Automatic MVA Evaluation

Thomas Keck
Moritz Gelb
Nils Braun

March 26, 2019

Abstract

Evaluation plots

Contents

1 Classifiers **3**
1.1 ../n_events_training=5000/trackfindingcdc_TrackQualityIndicator.weights.xml 3

2 Variables **3**
2.1 Importance 4
2.2 Correlation 5
2.3 has_matching_segment 5
2.4 adc_variance 6
2.5 adc_max 6
2.6 adc_min 7
2.7 empty_s_max 7
2.8 drift_length_max 8
2.9 adc_sum 8
2.10 s_range 9
2.11 empty_s_min 9
2.12 drift_length_sum 10
2.13 empty_s_sum 10
2.14 adc_mean 11
2.15 drift_length_variance 11
2.16 drift_length_min 12
2.17 drift_length_mean 12
2.18 sz_slope 13
2.19 empty_s_variance 13
2.20 empty_s_mean 14
2.21 pt 14
2.22 size 15

3 Classifier Plot **15**

4 ROC Plot **15**

5 Classification Results **16**
5.1 ../n_ 16

6 Diagonal Plot **16**
6.1 ../n_ 16

7 Spectators **16**

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
../n_events_training=5000/trackfindingcdc_TrackQualityIndicator.weights.xml	../n_

1.1 ../n_events_training=5000/trackfindingcdc_TrackQualityIndicator.weights.xml


```
<?xml version="1.0" encoding="utf-8"?>
<method>FastBDT</method>
<weightfile>./git_hash=cd1cfc6b5beb639a4b2839a73e8b8f24bea35572/n_events_training=5000/trackfindingcdc_TrackQualityIndicator.
     weights.xml</weightfile>
<treename>records</treename>
<target_variable>truth</target_variable>
<weight_variable/>
<signal_class>1</signal_class>
<max_events>0</max_events>
<number_feature_variables>20</number_feature_variables>
<variable0>size</variable0>
<variable1>pt</variable1>
<variable2>sz_slope</variable2>
<variable3>drift_length_mean</variable3>
<variable4>drift_length_variance</variable4>
<variable5>drift_length_max</variable5>
<variable6>drift_length_min</variable6>
<variable7>drift_length_sum</variable7>
<variable8>adc_mean</variable8>
<variable9>adc_variance</variable9>
<variable10>adc_max</variable10>
<variable11>adc_min</variable11>
<variable12>adc_sum</variable12>
<variable13>empty_s_mean</variable13>
<variable14>empty_s_variance</variable14>
<variable15>empty_s_max</variable15>
<variable16>empty_s_min</variable16>
<variable17>empty_s_sum</variable17>
<variable18>has_matching_segment</variable18>
<variable19>s_range</variable19>
<number_spectator_variables>0</number_spectator_variables>
<number_data_files>1</number_data_files>
<datafile0>./git_hash=cd1cfc6b5beb639a4b2839a73e8b8f24bea35572/n_events_training=5000/random_seed=traincdc_0/cdc_qe_records.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

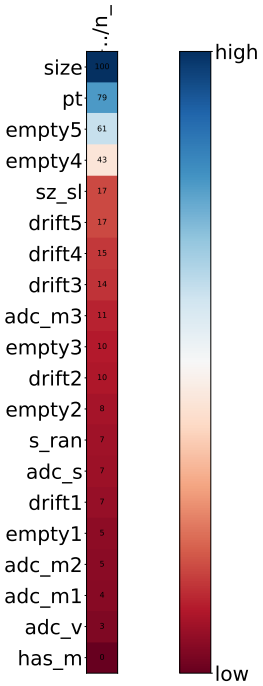
This section contains an overview of the importance and correlation of the variables used by the classifiers. And distribution plots of the variables on the independent dataset. The distributions are normed for signal and background separately, and only the region +- 3 sigma around the mean is shown.

Table 2: Abbreviations of variables

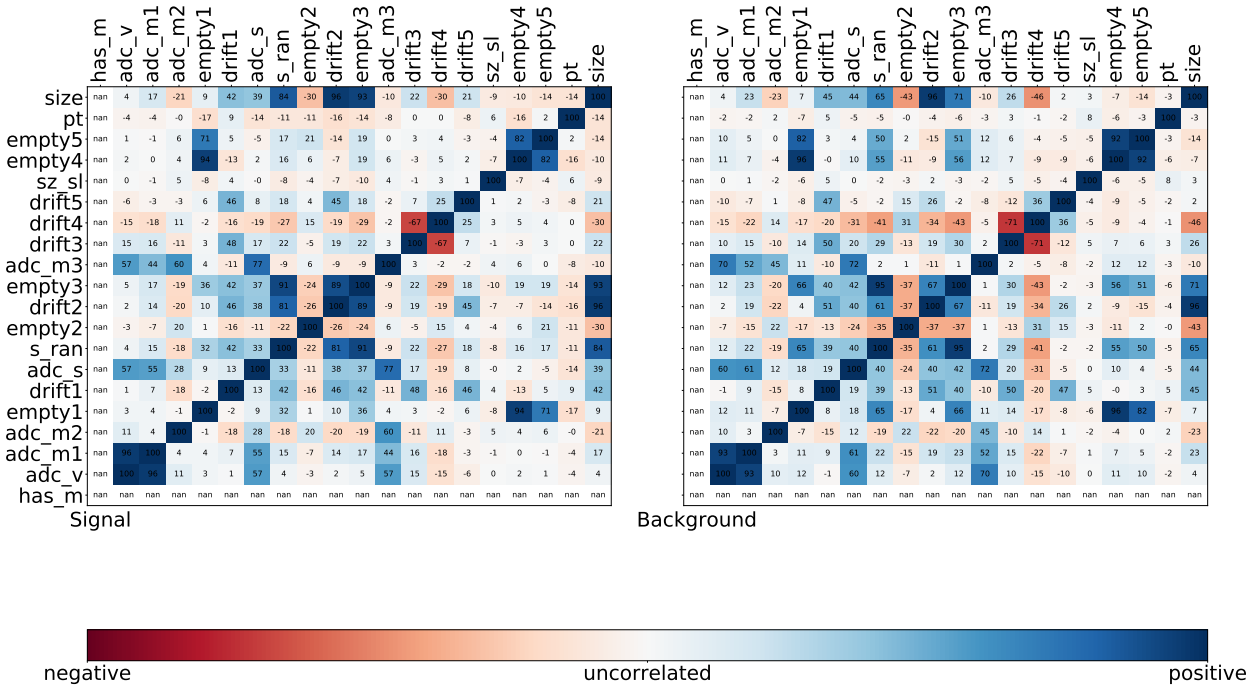
Variable	Abbreviation
has_matching_segment	has_m
adc_variance	adc_v
adc_max	adc_m1
adc_min	adc_m2

empty_s_max	empty1
drift_length_max	drift1
adc_sum	adc_s
s_range	s_ran
empty_s_min	empty2
drift_length_sum	drift2
empty_s_sum	empty3
adc_mean	adc_m3
drift_length_variance	drift3
drift_length_min	drift4
drift_length_mean	drift5
sz_slope	sz_sl
empty_s_variance	empty4
empty_s_mean	empty5
pt	pt
size	size

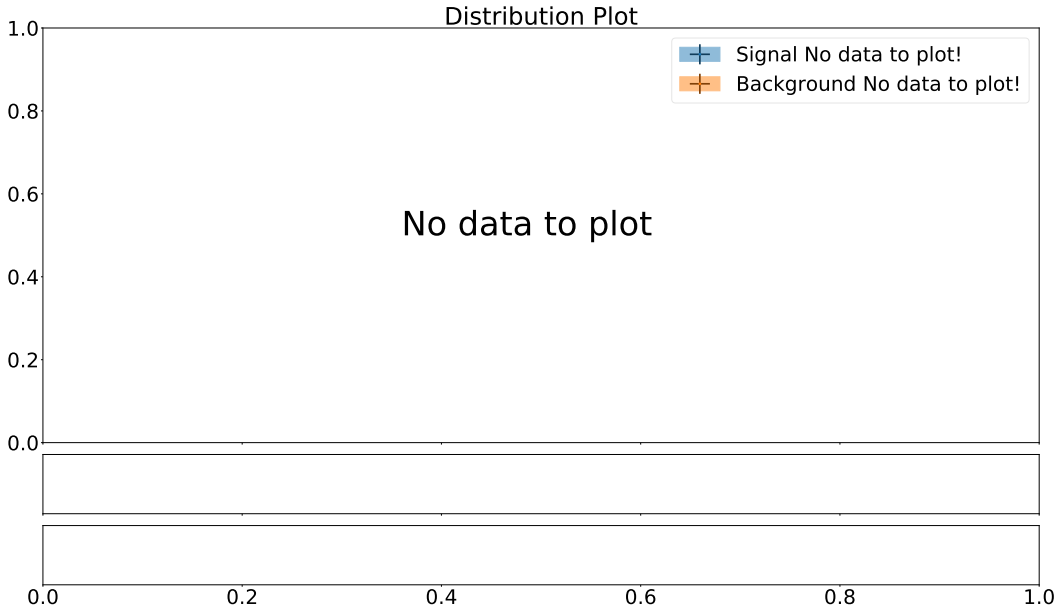
2.1 Importance



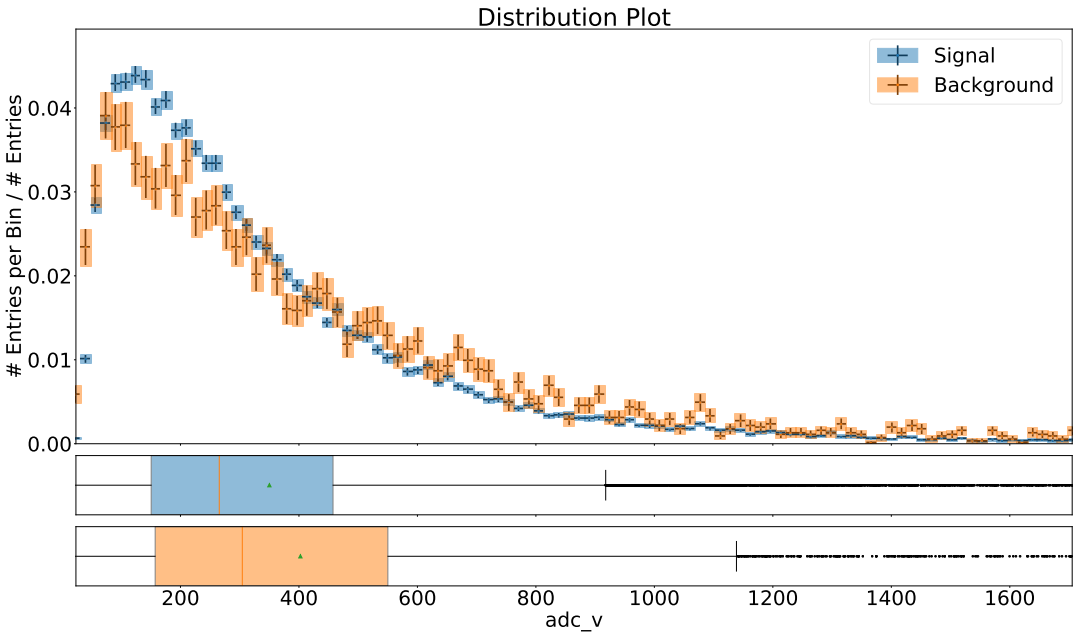
2.2 Correlation



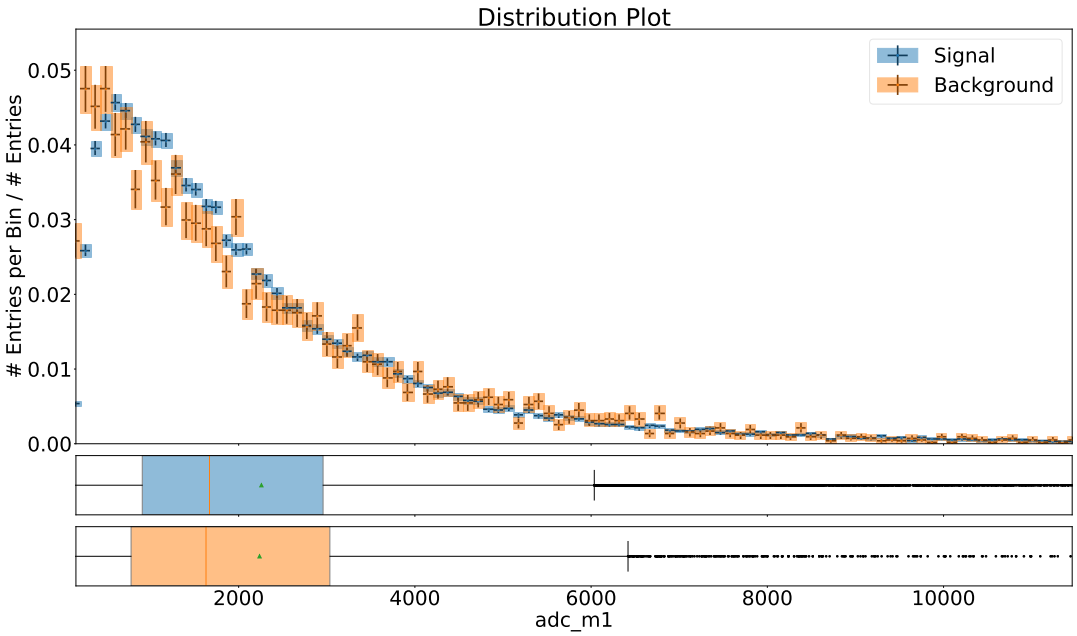
2.3 has_matching_segment



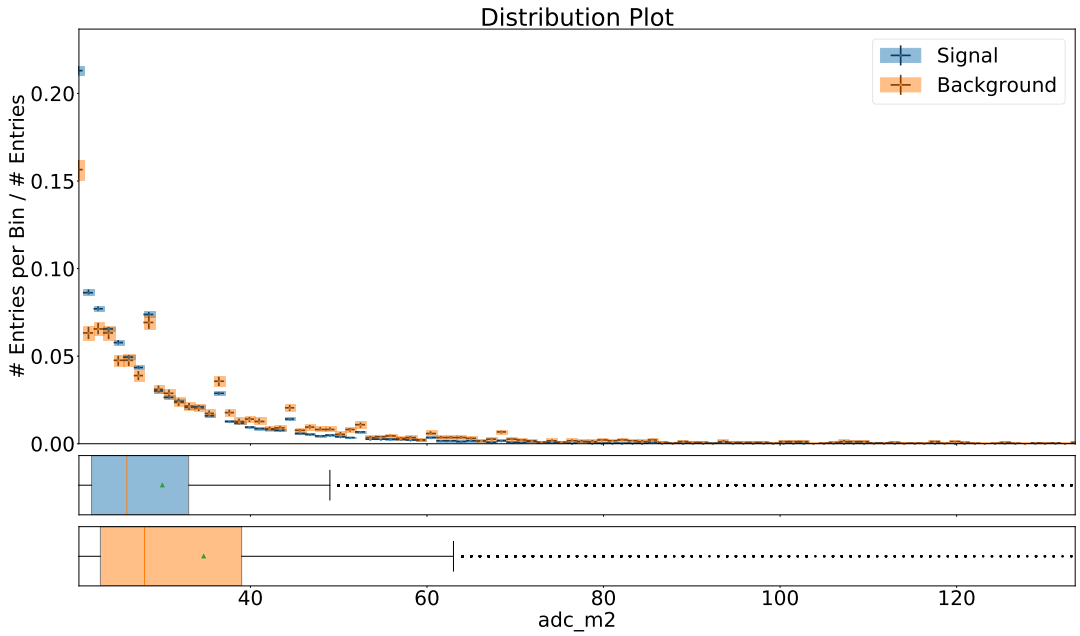
2.4 adc_variance



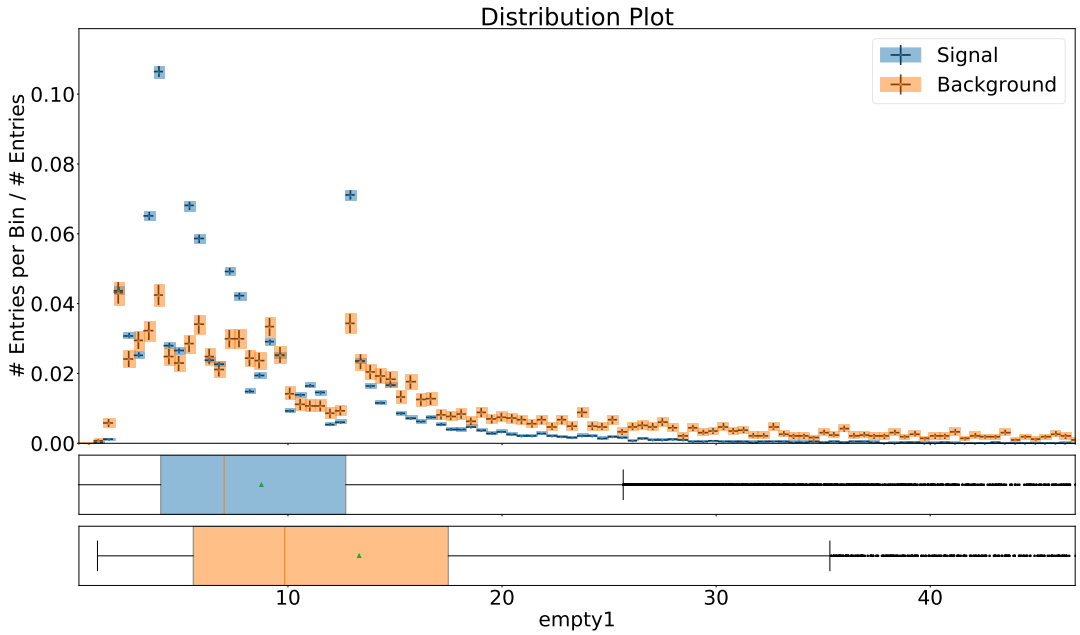
2.5 adc_max



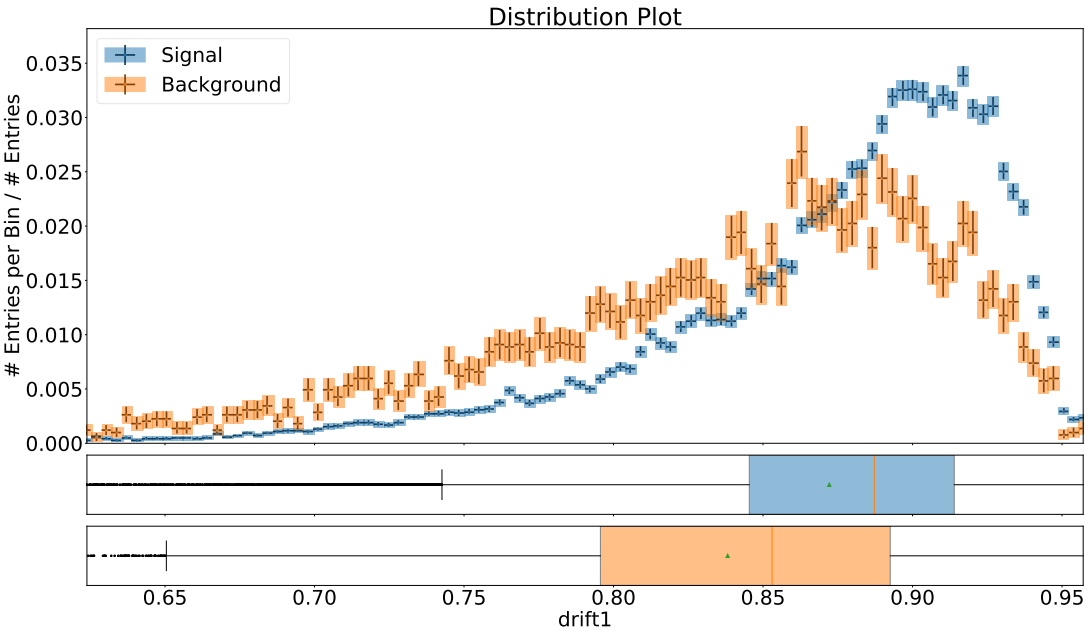
2.6 adc_min



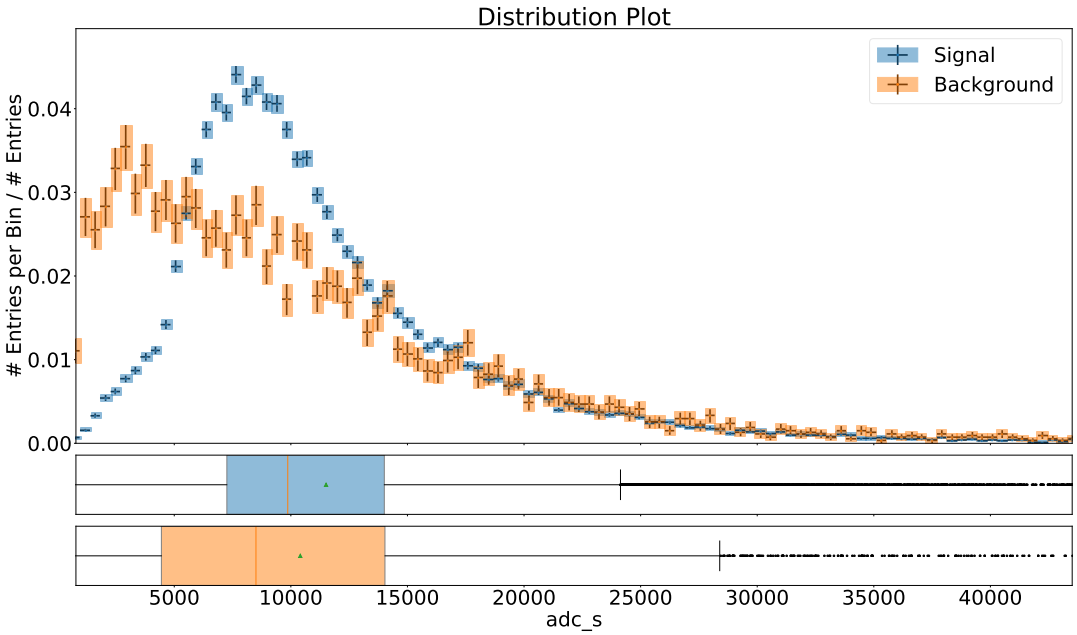
2.7 empty_s_max



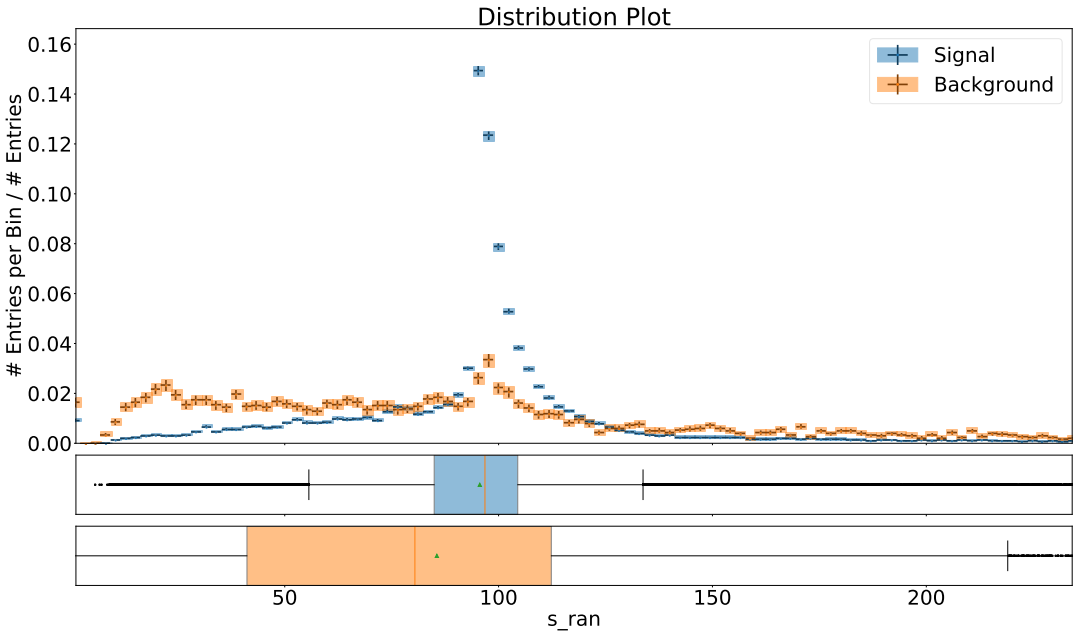
2.8 drift_length_max



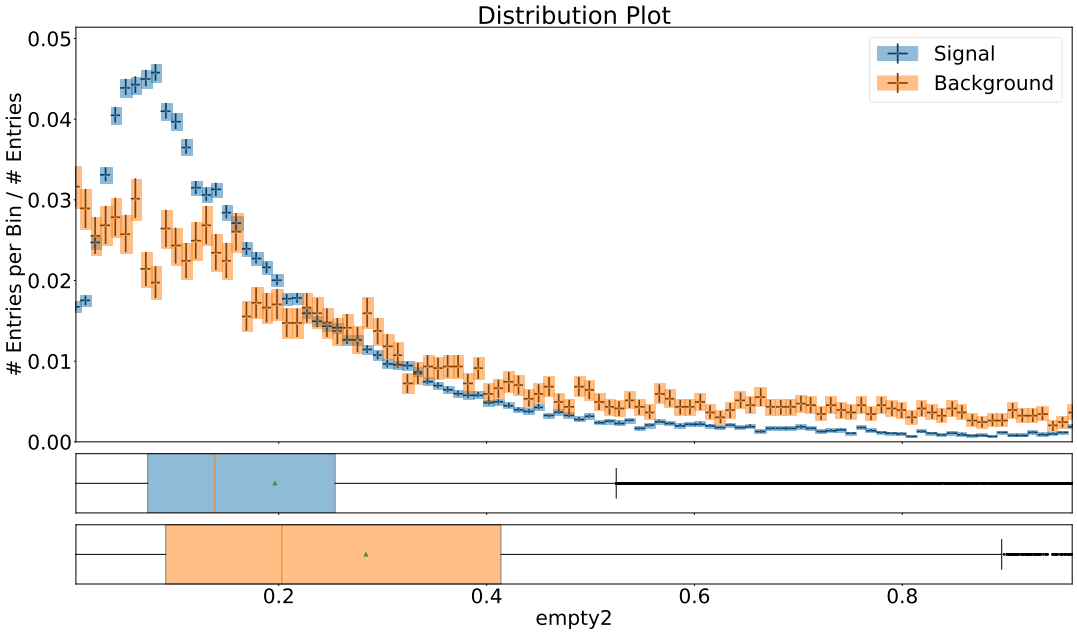
2.9 adc_sum



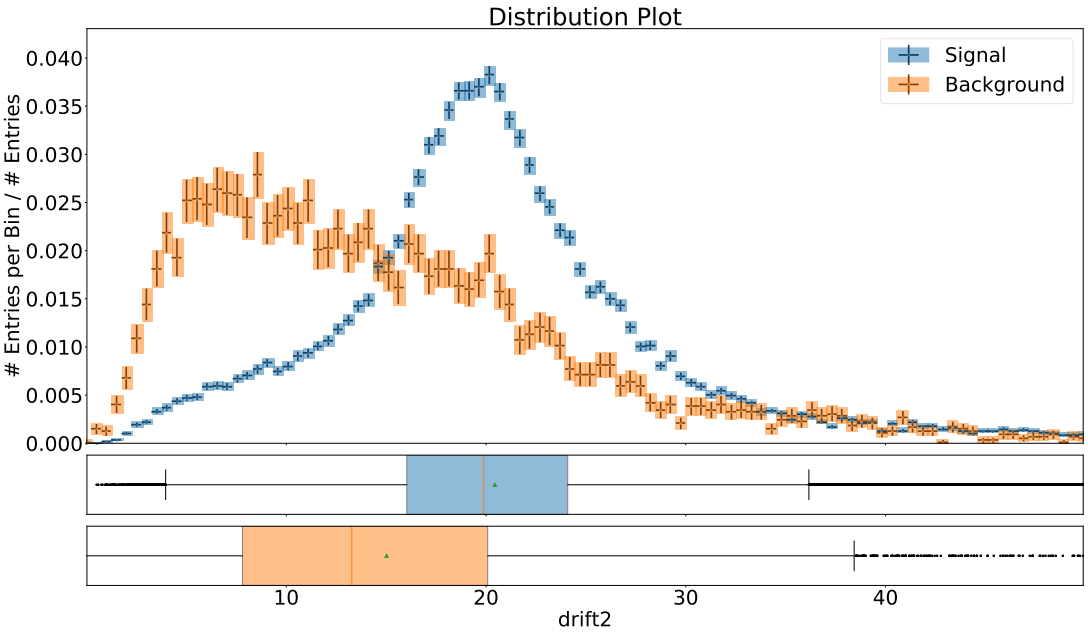
2.10 s_range



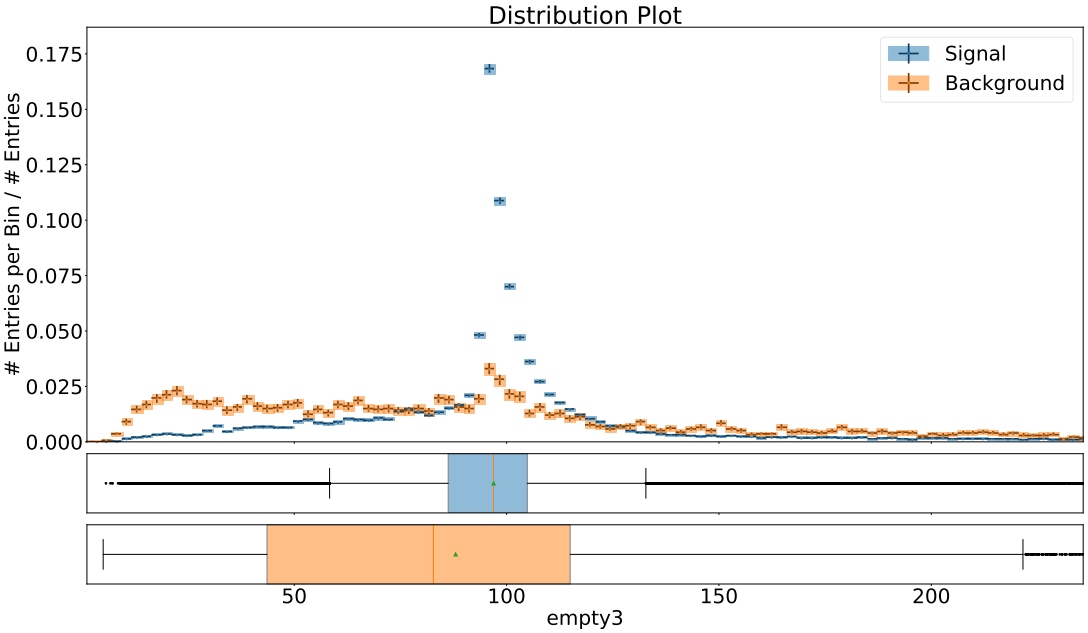
2.11 empty_s_min



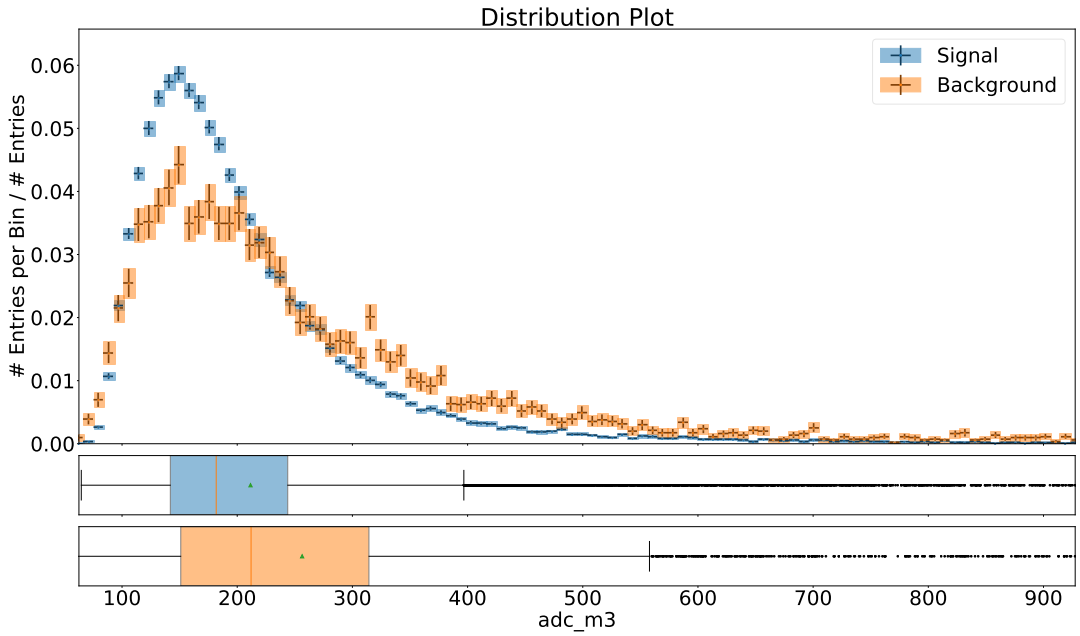
2.12 drift__length__sum



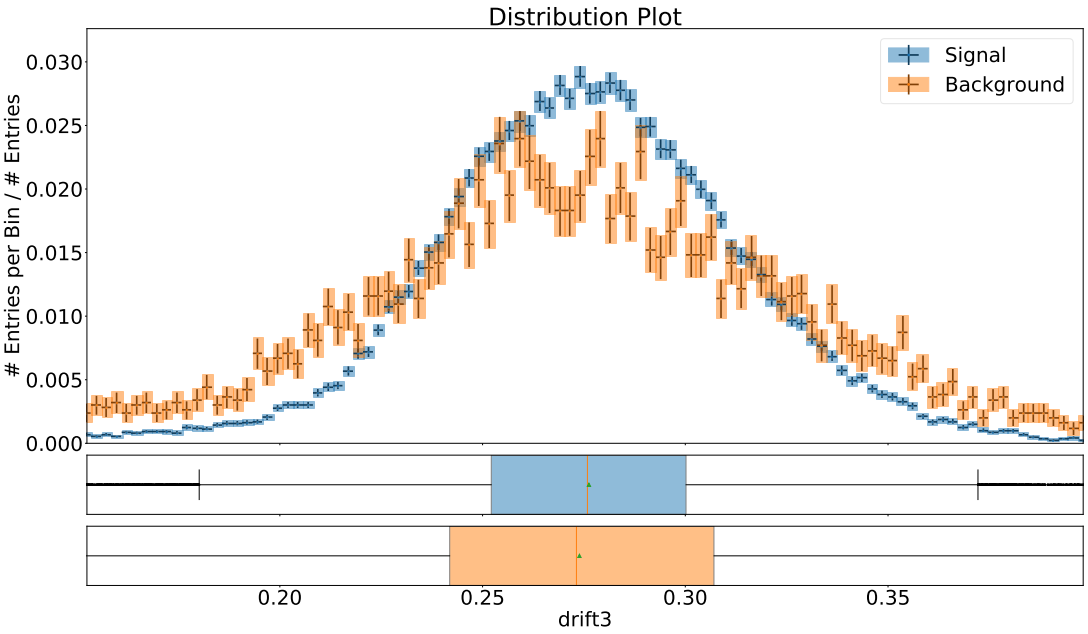
2.13 empty__s__sum



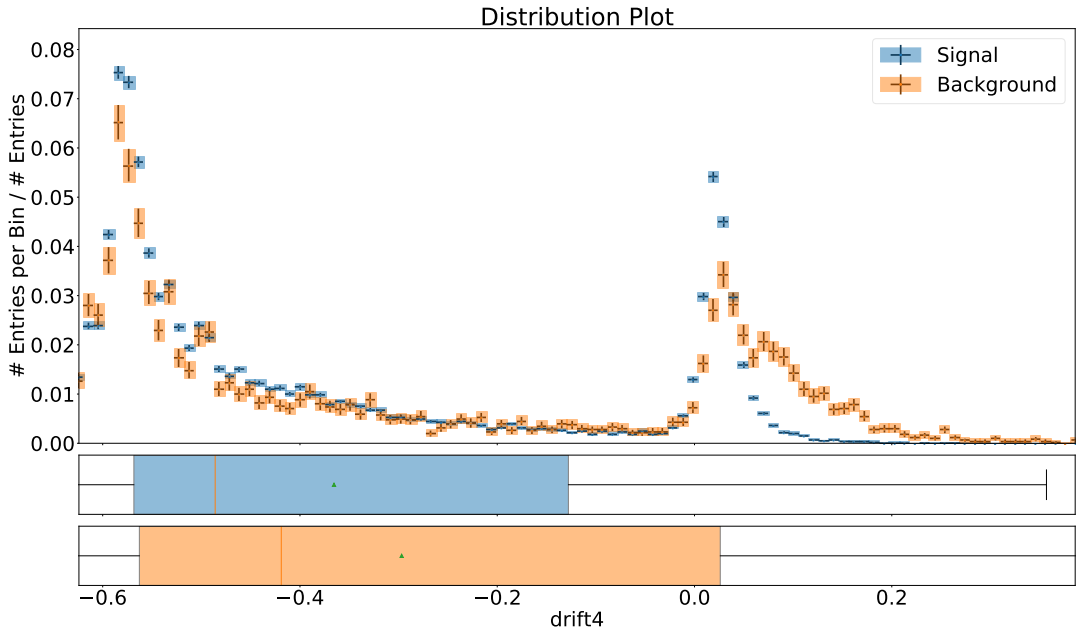
2.14 adc_mean



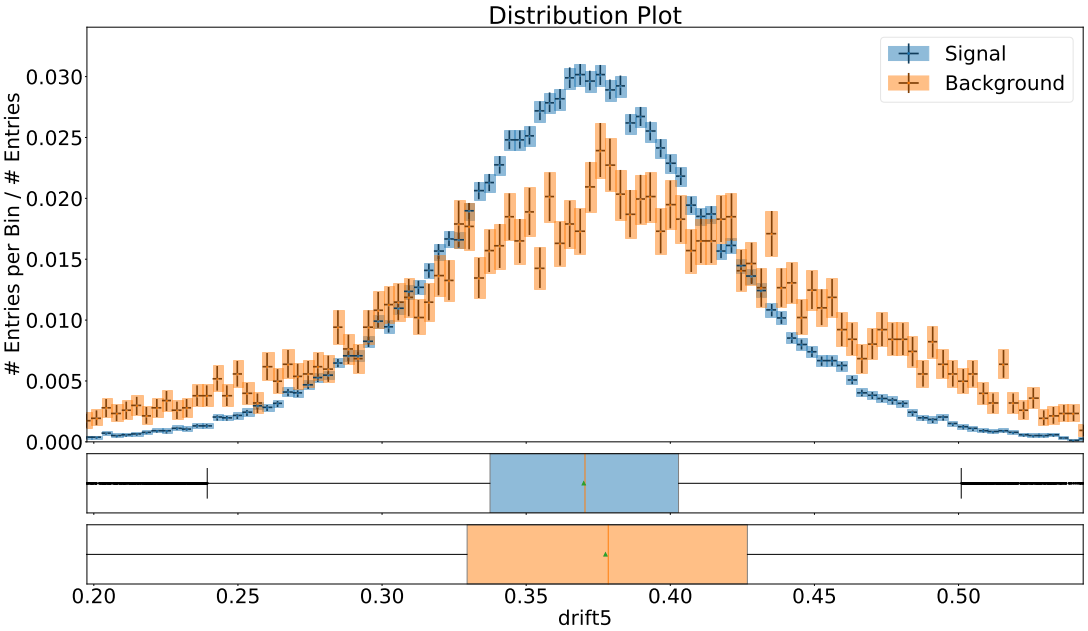
2.15 drift_length_variance



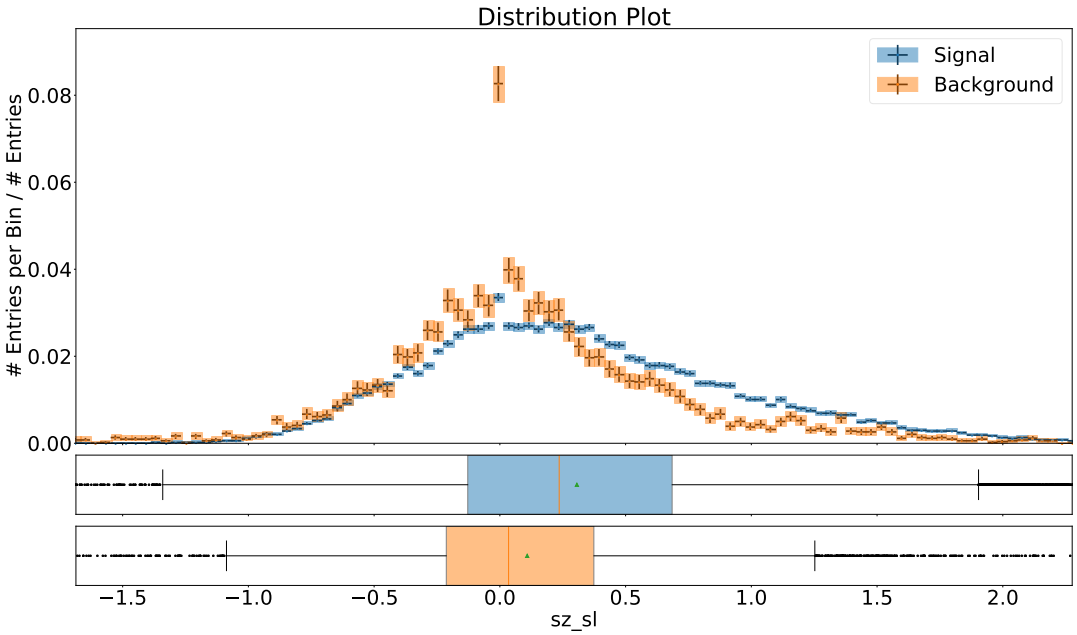
2.16 drift__length__min



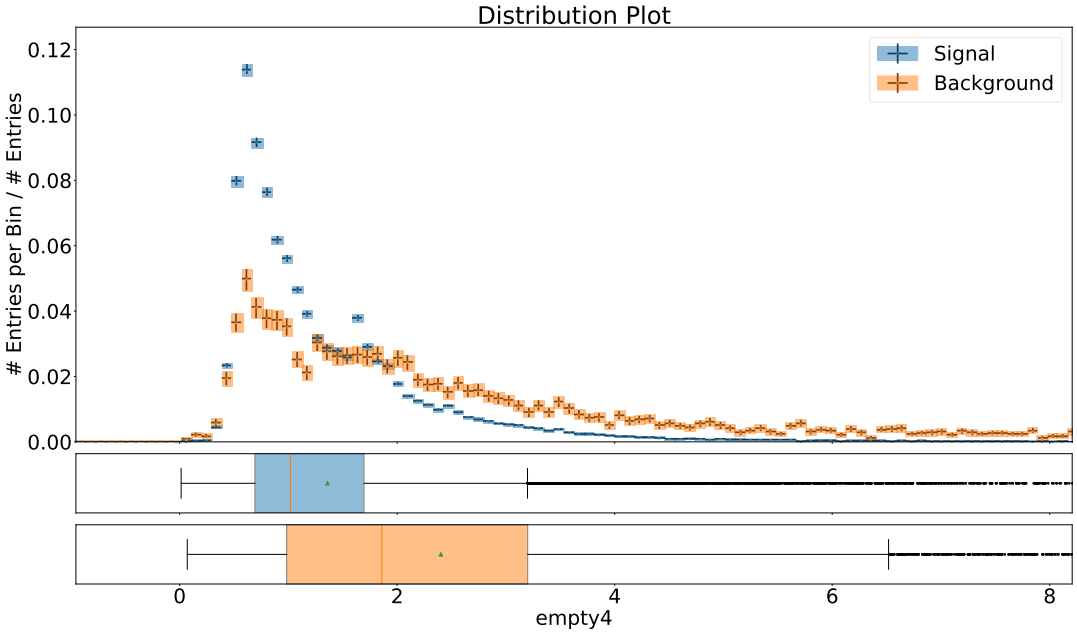
2.17 drift__length__mean



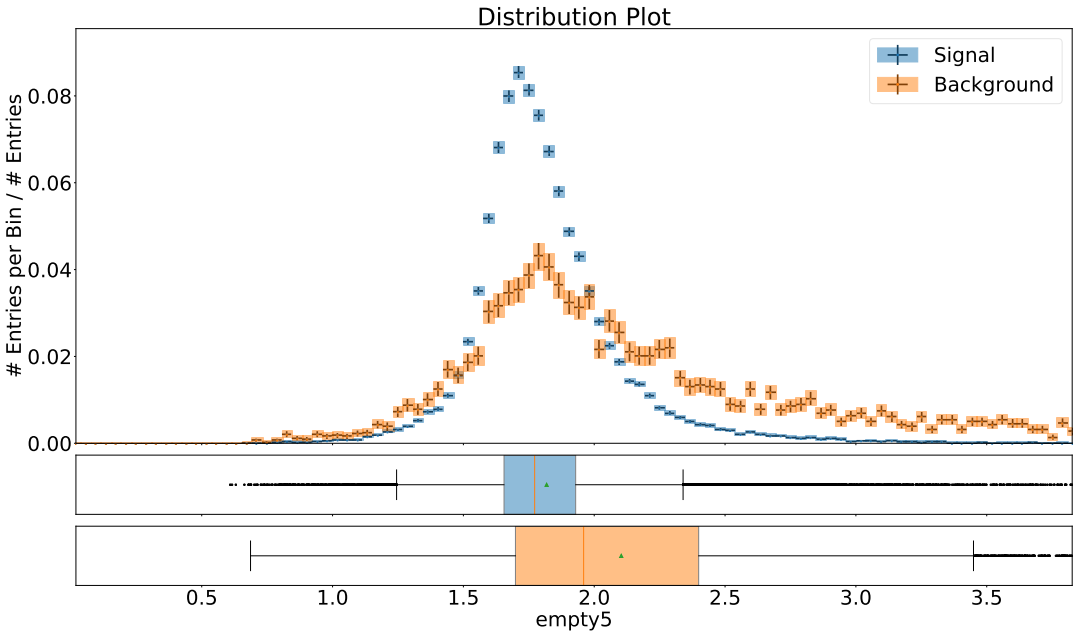
2.18 sz_slope



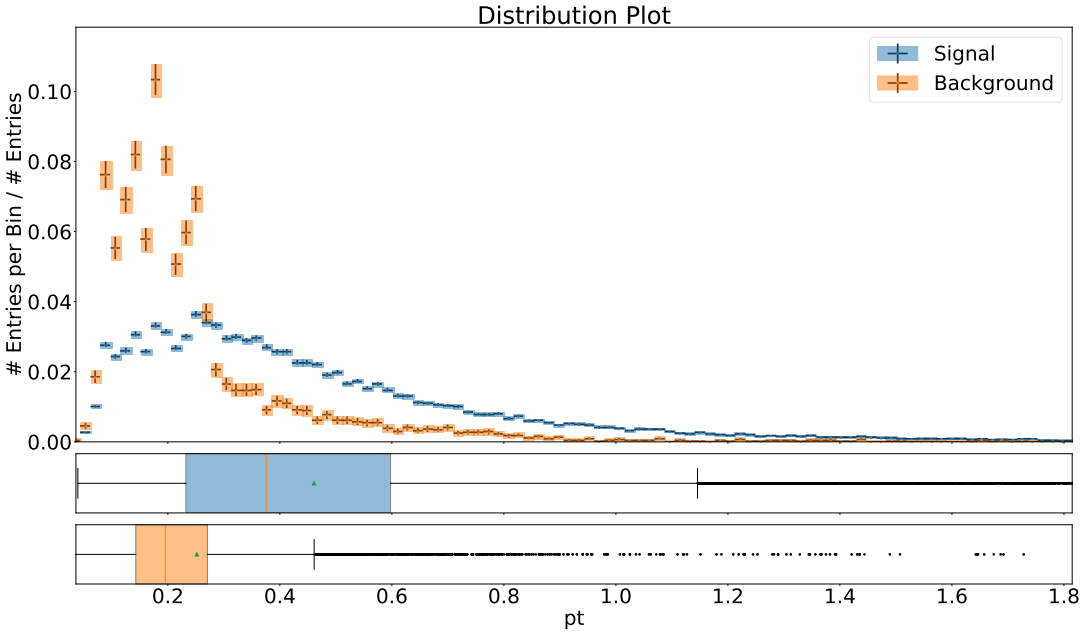
2.19 empty_s_variance



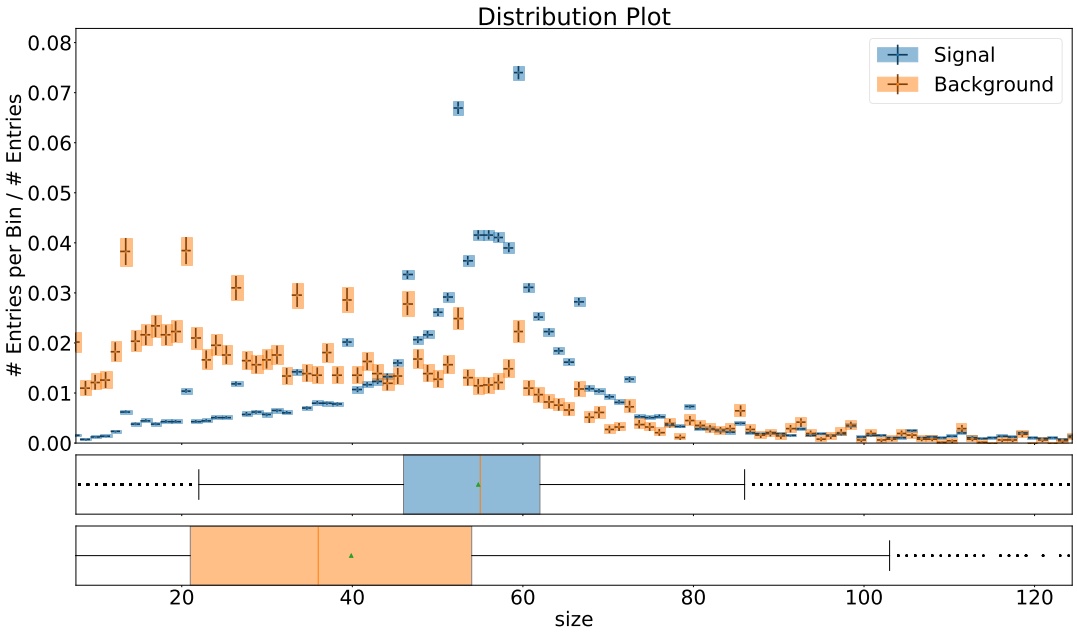
2.20 empty__s__mean



2.21 pt



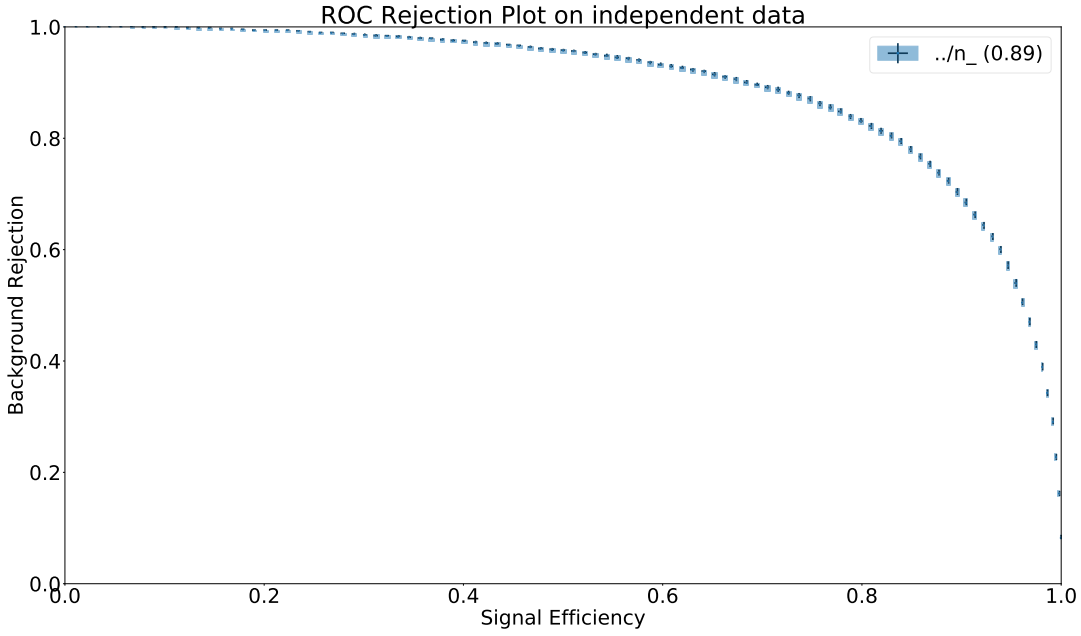
2.22 size



3 Classifier Plot

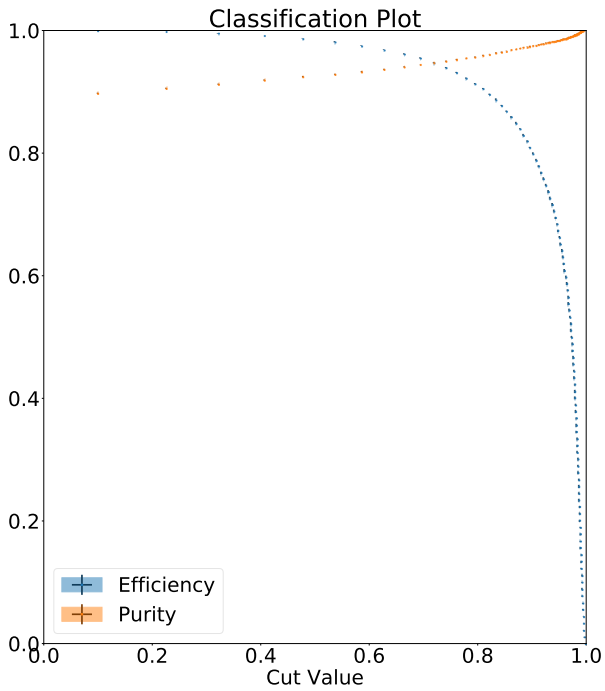
This section contains the receiver operating characteristics (ROC), purity projection, ...of the classifiers on training and independent data.The legend of each plot contains the shortened identifier and the area under the ROC curvein parenthesis.

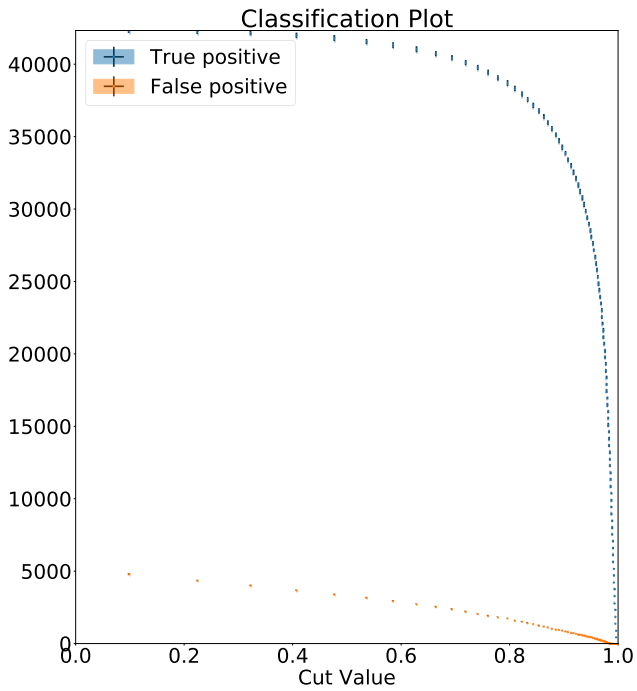
4 ROC Plot



5 Classification Results

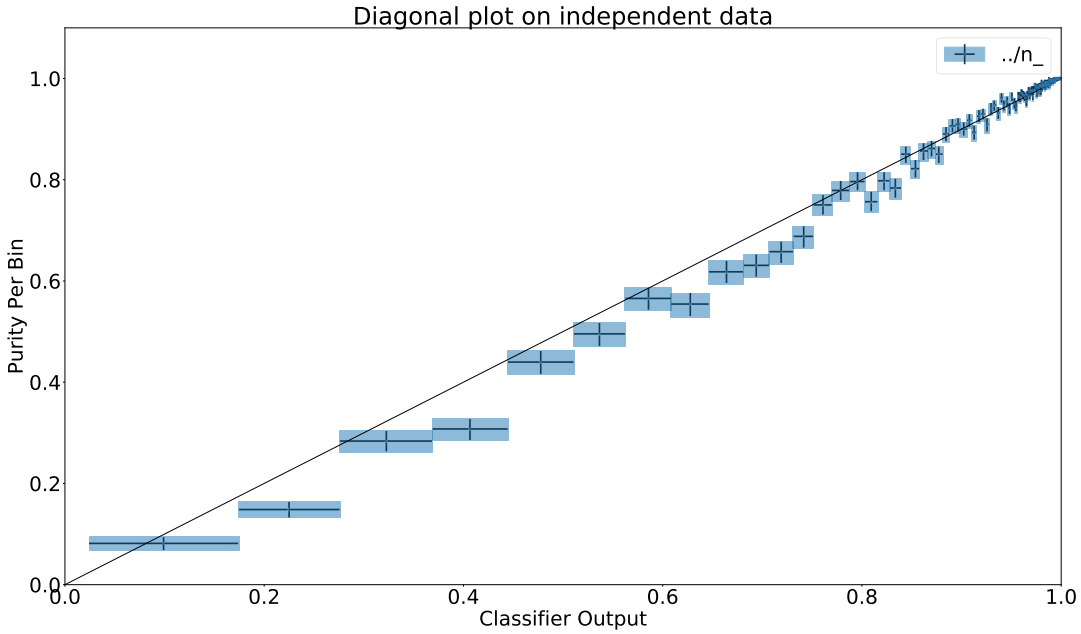
5.1 ../n_





6 Diagonal Plot

6.1 ../n_



7 Spectators

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

Table 3: Abbreviations of spectators

Spectator	Abbreviation