

Automatic MVA Evaluation

Thomas Keck
Moritz Gelb
Nils Braun

August 8, 2019

Abstract

Evaluation plots

Contents

1	Classifiers	3
1.1	mva_cs_id.xml	3
2	Variables	3
2.1	Importance	4
2.2	Correlation	5
2.3	CleoConeCS(3)	5
2.4	KSFVariables(hso00)	6
2.5	KSFVariables(hoo4)	6
2.6	CleoConeCS(5)	7
2.7	KSFVariables(hoo3)	7
2.8	CleoConeCS(4)	8
2.9	KSFVariables(mm2)	8
2.10	KSFVariables(hoo1)	9
2.11	CleoConeCS(6)	9
2.12	CleoConeCS(8)	10
2.13	CleoConeCS(7)	10
2.14	KSFVariables(et)	11
2.15	KSFVariables(hso14)	11
2.16	KSFVariables(hso20)	12
2.17	KSFVariables(hso22)	12
2.18	KSFVariables(hso10)	13
2.19	KSFVariables(hoo0)	13
2.20	thrustOm	14
2.21	KSFVariables(hso24)	14
2.22	KSFVariables(hso04)	15
2.23	cosTBz	15
2.24	CleoConeCS(2)	16
2.25	KSFVariables(hso12)	16
2.26	KSFVariables(hoo2)	17
2.27	KSFVariables(hso02)	17
2.28	R2	18
2.29	thrustBm	18
2.30	CleoConeCS(1)	19
2.31	cosTBTO	19
3	Classifier Plot	19
4	ROC Plot	20
5	Classification Results	21
5.1	mva_c	21
6	Diagonal Plot	21
6.1	mva_c	21
6.2	Overtraining Plot	22
7	Spectators	22
7.1	mbc	23
7.1.1	mbc with classifier mva_cs_id.xml	23

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
mva_cs_id.xml	mva_c

1.1 mva_cs_id.xml

```
<?xml version="1.0" encoding="utf-8"?>
<method>FastBDT</method>
<weightfile>mva_cs_id.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>29</number_feature_variables>
<variable0>R2</variable0>
<variable1>thrustBm</variable1>
<variable2>thrustOm</variable2>
<variable3>cosTBTo</variable3>
<variable4>cosTBz</variable4>
<variable5>KSFVWVariables(et)</variable5>
<variable6>KSFVWVariables(mm2)</variable6>
<variable7>KSFVWVariables(hso00)</variable7>
<variable8>KSFVWVariables(hso02)</variable8>
<variable9>KSFVWVariables(hso04)</variable9>
<variable10>KSFVWVariables(hso10)</variable10>
<variable11>KSFVWVariables(hso12)</variable11>
<variable12>KSFVWVariables(hso14)</variable12>
<variable13>KSFVWVariables(hso20)</variable13>
<variable14>KSFVWVariables(hso22)</variable14>
<variable15>KSFVWVariables(hso24)</variable15>
<variable16>KSFVWVariables(hoo0)</variable16>
<variable17>KSFVWVariables(hoo1)</variable17>
<variable18>KSFVWVariables(hoo2)</variable18>
<variable19>KSFVWVariables(hoo3)</variable19>
<variable20>KSFVWVariables(hoo4)</variable20>
<variable21>CleoConeCS(1)</variable21>
<variable22>CleoConeCS(2)</variable22>
<variable23>CleoConeCS(3)</variable23>
<variable24>CleoConeCS(4)</variable24>
<variable25>CleoConeCS(5)</variable25>
<variable26>CleoConeCS(6)</variable26>
<variable27>CleoConeCS(7)</variable27>
<variable28>CleoConeCS(8)</variable28>
<number_spectator_variables>1</number_spectator_variables>
<spectator0>mbc</spectator0>
<number_data_files>1</number_data_files>
<datafile0>-/..//ganiev/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>0.1000000000000001</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

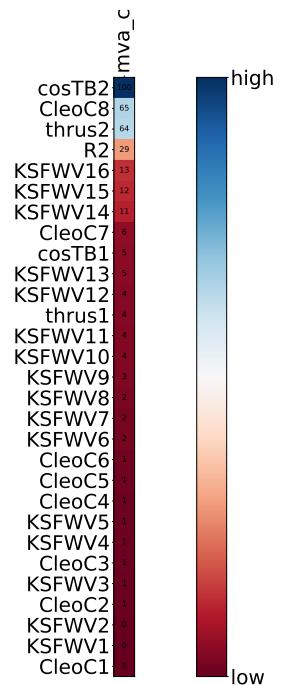
This section contains an overview of the importance and correlation of the variables used by the classifiers. And distribution plots of the variables on the independent dataset. The distributions are normed for signal and background separately, and only the region ± 3 sigma around the mean is shown.

Table 2: Abbreviations of variables

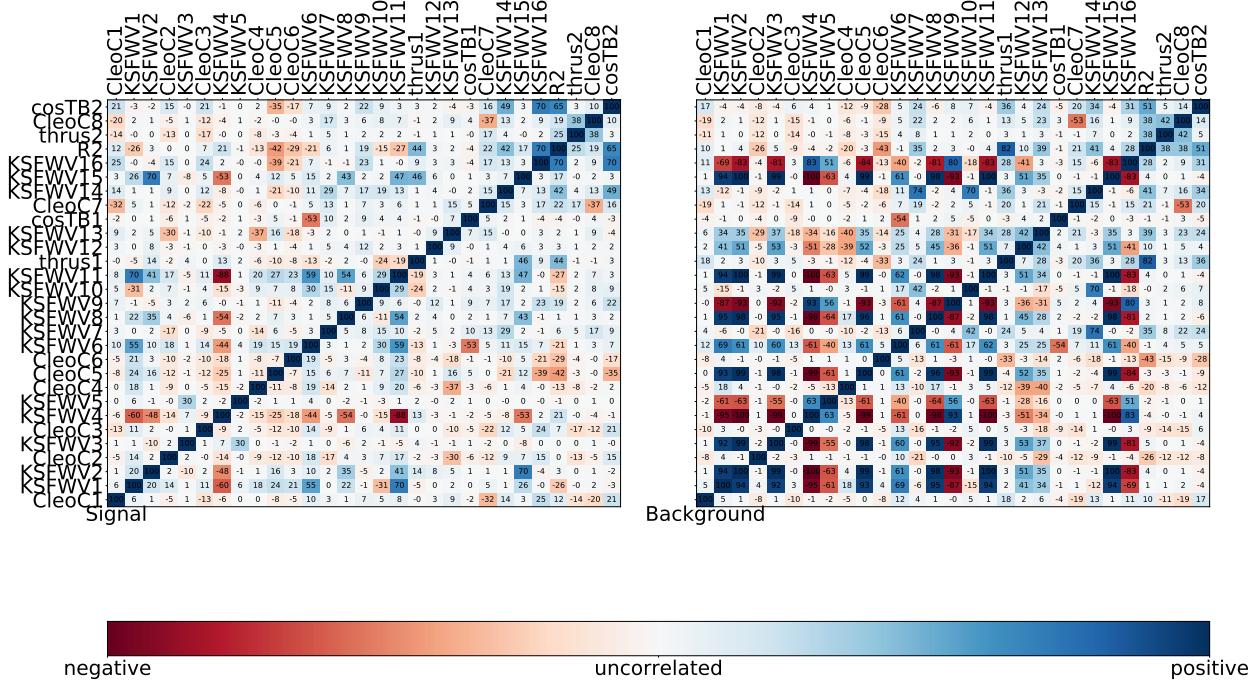
Variable	Abbreviation
----------	--------------

CleoConeCS(3)	CleoC1
KSFVariables(hso00)	KSFV1
KSFVariables(hoo4)	KSFV2
CleoConeCS(5)	CleoC2
KSFVariables(hoo3)	KSFV3
CleoConeCS(4)	CleoC3
KSFVariables(mm2)	KSFV4
KSFVariables(hoo1)	KSFV5
CleoConeCS(6)	CleoC4
CleoConeCS(8)	CleoC5
CleoConeCS(7)	CleoC6
KSFVariables(et)	KSFV6
KSFVariables(hso14)	KSFV7
KSFVariables(hso20)	KSFV8
KSFVariables(hso22)	KSFV9
KSFVariables(hso10)	KSFV10
KSFVariables(hoo0)	KSFV11
thrustOm	thrus1
KSFVariables(hso24)	KSFV12
KSFVariables(hso04)	KSFV13
cosTBz	cosTB1
CleoConeCS(2)	CleoC7
KSFVariables(hso12)	KSFV14
KSFVariables(hoo2)	KSFV15
KSFVariables(hso02)	KSFV16
R2	R2
thrustBm	thrus2
CleoConeCS(1)	CleoC8
cosTBTO	cosTB2

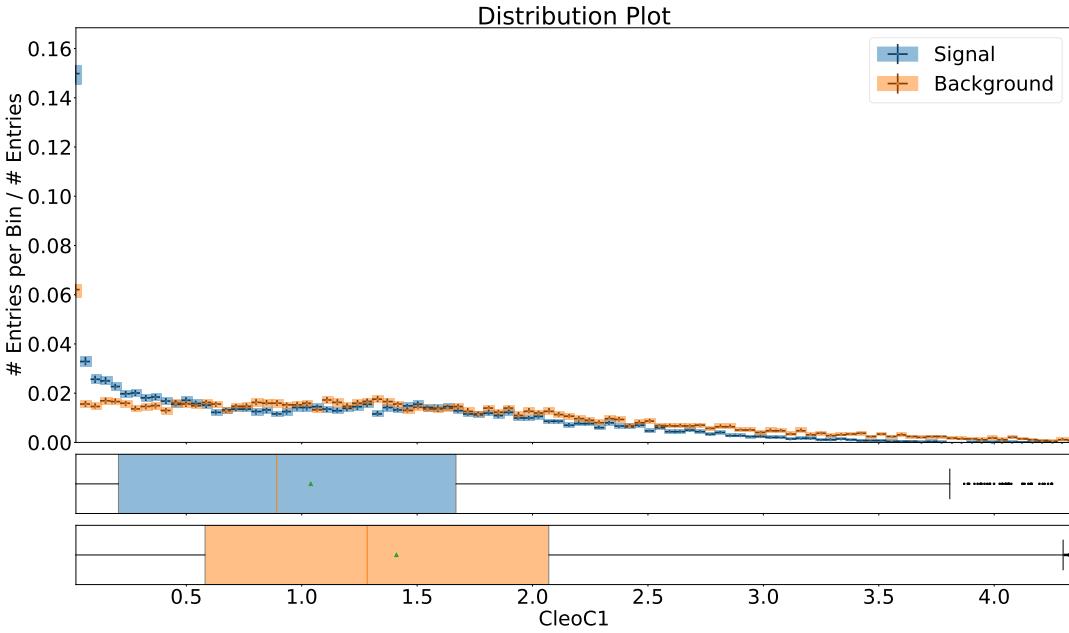
2.1 Importance



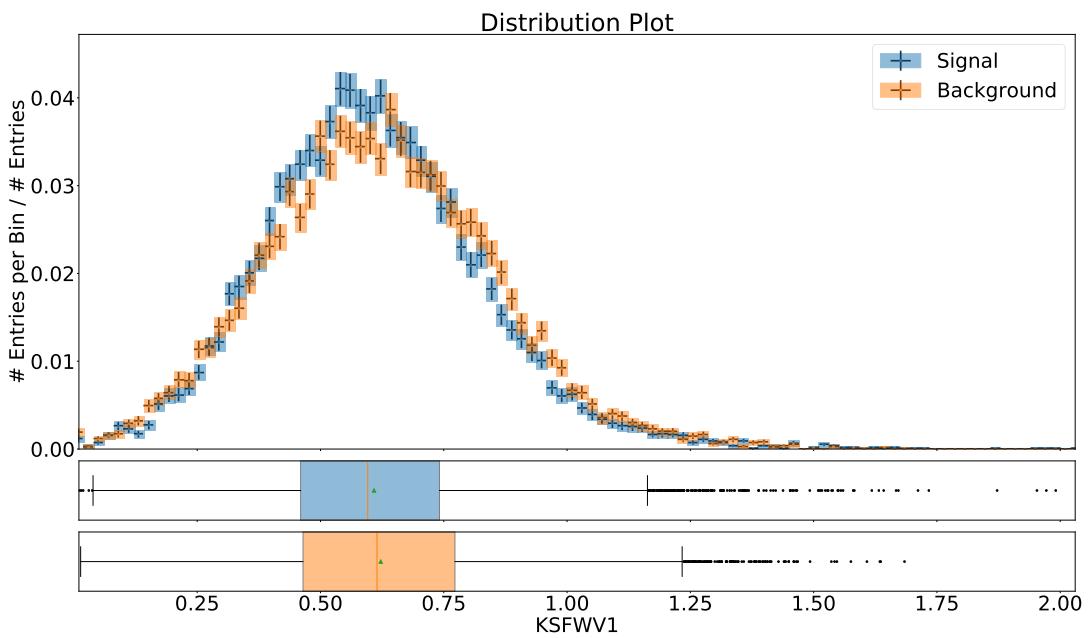
2.2 Correlation



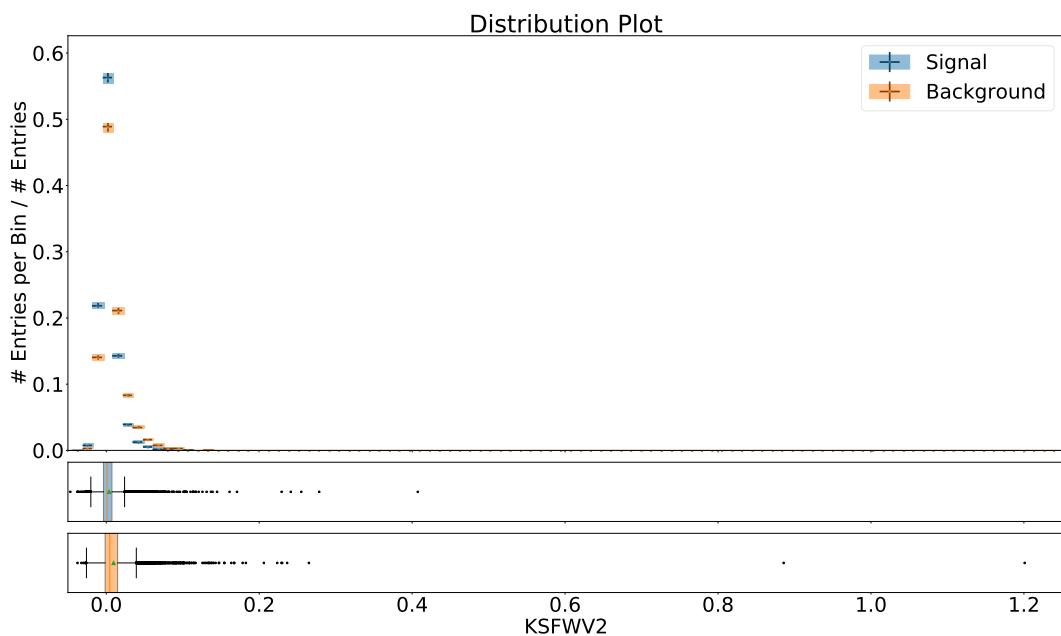
2.3 CleoConeCS(3)



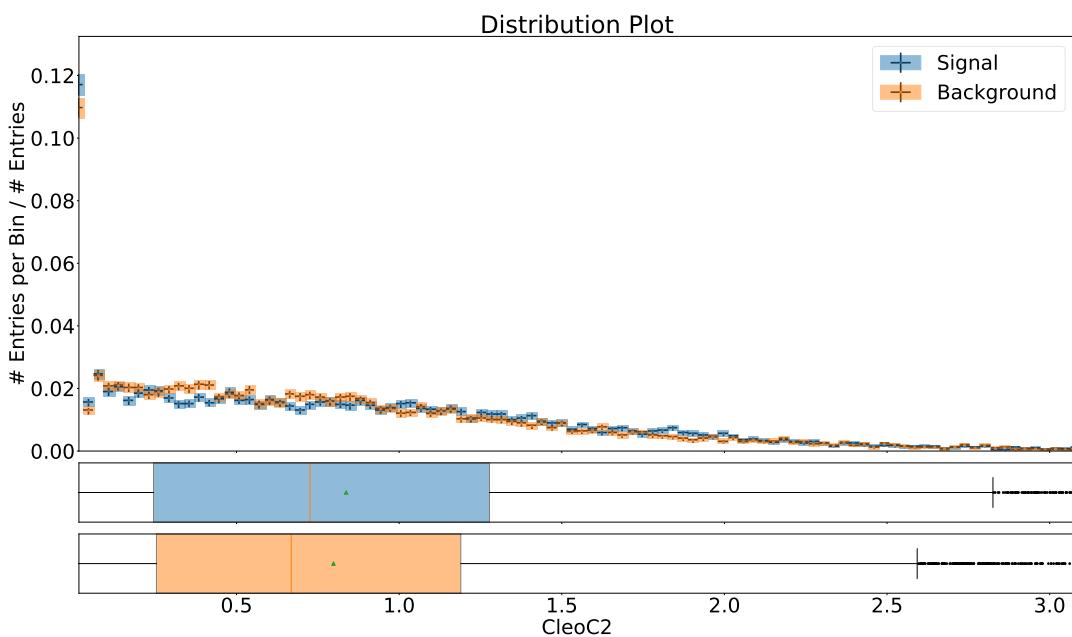
2.4 KSFVVariables(hso00)



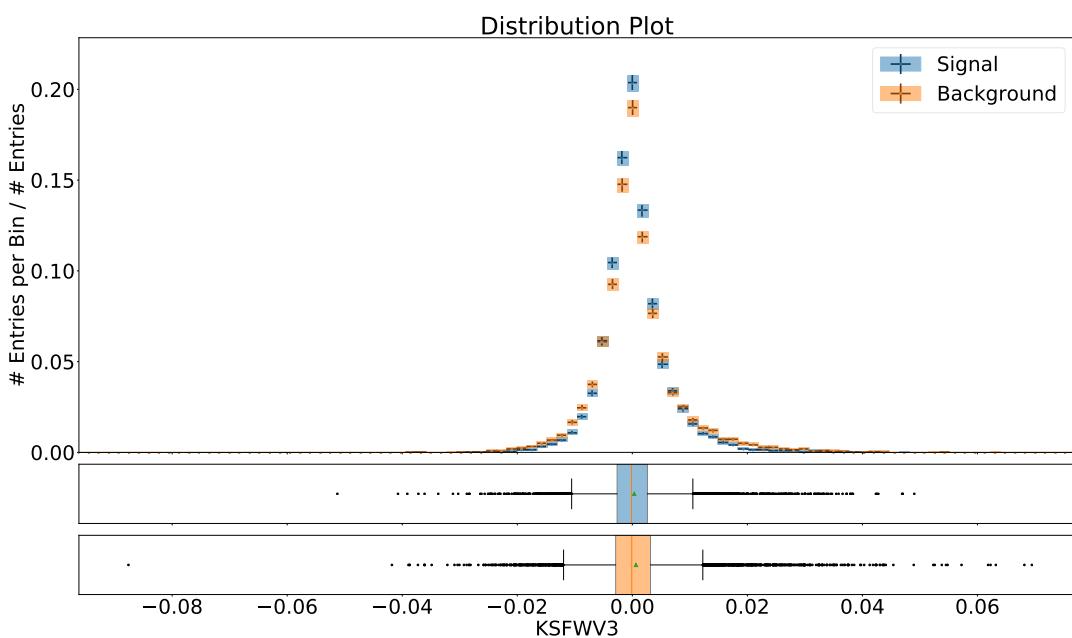
2.5 KSFVVariables(hoo4)



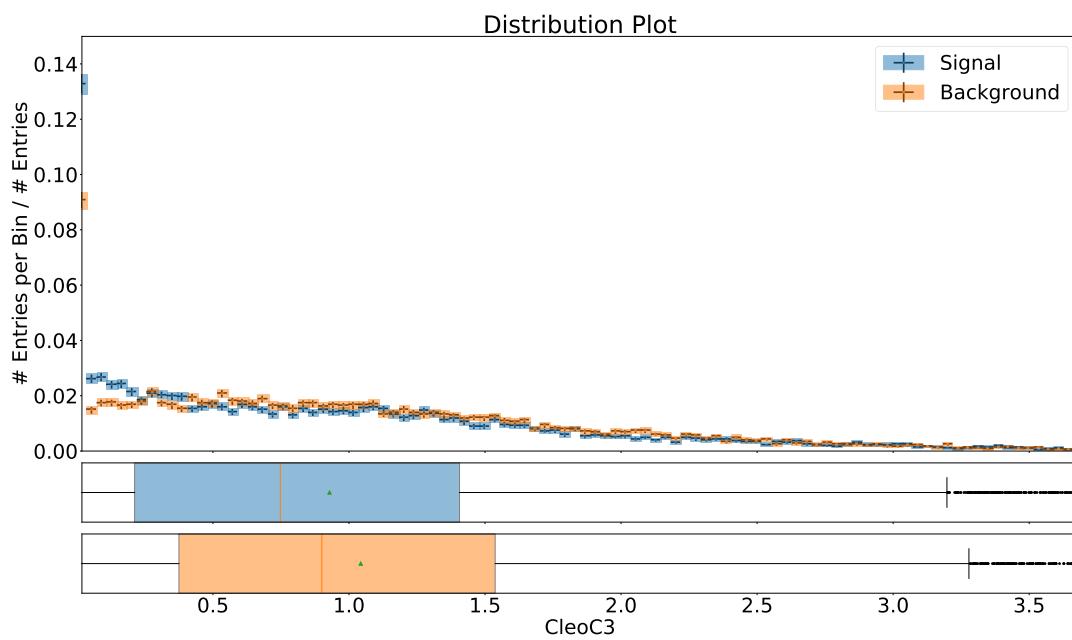
2.6 CleoConeCS(5)



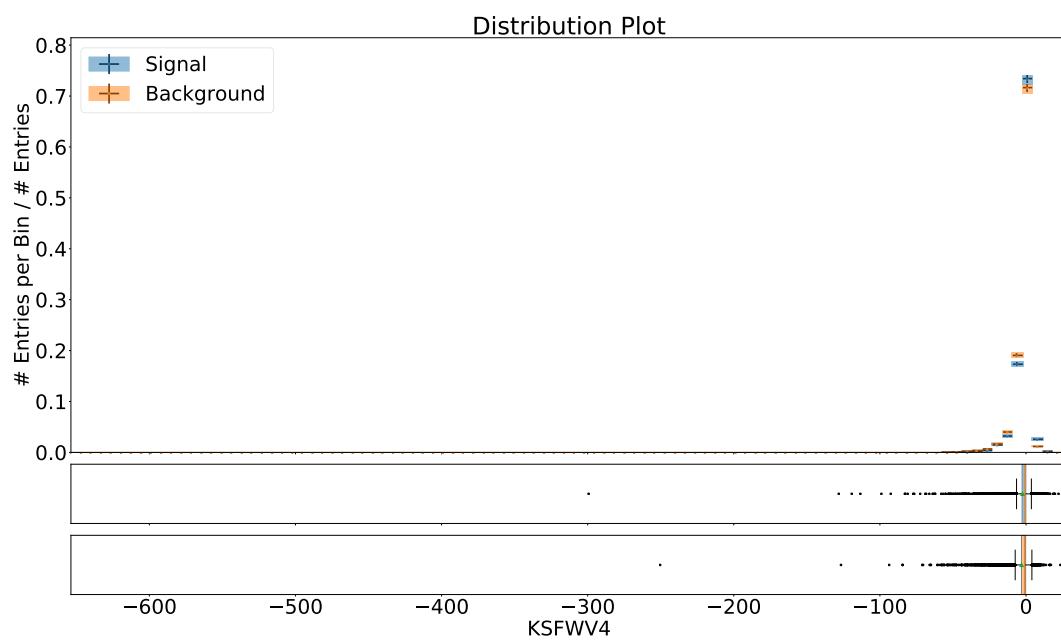
2.7 KSFVVariables(hoo3)



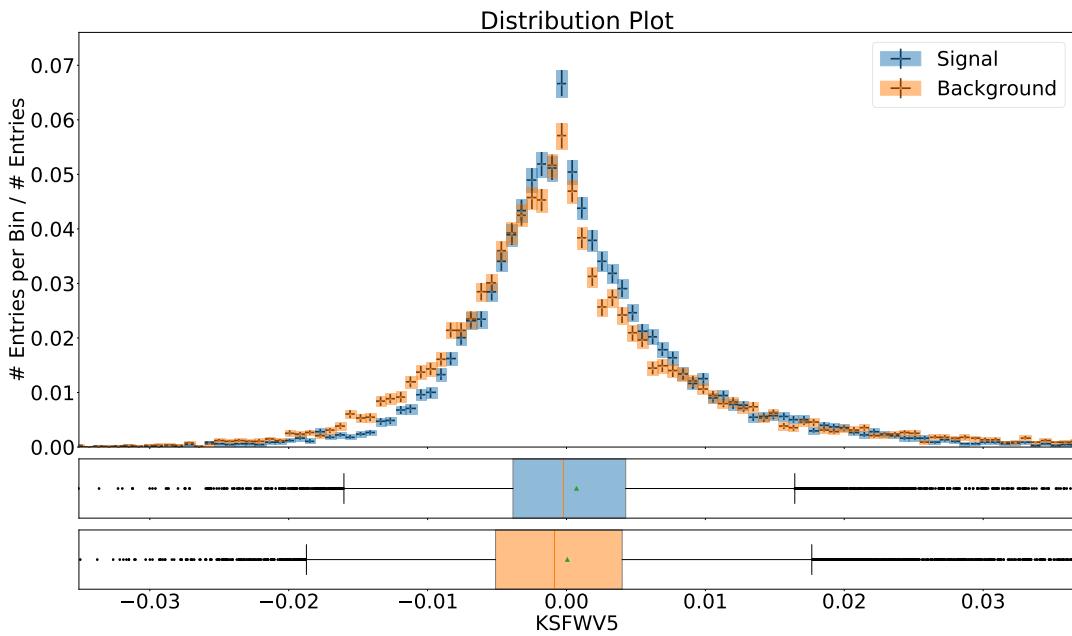
2.8 CleoConeCS(4)



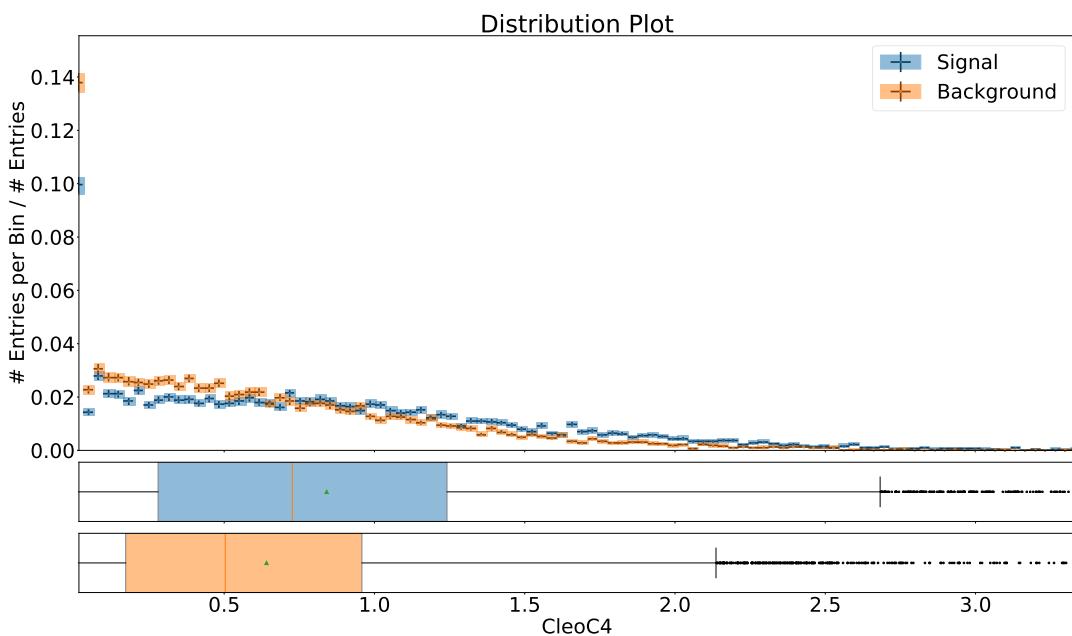
2.9 KSFVVariables(mm2)



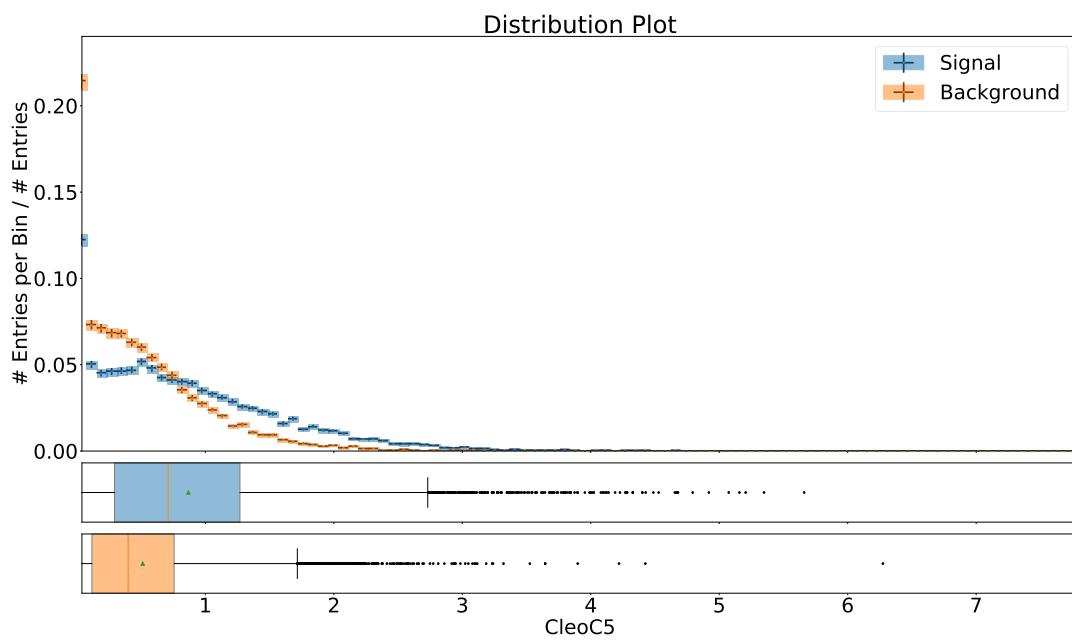
2.10 KSFVWVariables(hoo1)



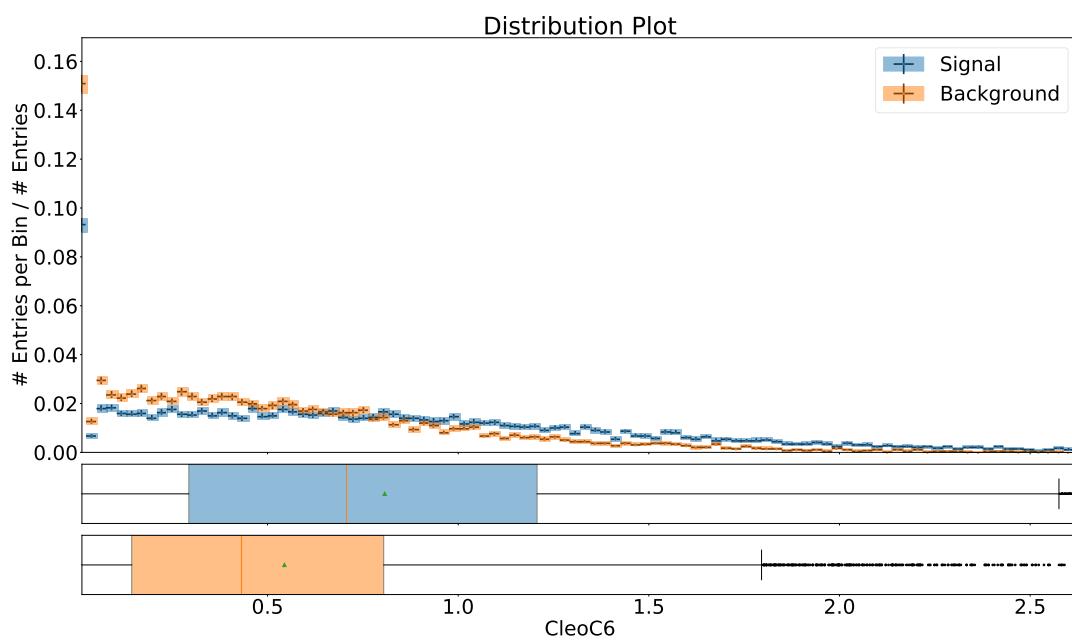
2.11 CleoConeCS(6)



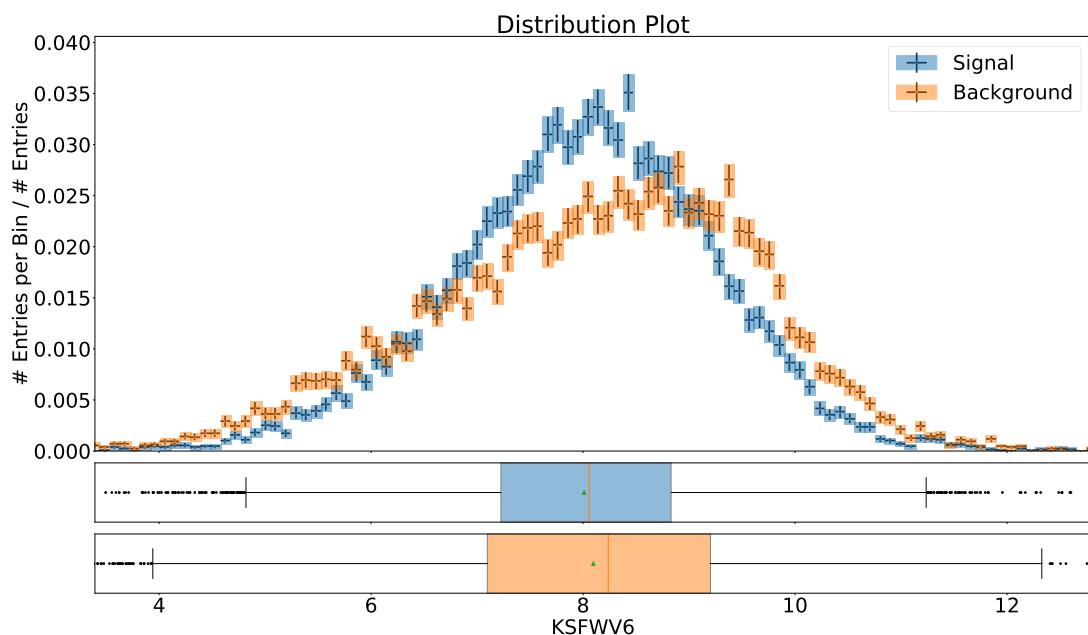
2.12 CleoConeCS(8)



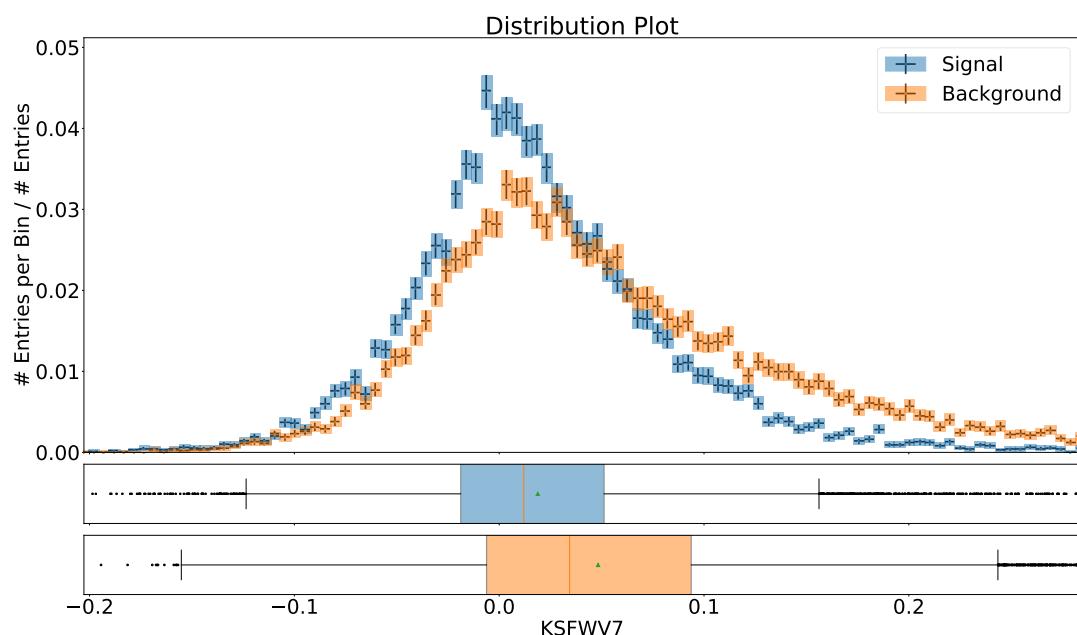
2.13 CleoConeCS(7)



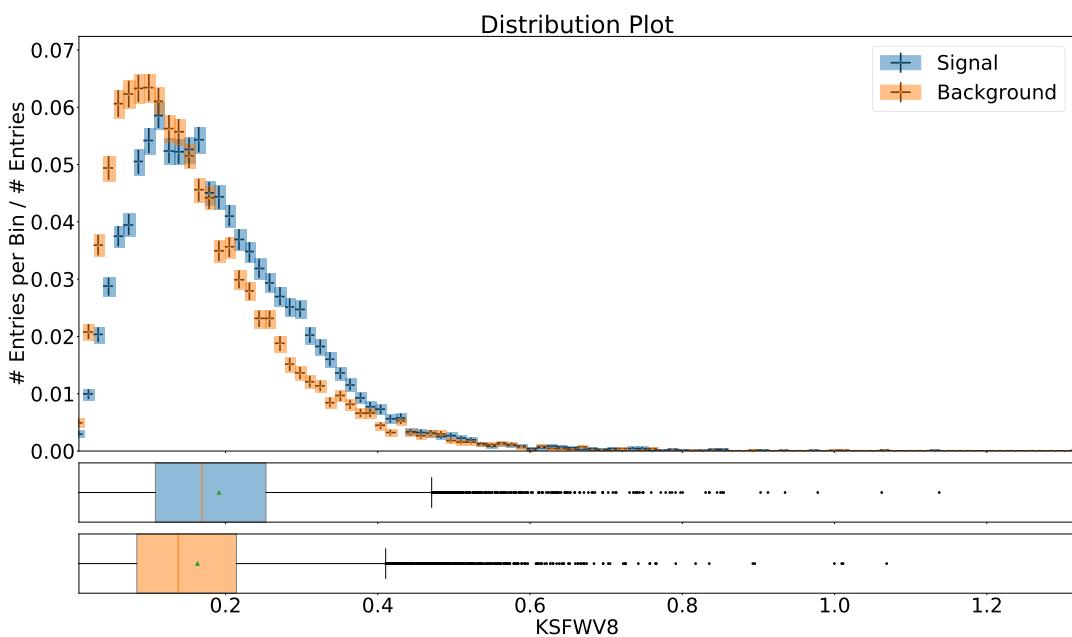
2.14 KSFVVariables(et)



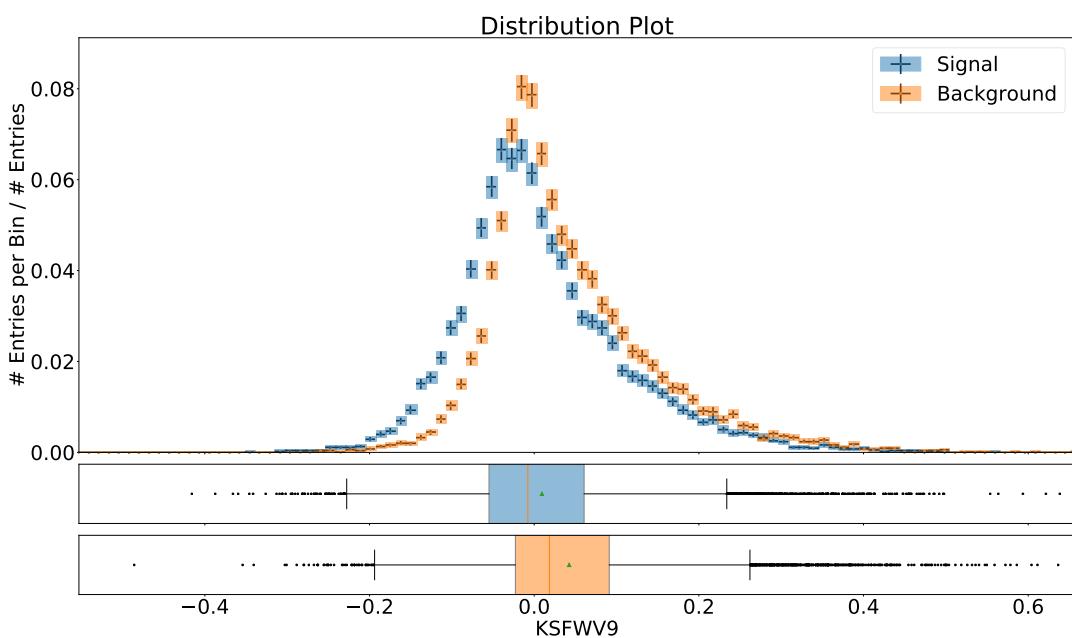
2.15 KSFVVariables(hso14)



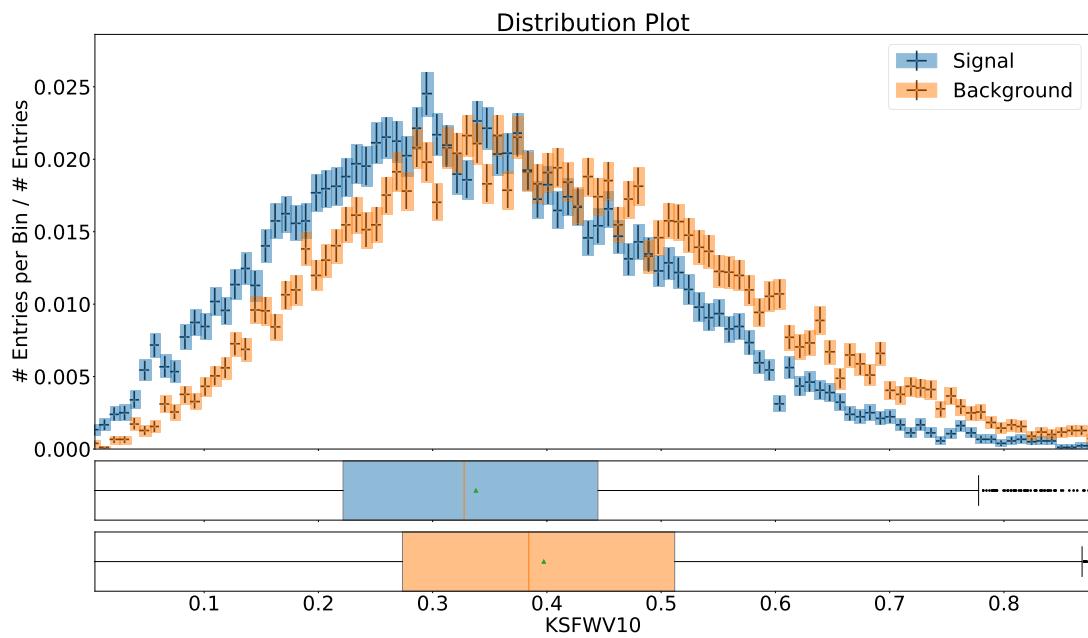
2.16 KSFVVariables(hso20)



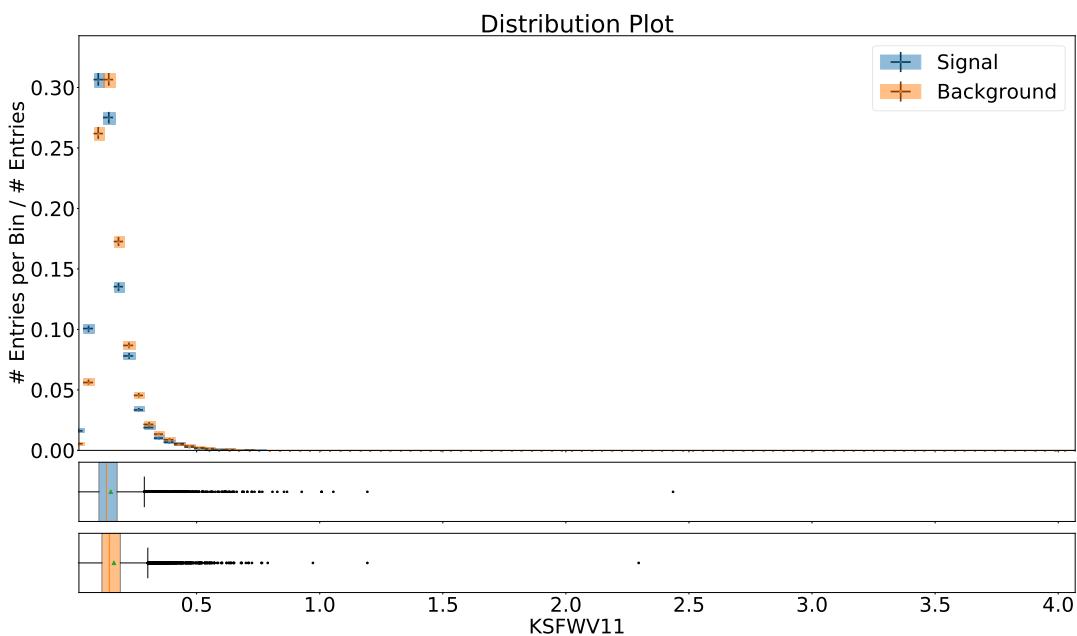
2.17 KSFVVariables(hso22)



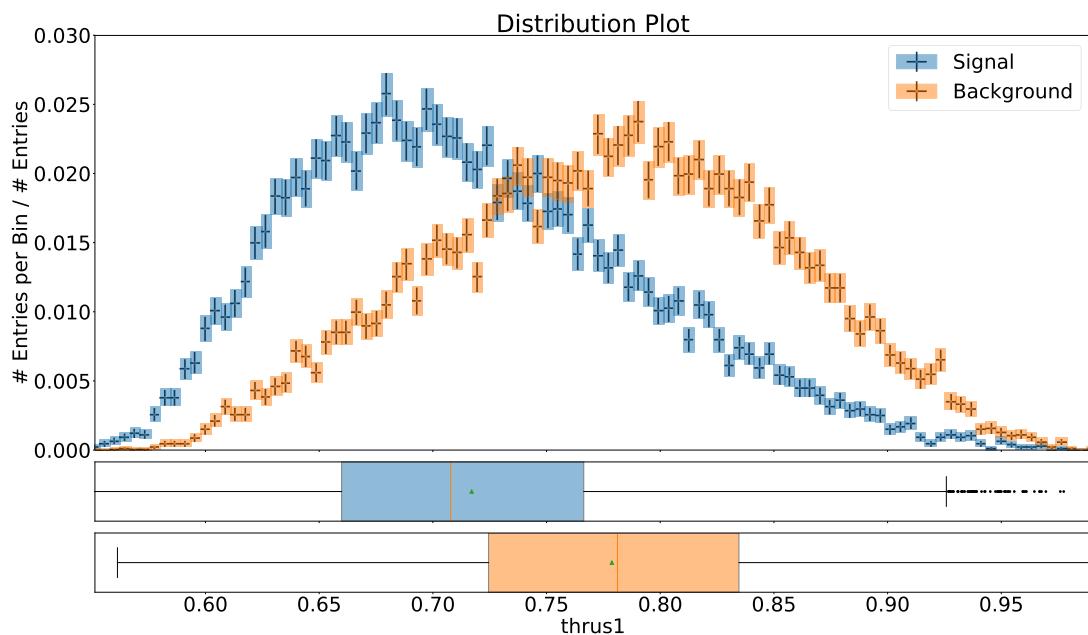
2.18 KSFVVariables(hso10)



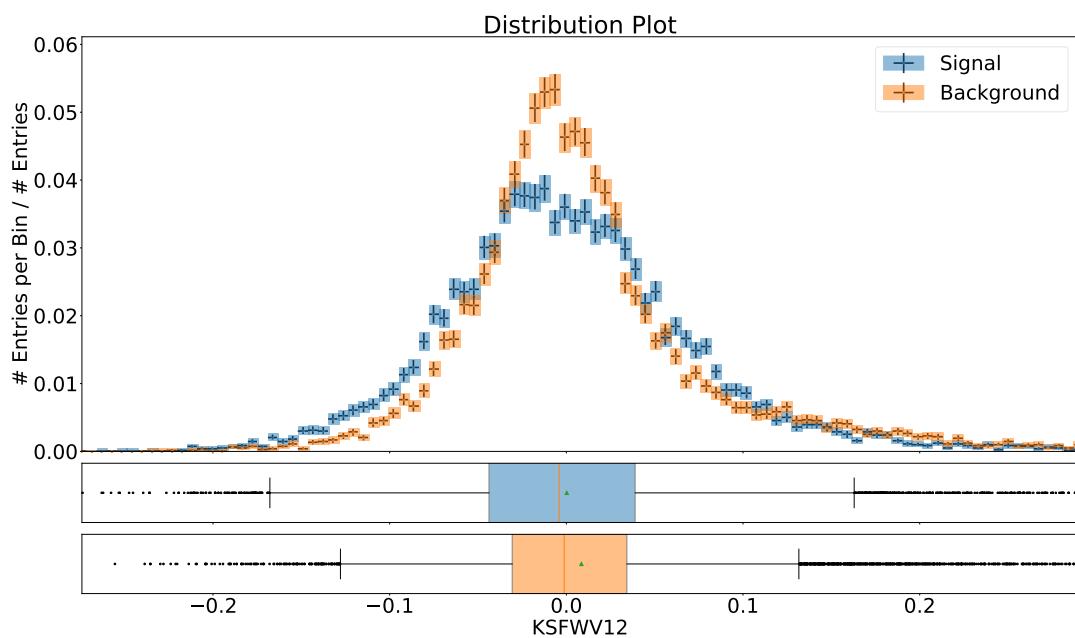
2.19 KSFVVariables(hoo0)



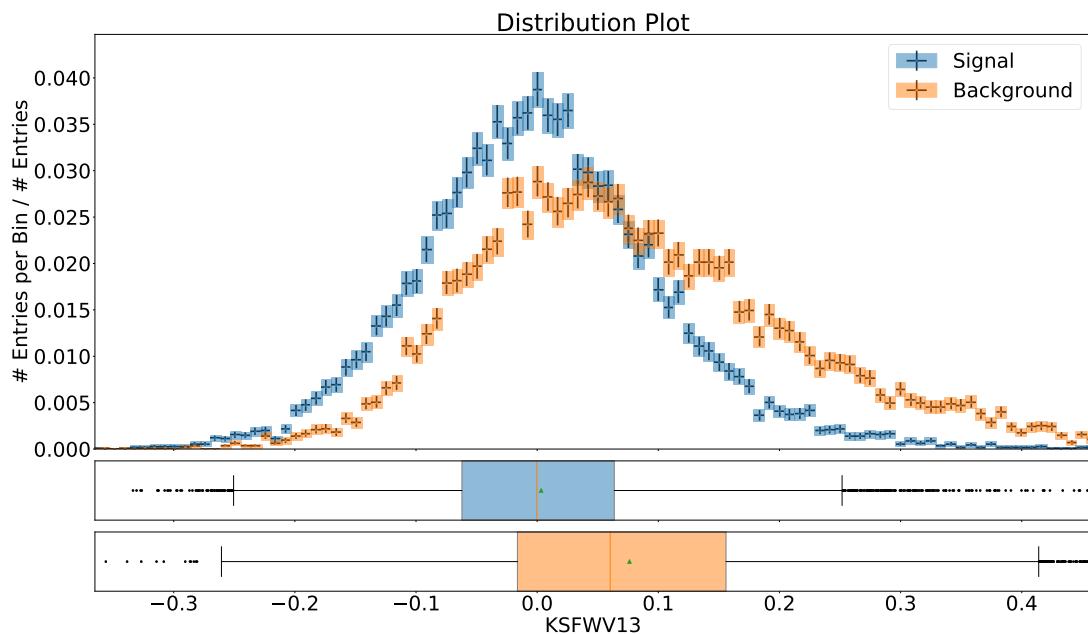
2.20 thrustOm



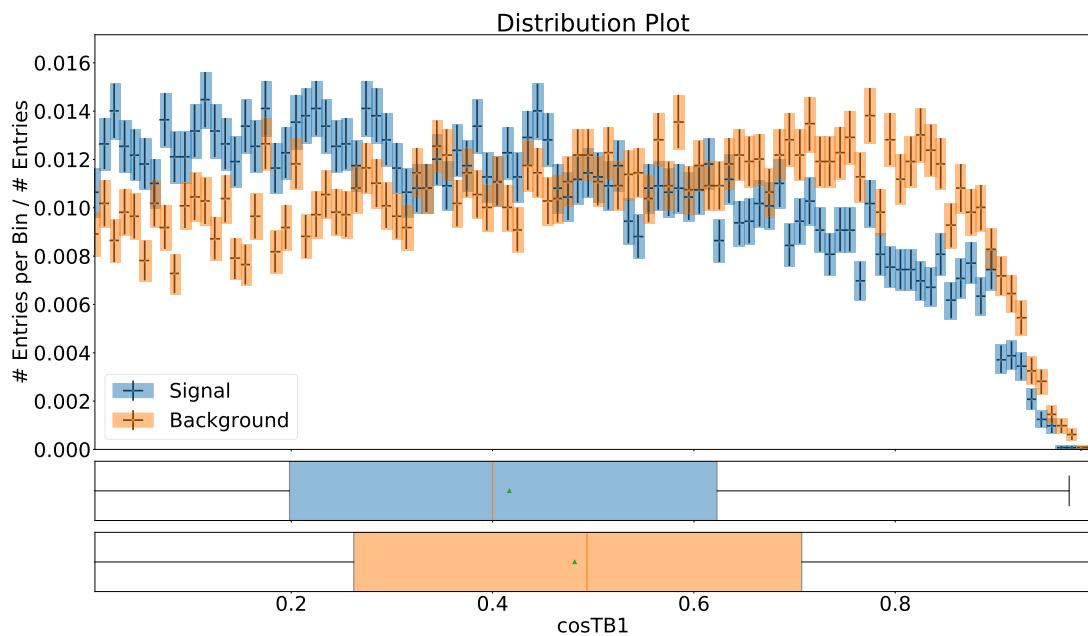
2.21 KSFVVariables(hso24)



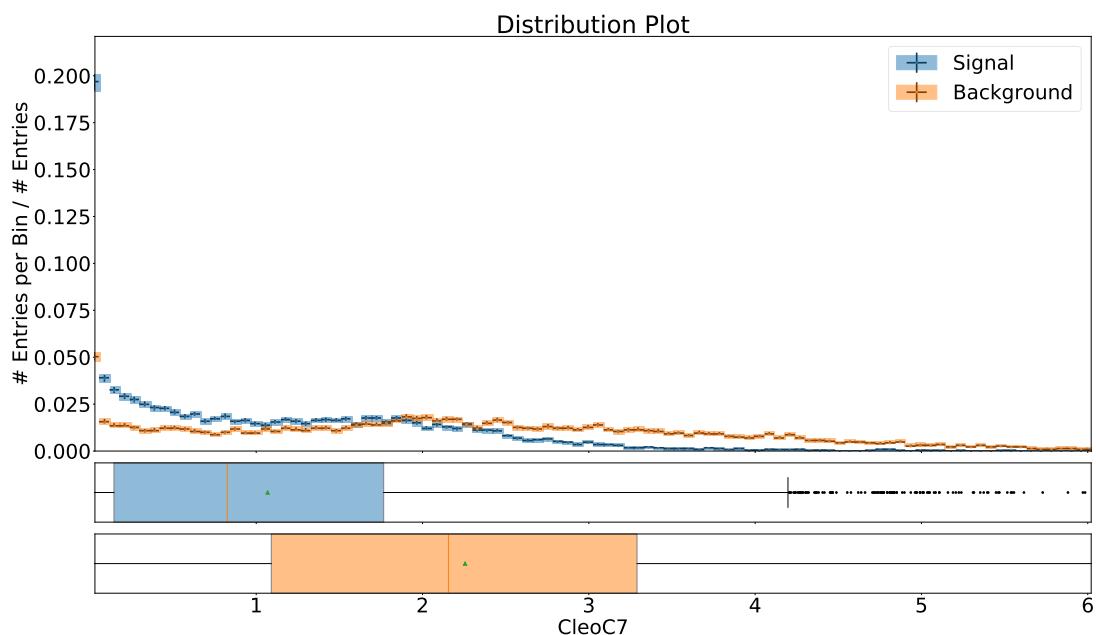
2.22 KSFVWVariables(hso04)



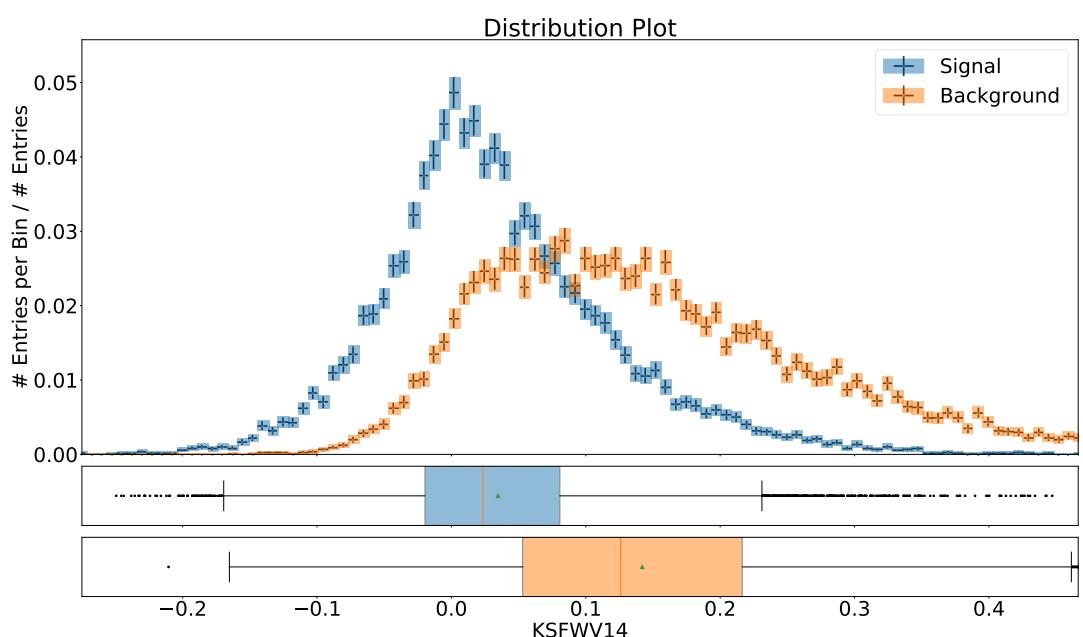
2.23 cosTBz



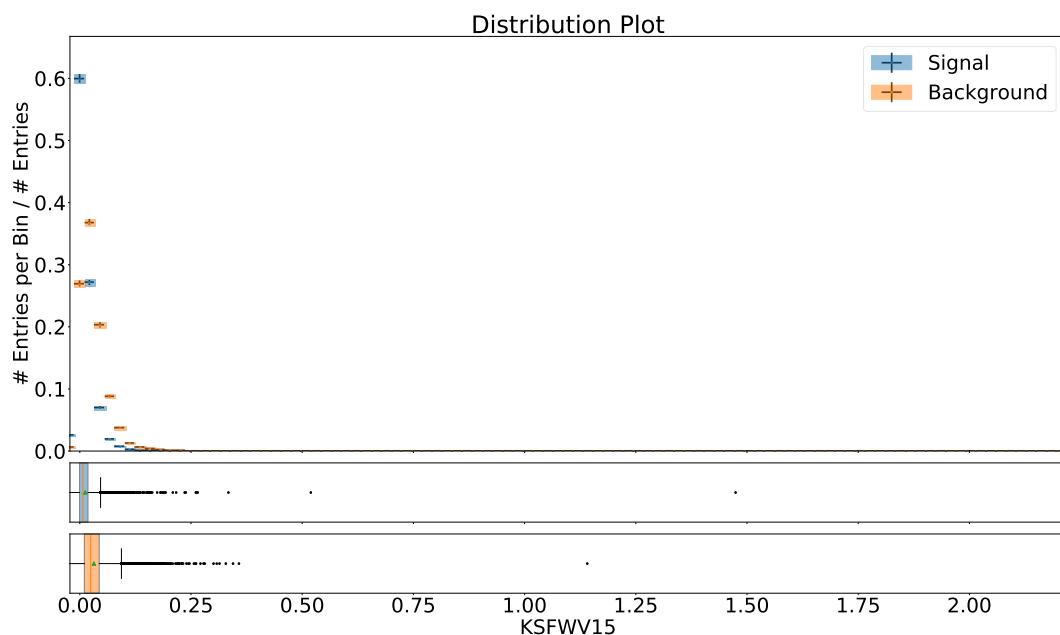
2.24 CleoConeCS(2)



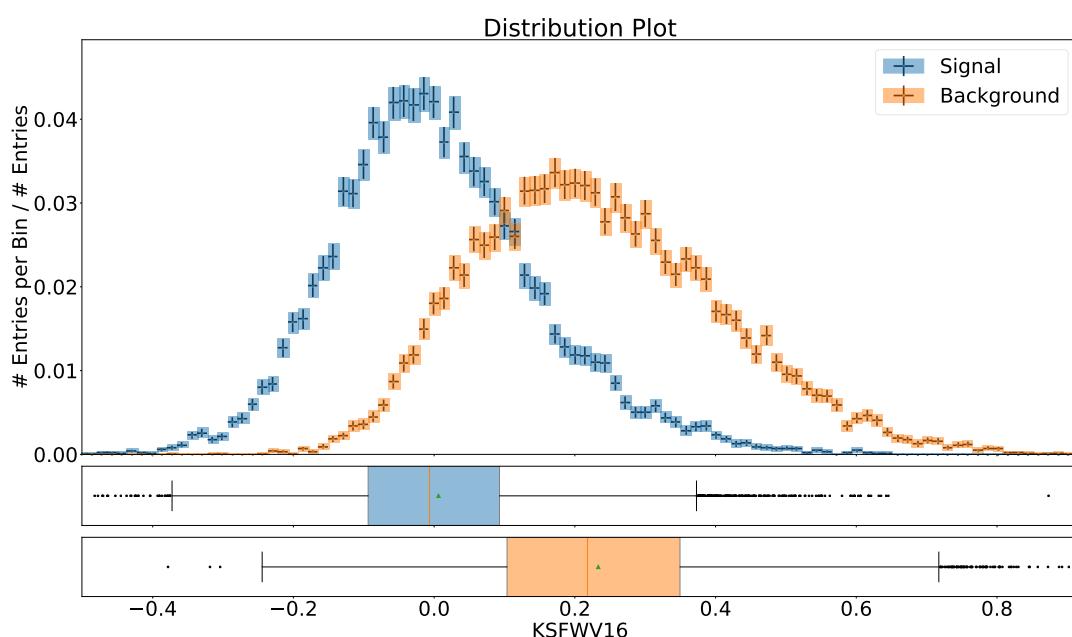
2.25 KSFVVariables(hso12)



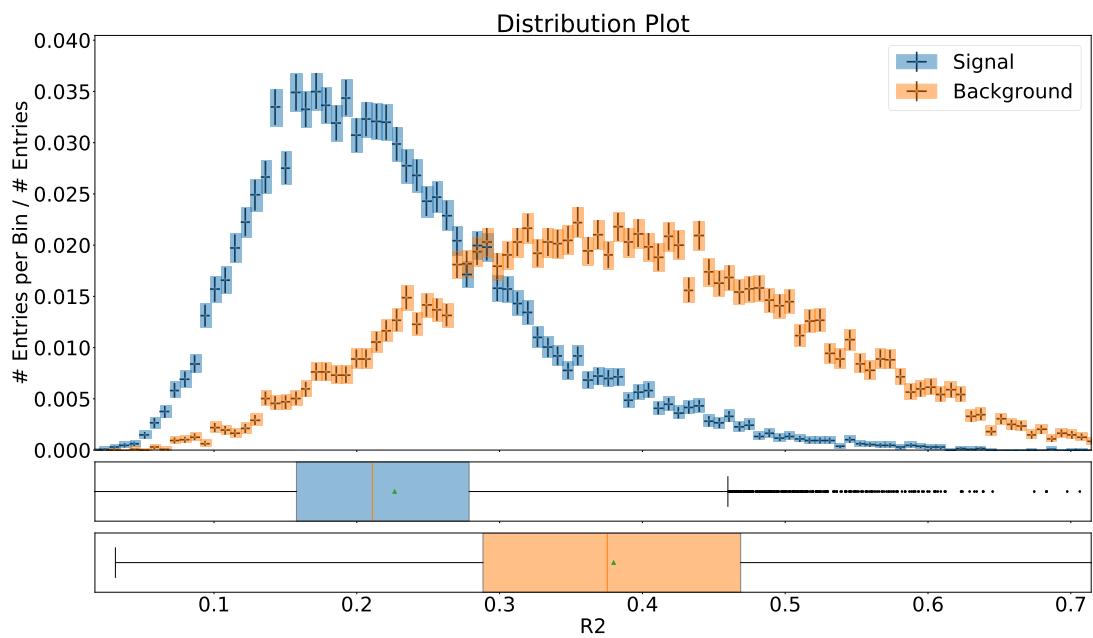
2.26 KSFVVariables(hoo2)



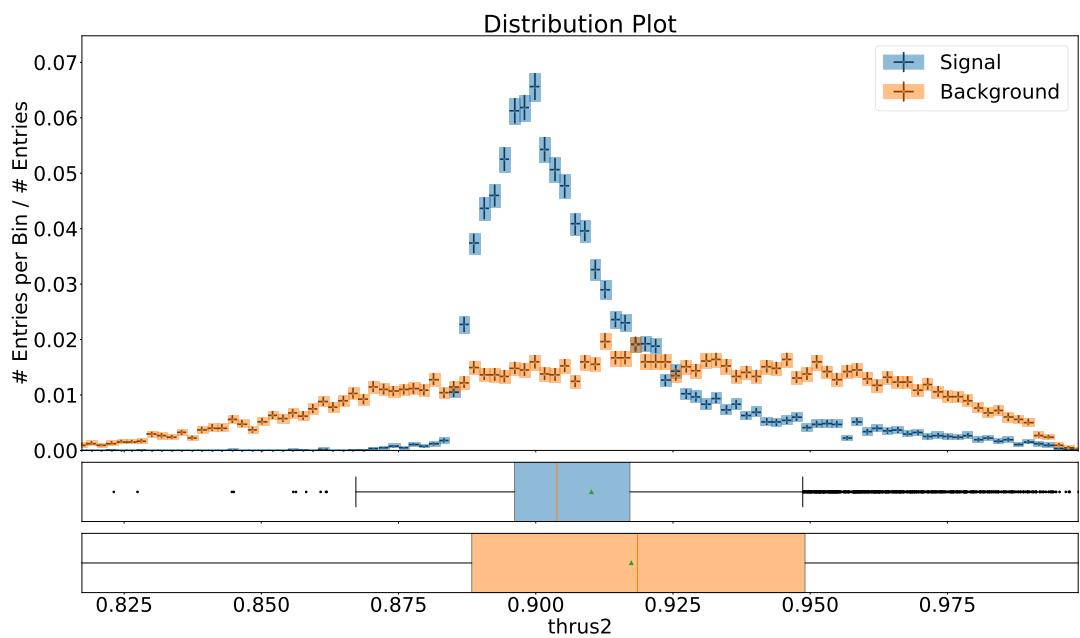
2.27 KSFVVariables(hso02)



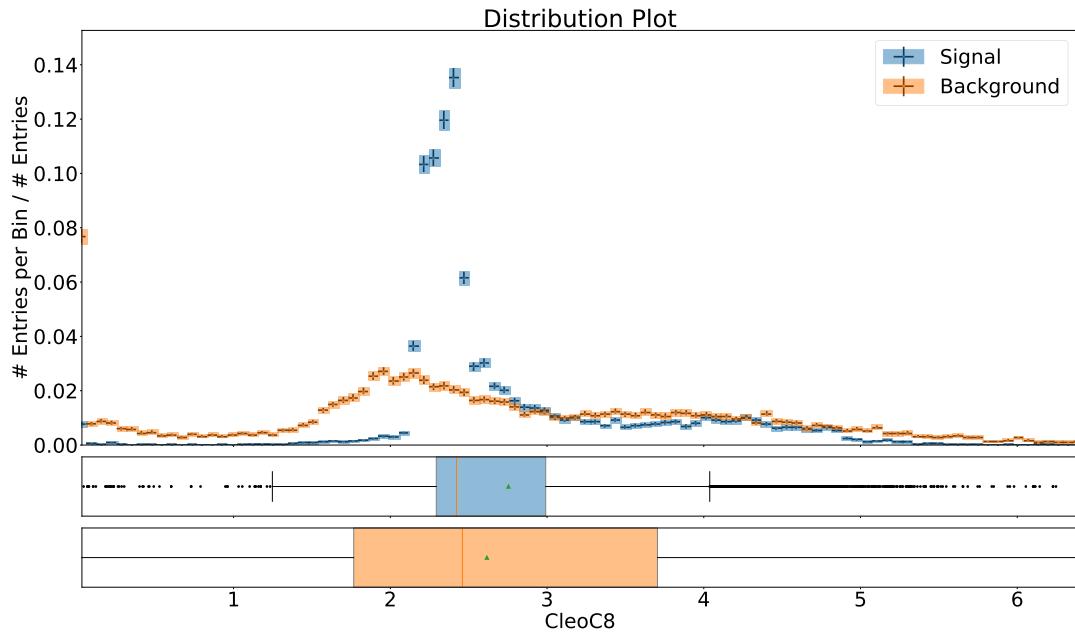
2.28 R2



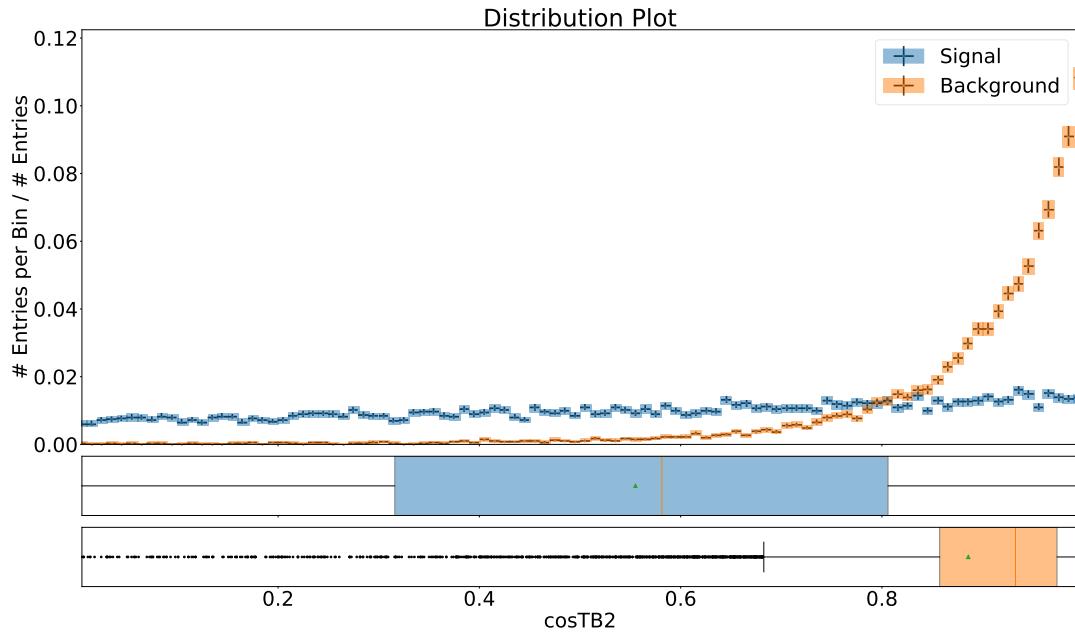
2.29 thrustBm



2.30 CleoConeCS(1)



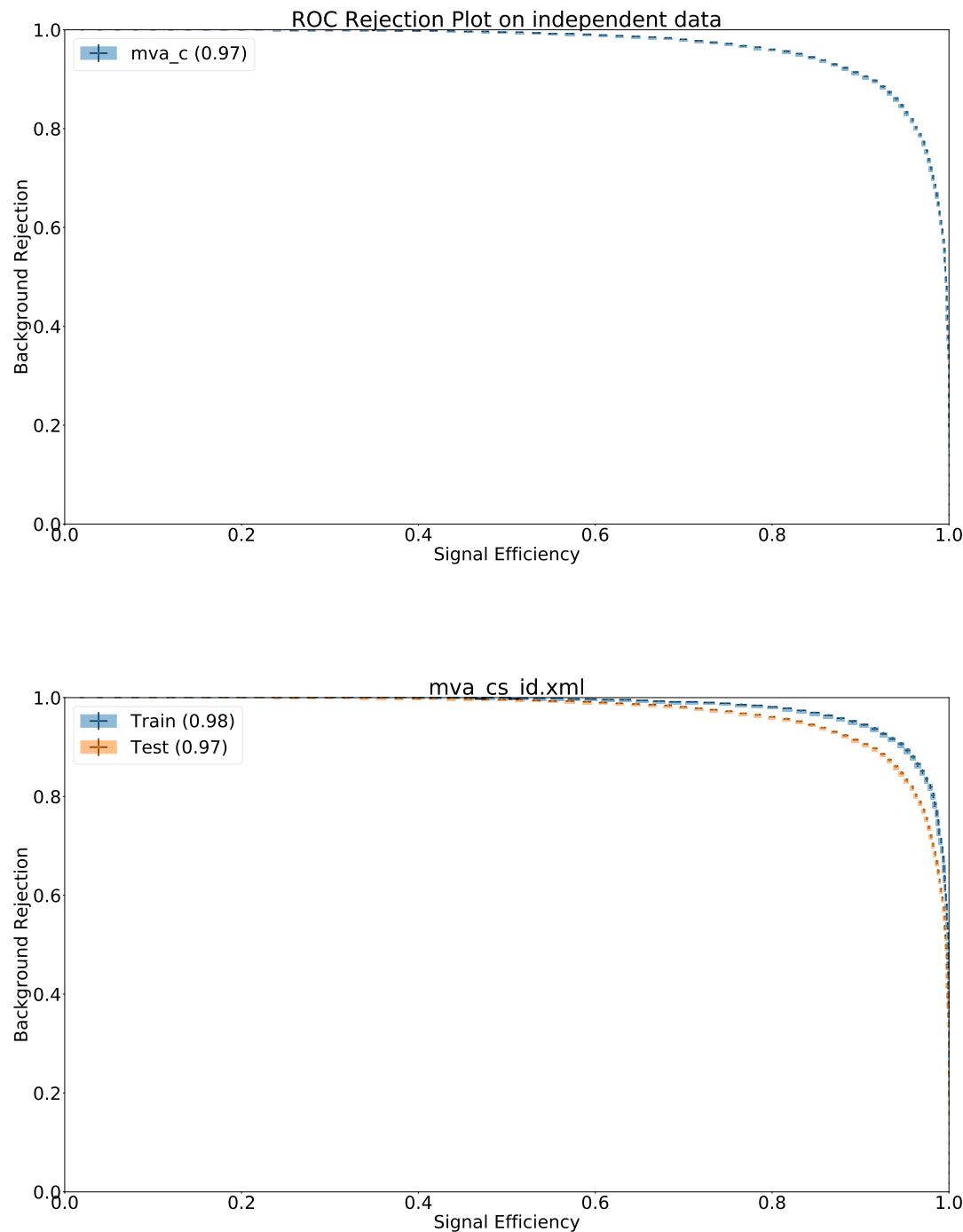
2.31 cosTBTO



3 Classifier Plot

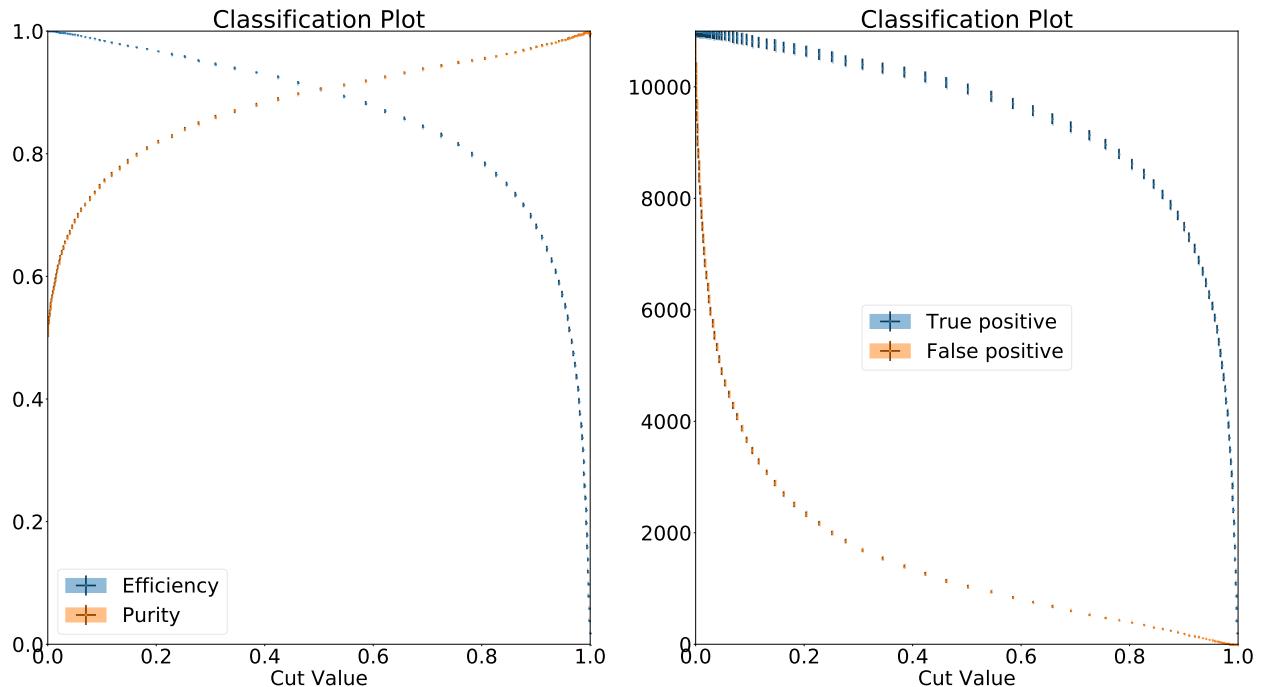
This section contains the receiver operating characteristics (ROC), purity projection, ...of the classifiers on training and independent data. The legend of each plot contains the shortened identifier and the area under the ROC curve in parenthesis.

4 ROC Plot



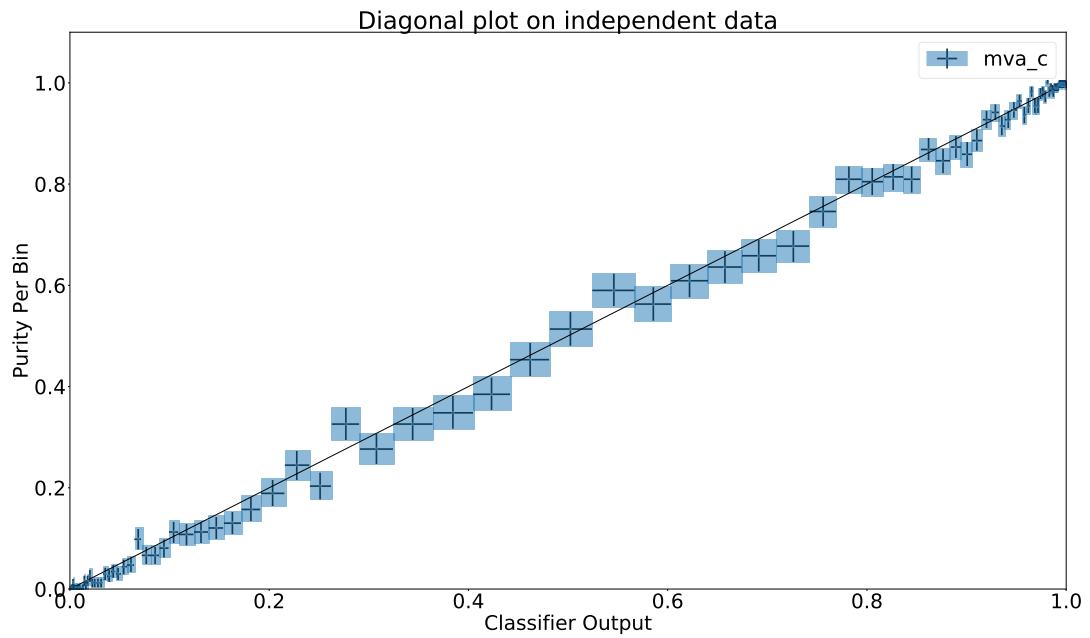
5 Classification Results

5.1 mva_c

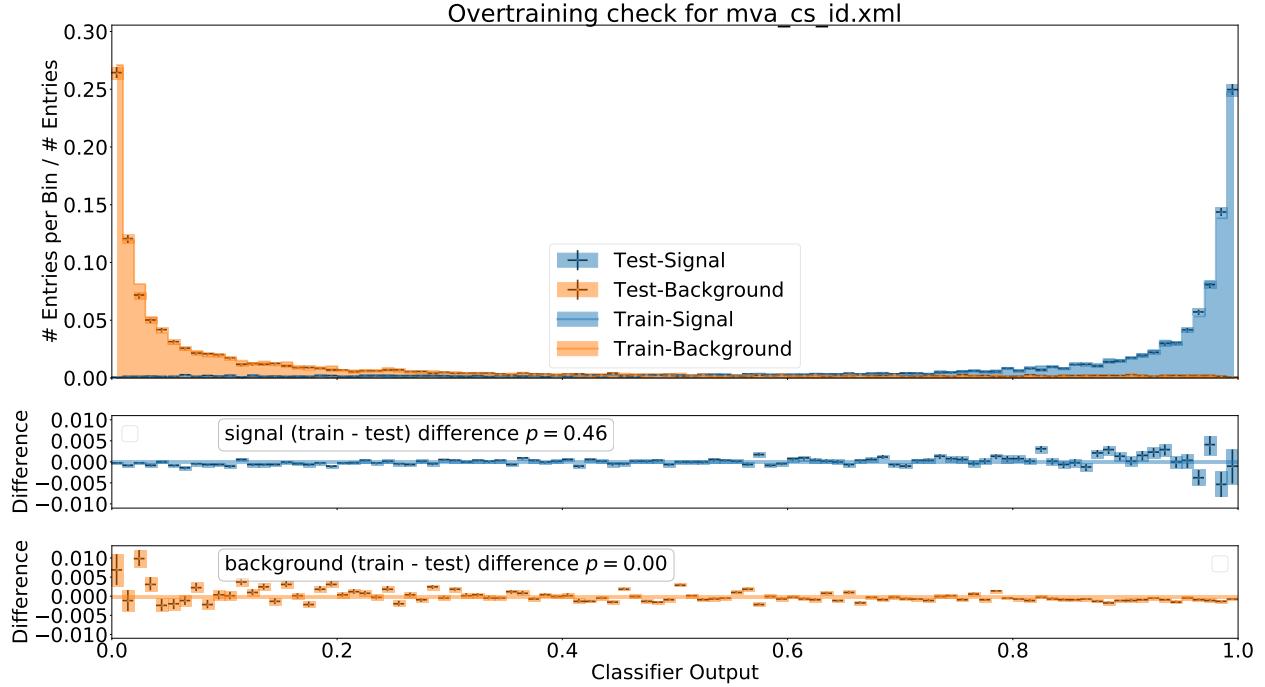


6 Diagonal Plot

6.1 mva_c



6.2 Overtraining Plot



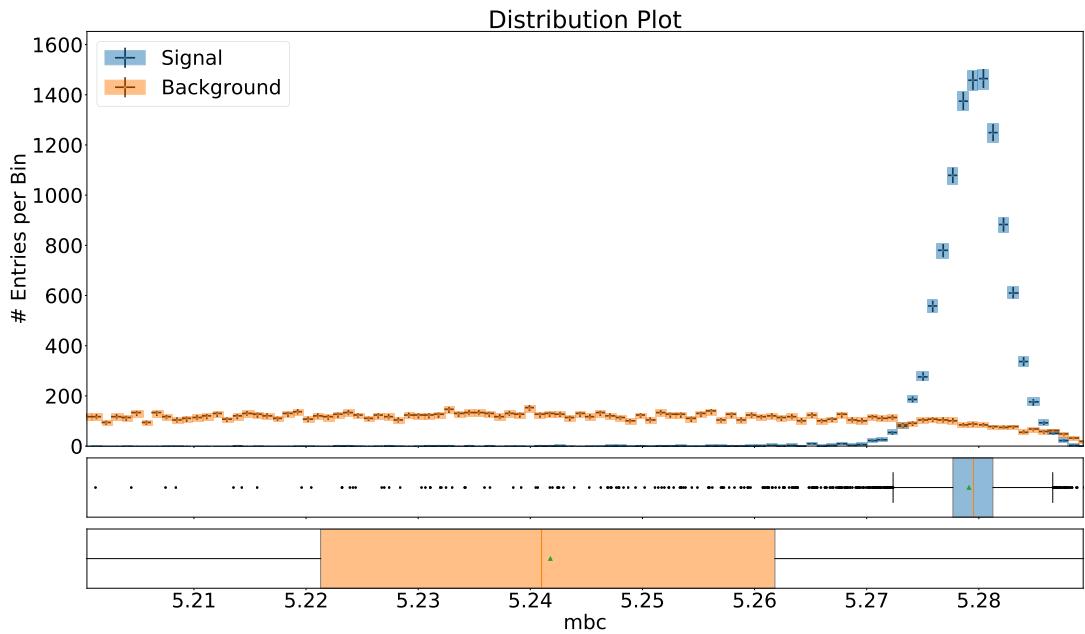
7 Spectators

This section contains the distribution and dependence on the classifier 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.xml

