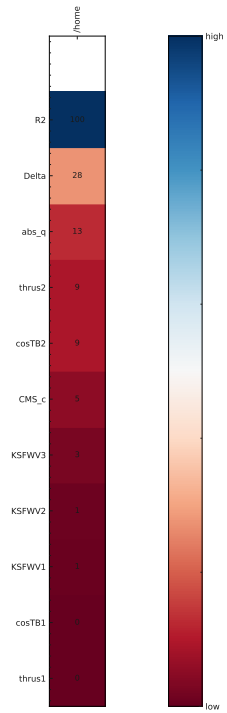
2 Variables

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.
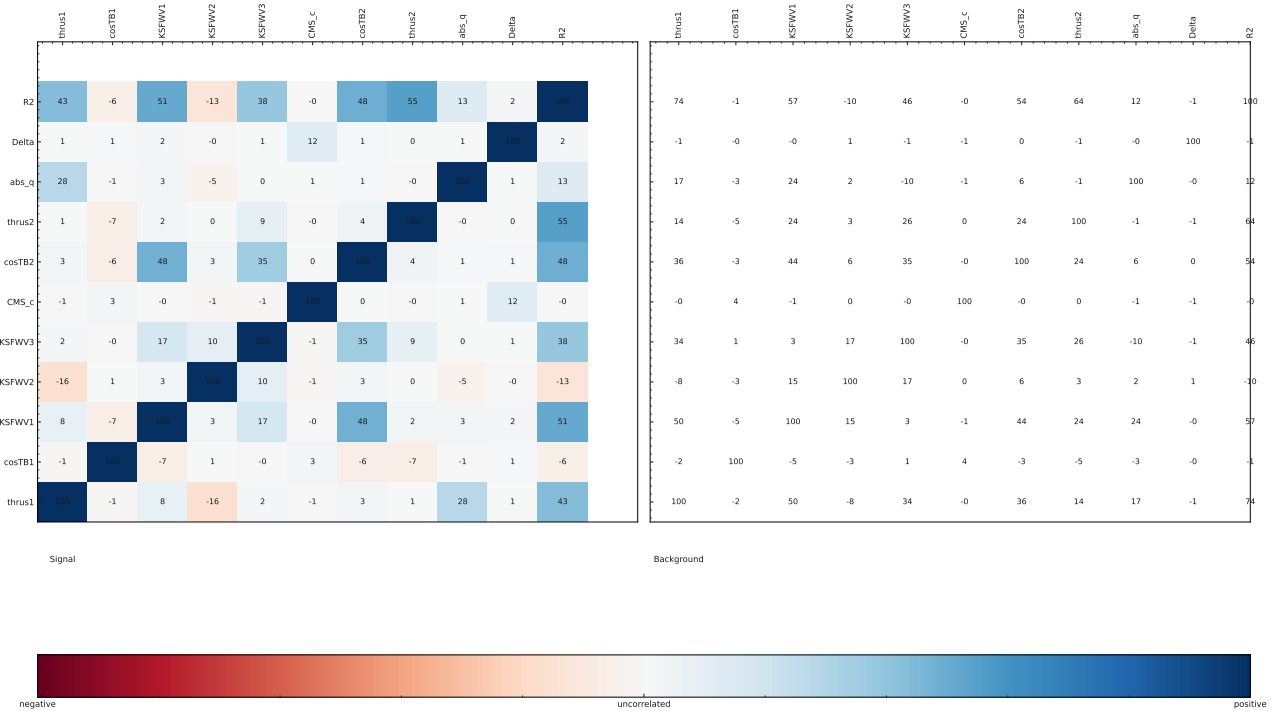
Table 2: Abbreviations of variables

Variable	Abbreviation
thrustOm	thrus1
cosTBz	cosTB1
KSFWVariables(hso02)	KSFWV1
KSFWVariables(hoo0)	KSFWV2
KSFWVariables(hso12)	KSFWV3
CMS_cosTheta	CMS_c
cosTBTO	cosTB2
thrustBm	thrus2
abs_qr	abs_q
DeltaZ	Delta

2.1 Importance

2.2 Correlation

Signal

	thrus1	cosTB1	KSFWV1	KSFWV2	KSFWV3	CMS_c	cosTB2	thrus2	abs_q	Delta	R2
R2	43	-6	51	-13	38	-0	48	55	13	2	100
Delta	1	1	2	-0	1	12	1	0	1	100	2
abs_q	28	-1	3	-5	0	1	1	-0	100	1	13
thrus2	1	-7	2	0	9	-0	4	100	-0	0	55
cosTB2	3	-6	48	3	35	0	100	4	1	1	48
CMS_c	-1	3	-0	-1	-1	100	0	-0	1	12	-0
KSFWV3	2	-0	17	10	100	-1	35	9	0	1	38
KSFWV2	-16	1	3	100	10	-1	3	0	-5	-0	-13
KSFWV1	8	-7	100	3	17	-0	48	2	3	2	51
cosTB1	-1	100	-7	1	-0	3	-6	-7	-1	1	-6
thrus1	100	-1	8	-16	2	-1	3	1	28	1	43

Background

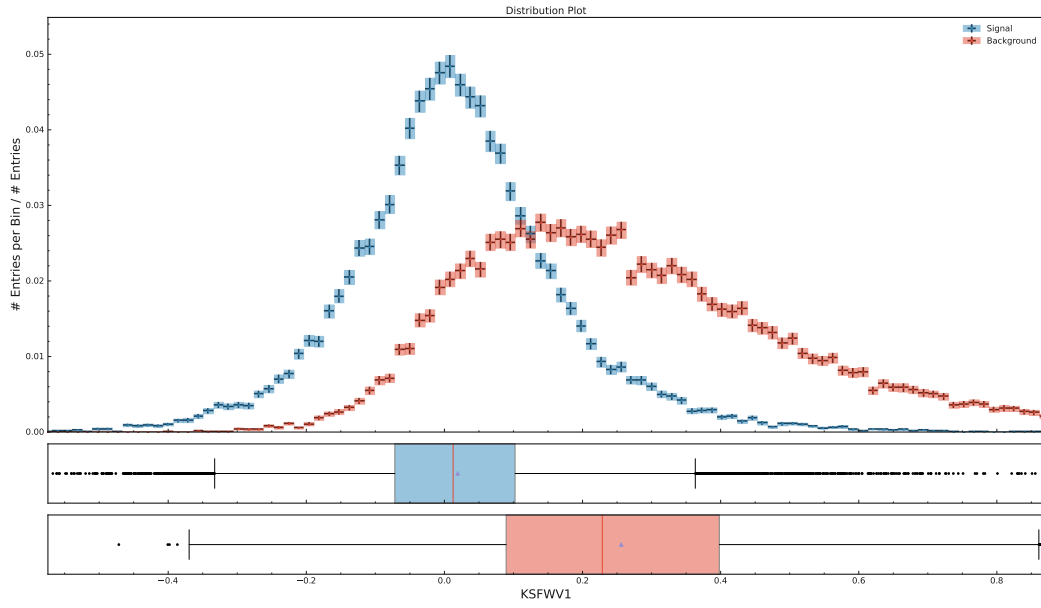|	thrus1	cosTB1	KSFWV1	KSFWV2	KSFWV3	CMS_c	cosTB2	thrus2	abs_q	Delta	R2
R2	74	-1	57	-10	46	-0	54	64	12	-1	100
Delta	-1	-0	-0	1	-1	-1	0	-1	-0	100	-1
abs_q	17	-3	24	2	-10	-1	6	-1	100	-0	12
thrus2	14	-5	24	3	26	0	24	100	-1	-1	64
cosTB2	36	-3	44	6	35	-0	100	24	6	0	54
CMS_c	-0	4	-1	0	-0	100	-0	0	-1	-1	-0
KSFWV3	34	1	3	17	100	-0	35	26	-10	-1	46
KSFWV2	-8	-3	15	100	17	0	6	3	2	1	-10
KSFWV1	50	-5	100	15	3	-1	44	24	24	-0	57
cosTB1	-2	100	-5	-3	1	4	-3	-5	-3	-0	-1
thrus1	100	-2	50	-8	34	-0	36	14	17	-1	74

negative   uncorrelated   positive

2.3 thrustOm
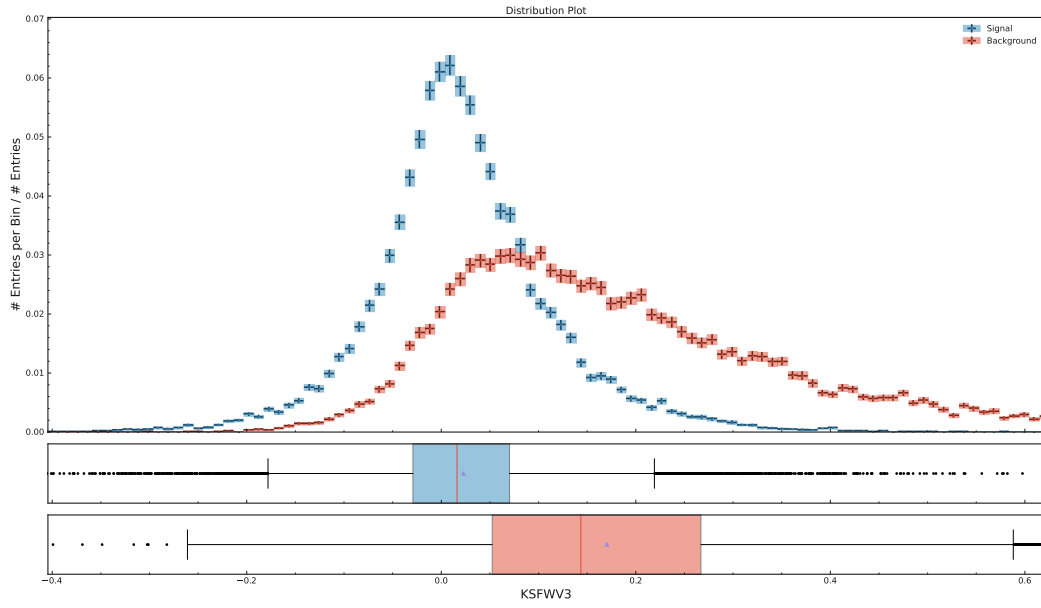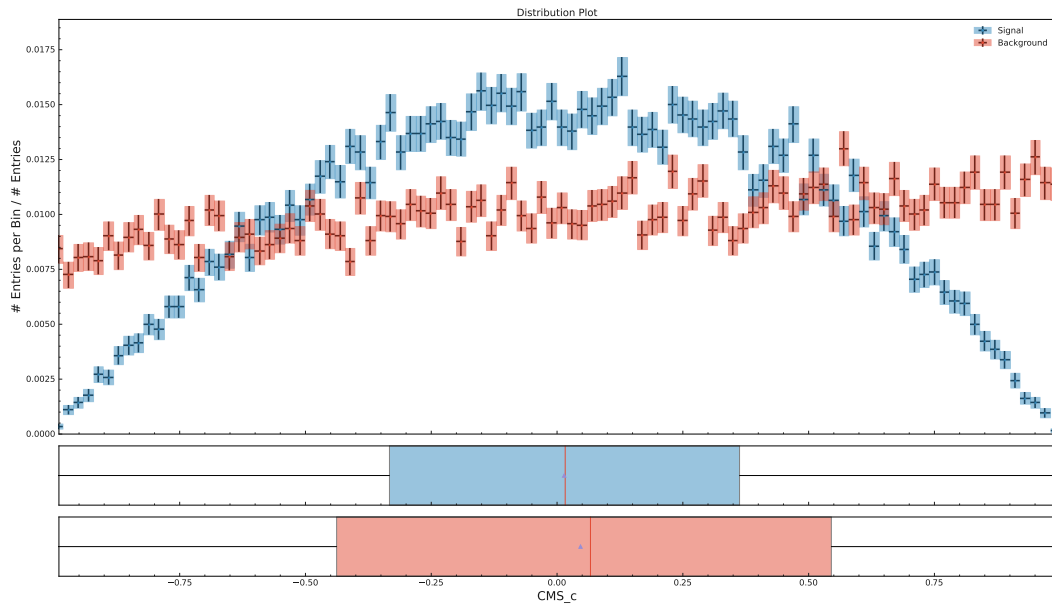

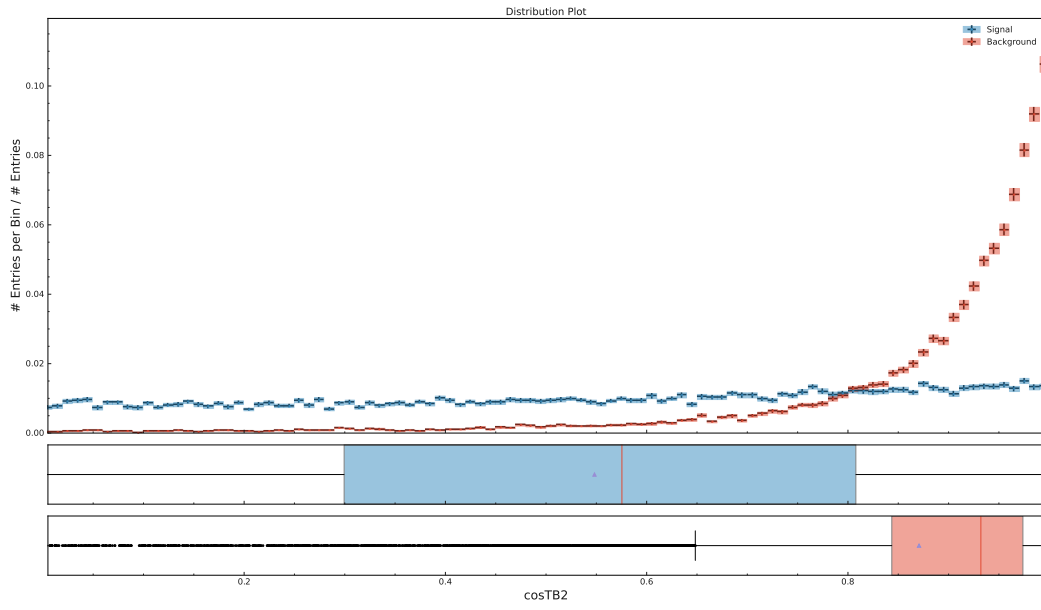## 2.4 cosTBz

2.5 KSFWVariables(hso02)
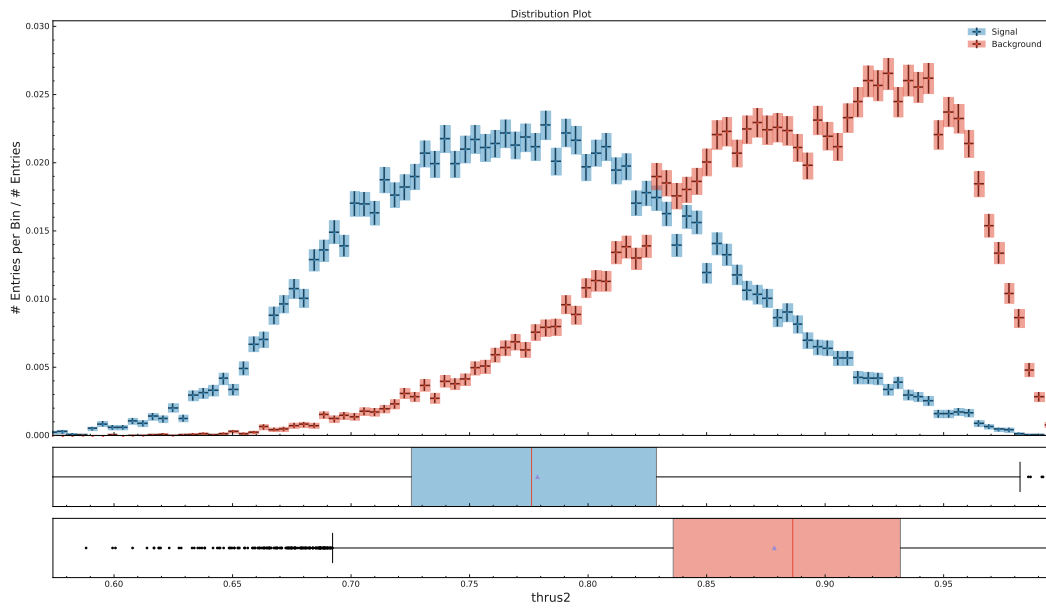

2.6 KSFWVariables(hoo0)

2.7 KSFWVariables(hso12)

2.8 CMS_cosTheta
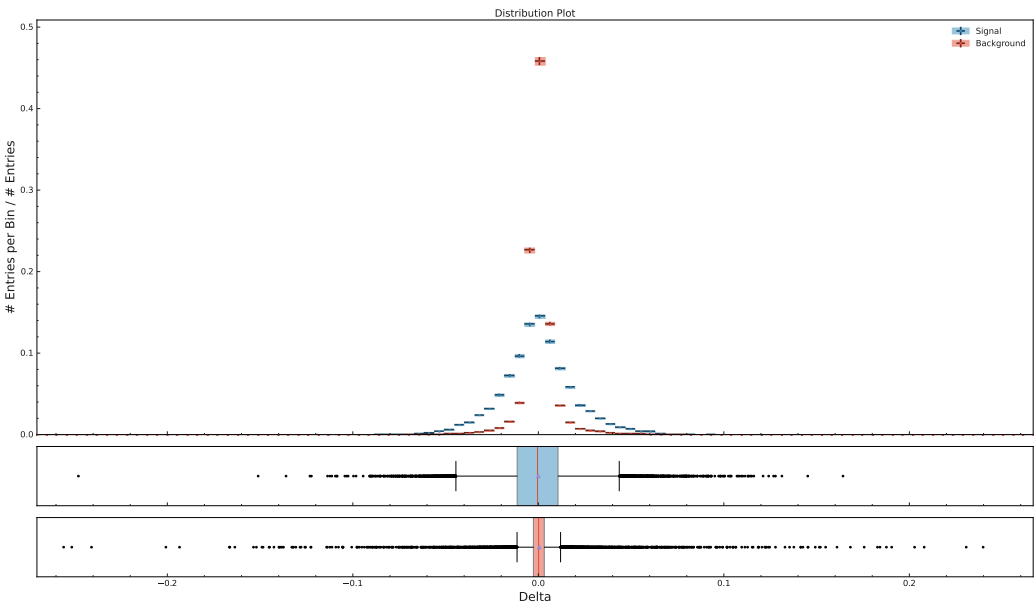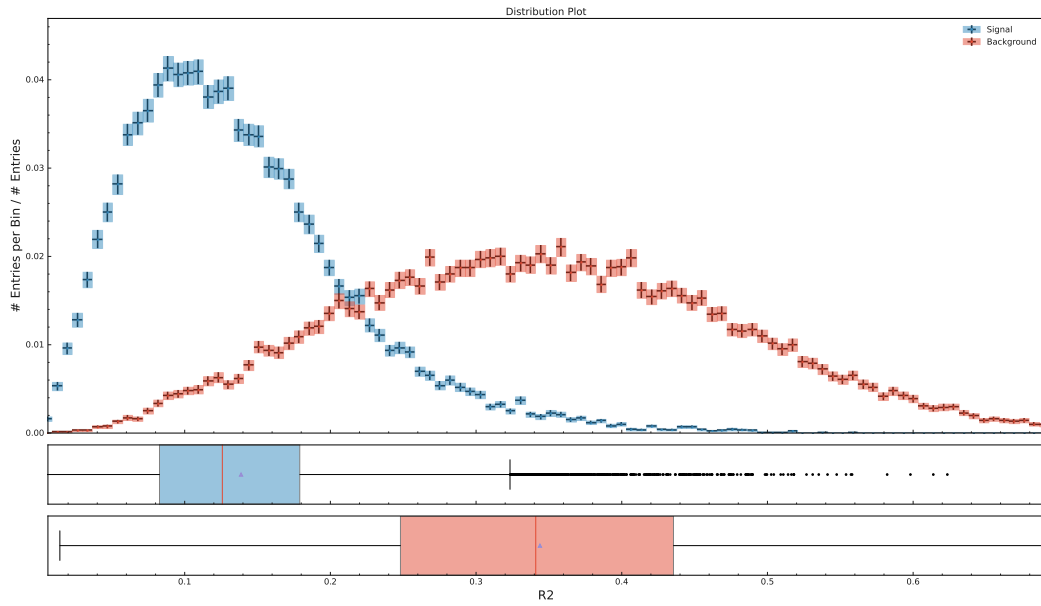
2.9 cosTBTO

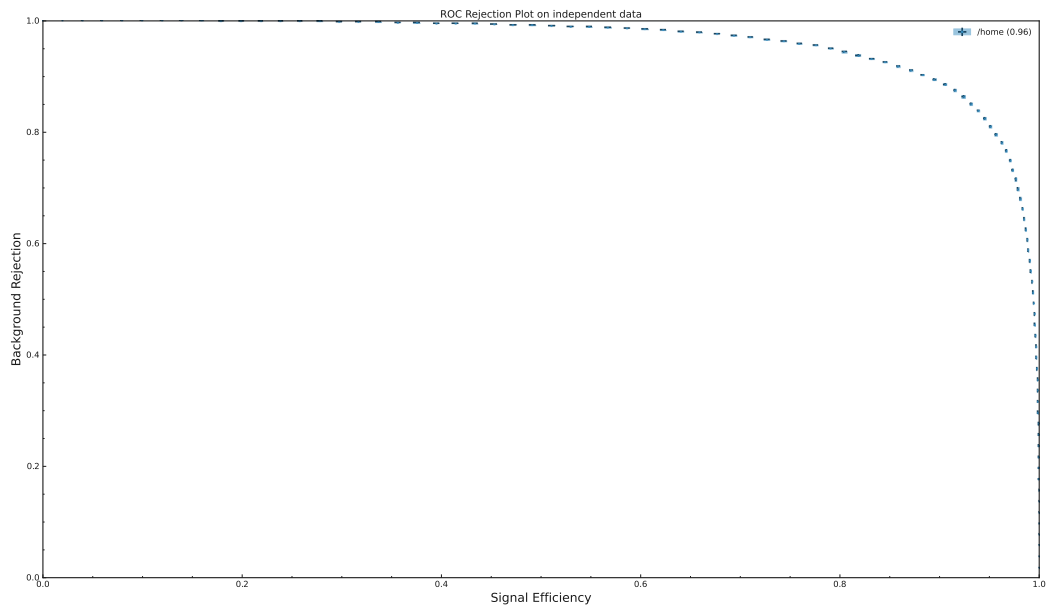## 2.10 thrustBm
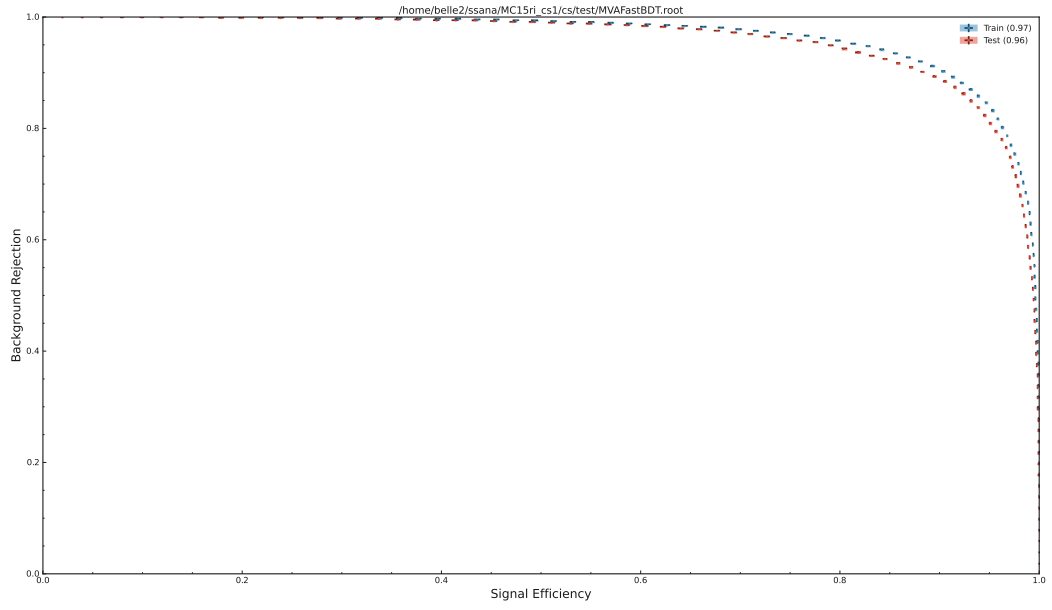
2.11   abs_qr
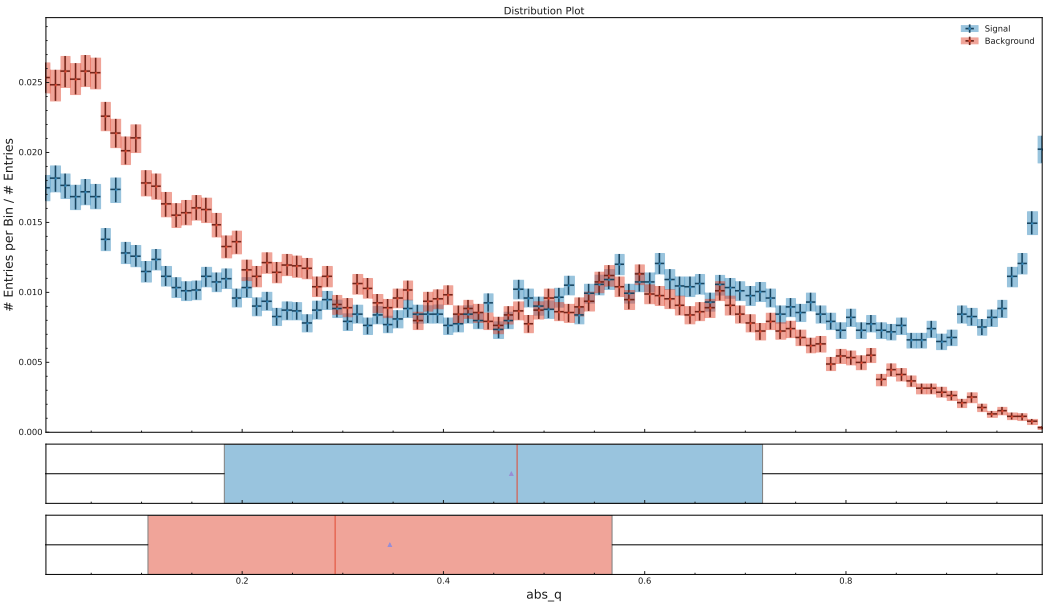
2.12   DeltaZ

2.13   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.
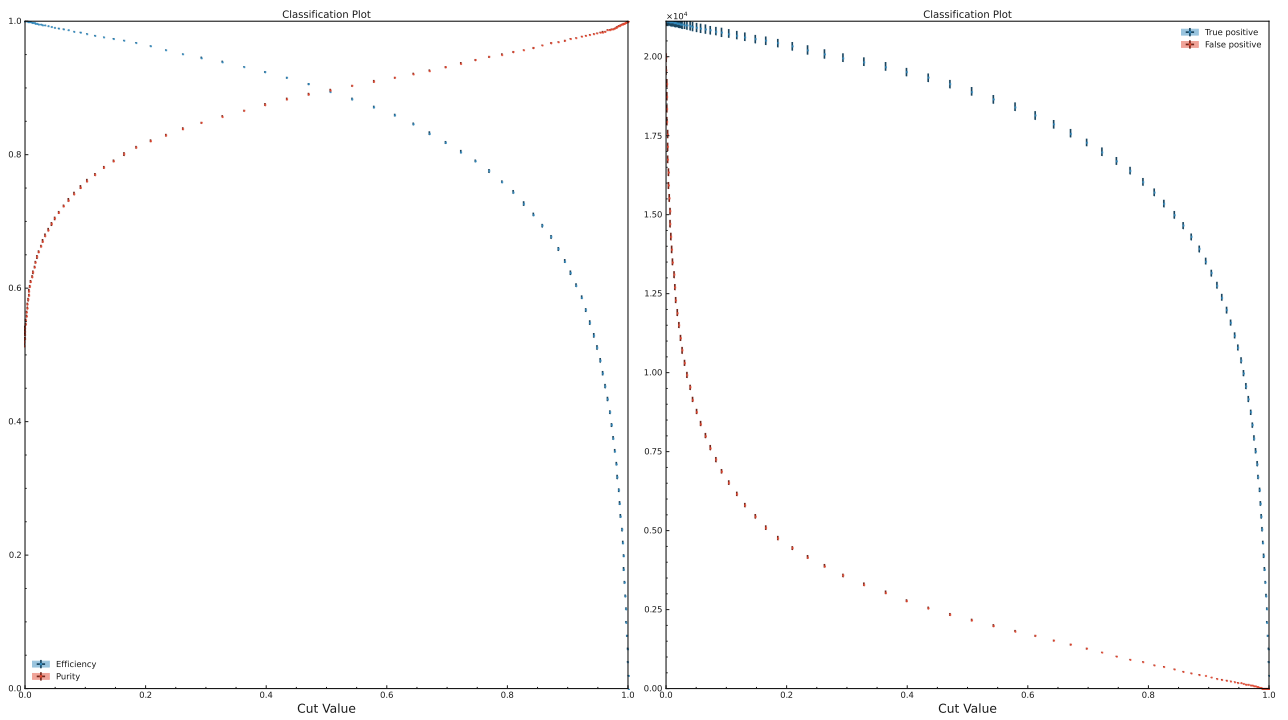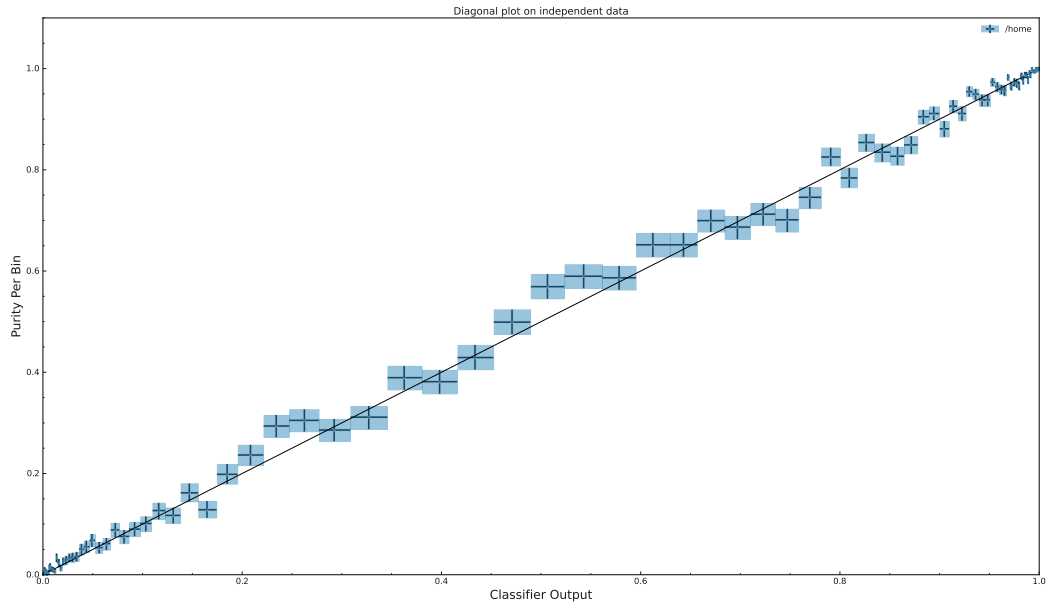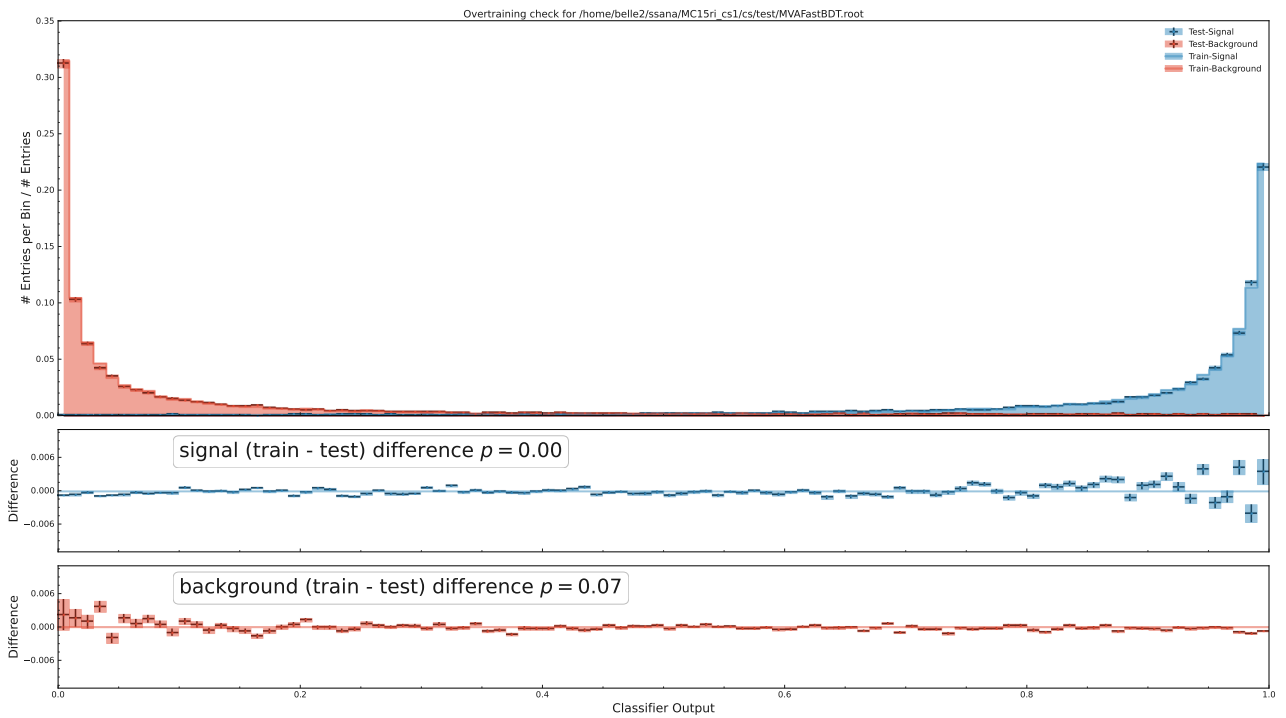
4   ROC Plot

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