Automatic MVA Evaluation

Thomas Keck
Moritz Gelb
Nils Braun

March 26, 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
../n_events_training=5000/trackfindingvxd_TrackQualityIndicator.weights.xml	../n_

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


```
<?xml version="1.0" encoding="utf-8"?>
<method>FastBDT</method>
<weightfile>./git_hash=cd1cfc6b5beb639a4b2839a73e8b8f24bea35572/n_events_training=5000/trackfindingvxd_TrackQualityIndicator.
     weights.xml</weightfile>
<treename>tree</treename>
<target_variable>truth</target_variable>
<weight_variable/>
<signal_class>1</signal_class>
<max_events>0</max_events>
<number_feature_variables>27</number_feature_variables>
<variable0>tripletFit_QI</variable0>
<variable1>tripletFit_Chi2</variable1>
<variable2>tripletFit_Pt</variable2>
<variable3>tripletFit_PMag</variable3>
<variable4>tripletFit_P_Mag</variable4>
<variable5>tripletFit_P_Eta</variable5>
<variable6>tripletFit_P_Phi</variable6>
<variable7>tripletFit_P_X</variable7>
<variable8>tripletFit_P_Y</variable8>
<variable9>tripletFit_P_Z</variable9>
<variable10>NSpacePoints</variable10>
<variable11>charge_max</variable11>
<variable12>charge_min</variable12>
<variable13>charge_mean</variable13>
<variable14>charge_std</variable14>
<variable15>seedCharge_max</variable15>
<variable16>seedCharge_min</variable16>
<variable17>seedCharge_mean</variable17>
<variable18>seedCharge_std</variable18>
<variable19>size_max</variable19>
<variable20>size_min</variable20>
<variable21>size_mean</variable21>
<variable22>size_std</variable22>
<variable23>energyLoss_max</variable23>
<variable24>energyLoss_min</variable24>
<variable25>energyLoss_mean</variable25>
<variable26>energyLoss_std</variable26>
<number_spectator_variables>0</number_spectator_variables>
<number_data_files>1</number_data_files>
<datafile0>./git_hash=cd1cfc6b5beb639a4b2839a73e8b8f24bea35572/n_events_training=5000/random_seed=trainvxd_0/vxd_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

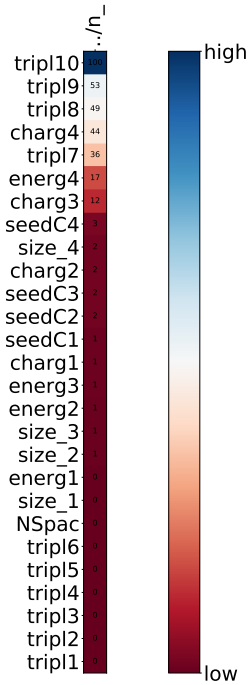
This section contains an overview of the importance and correlation of the variables used by the classifiers. And distribution plots of the variables on the independent dataset. The distributions are normed for signal and background separately, and only the region +- 3 sigma around the mean is shown.

Table 2: Abbreviations of variables

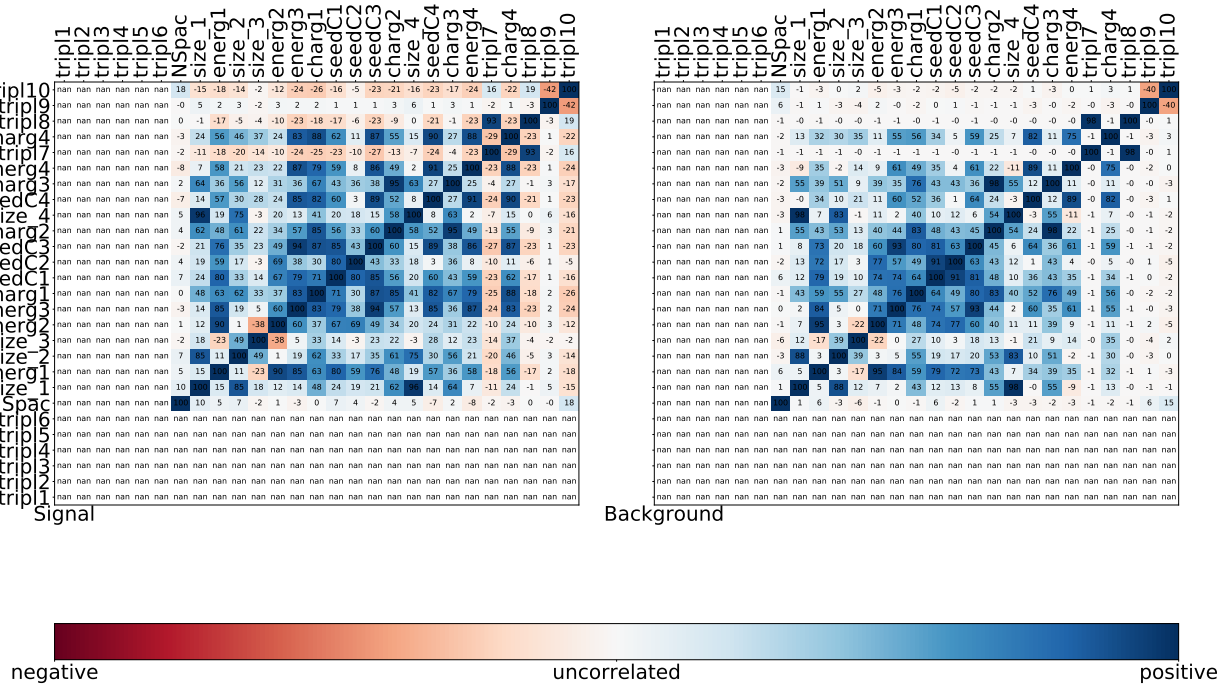
Variable	Abbreviation

tripletFit_P_Mag	tripl1
tripletFit_P_Eta	tripl2
tripletFit_P_Phi	tripl3
tripletFit_P_X	tripl4
tripletFit_P_Y	tripl5
tripletFit_P_Z	tripl6
NSpacePoints	NSpac
size_max	size_1
energyLoss_max	energ1
size_mean	size_2
size_min	size_3
energyLoss_std	energ2
energyLoss_mean	energ3
charge_mean	charg1
seedCharge_max	seedC1
seedCharge_std	seedC2
seedCharge_mean	seedC3
charge_max	charg2
size_std	size_4
seedCharge_min	seedC4
charge_std	charg3
energyLoss_min	energ4
tripletFit_Pt	tripl7
charge_min	charg4
tripletFit_PMag	tripl8
tripletFit_Chi2	tripl9
tripletFit_QI	tripl10

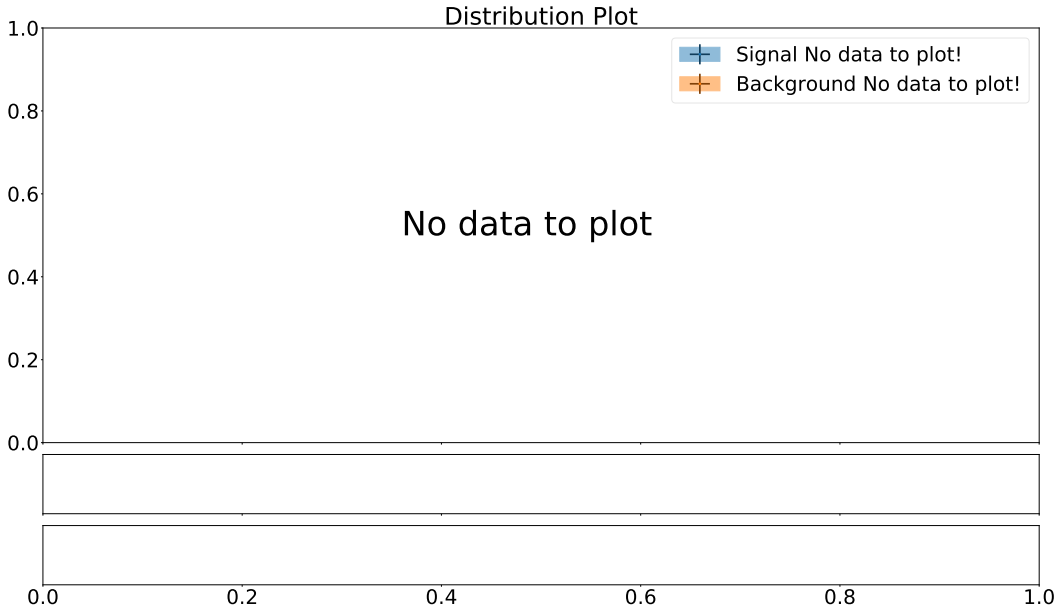
2.1 Importance



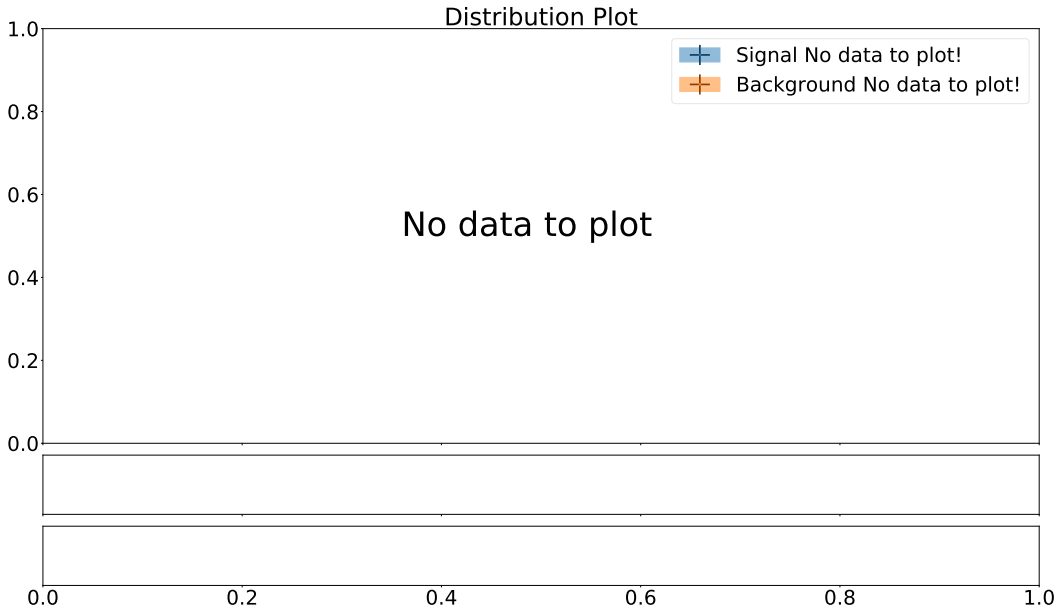
2.2 Correlation



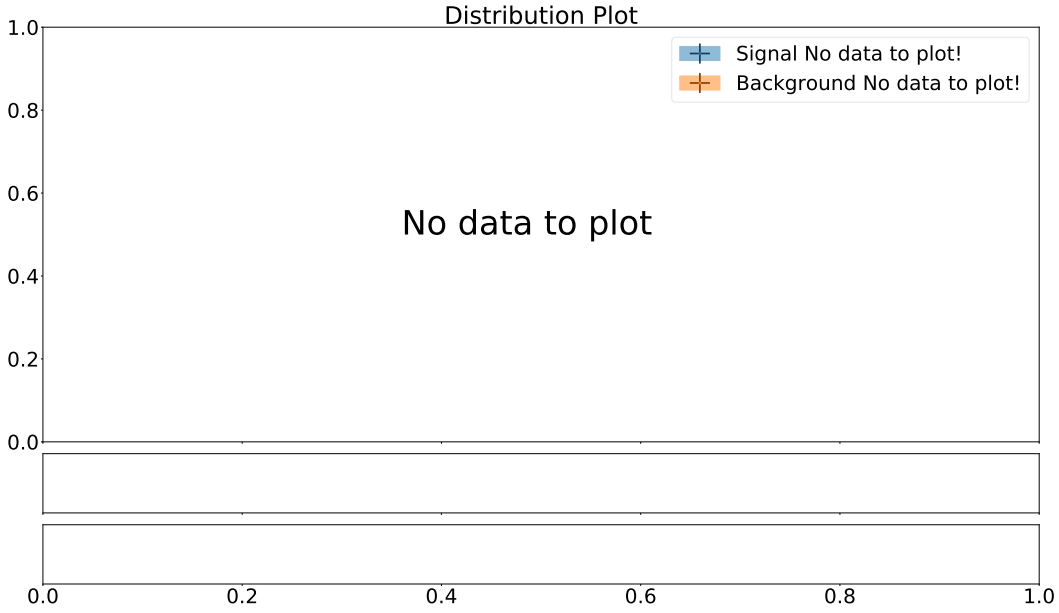
2.3 tripletFit_P_Mag



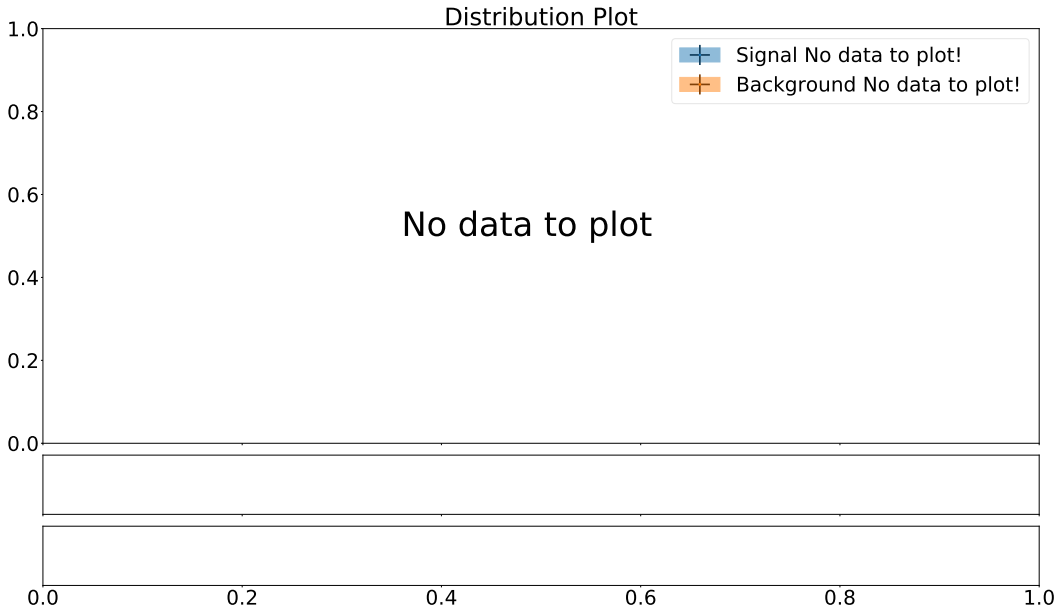
2.4 tripletFit_P_Eta



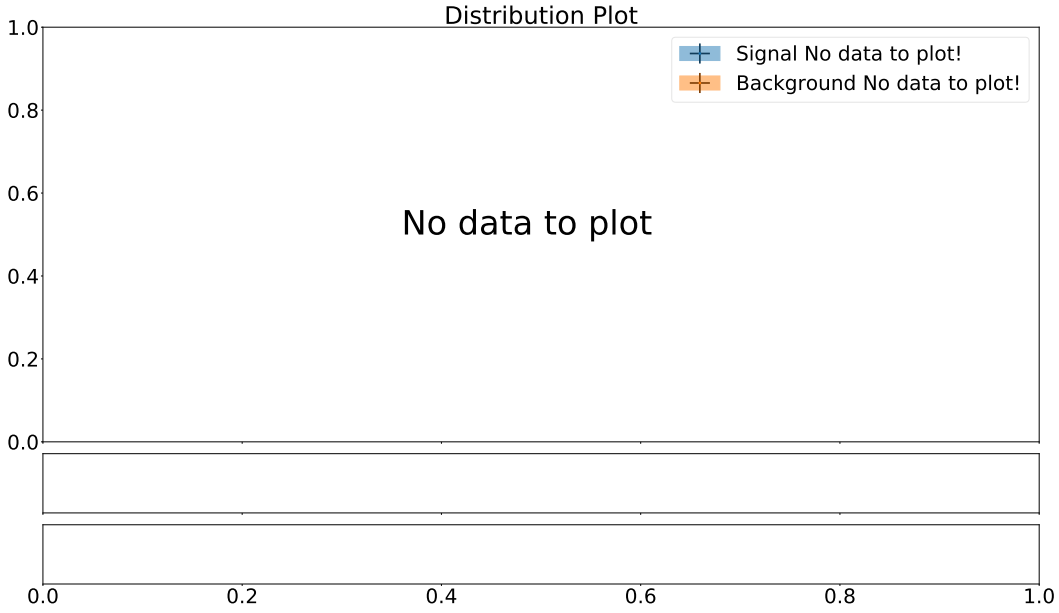
2.5 tripletFit_P_Phi



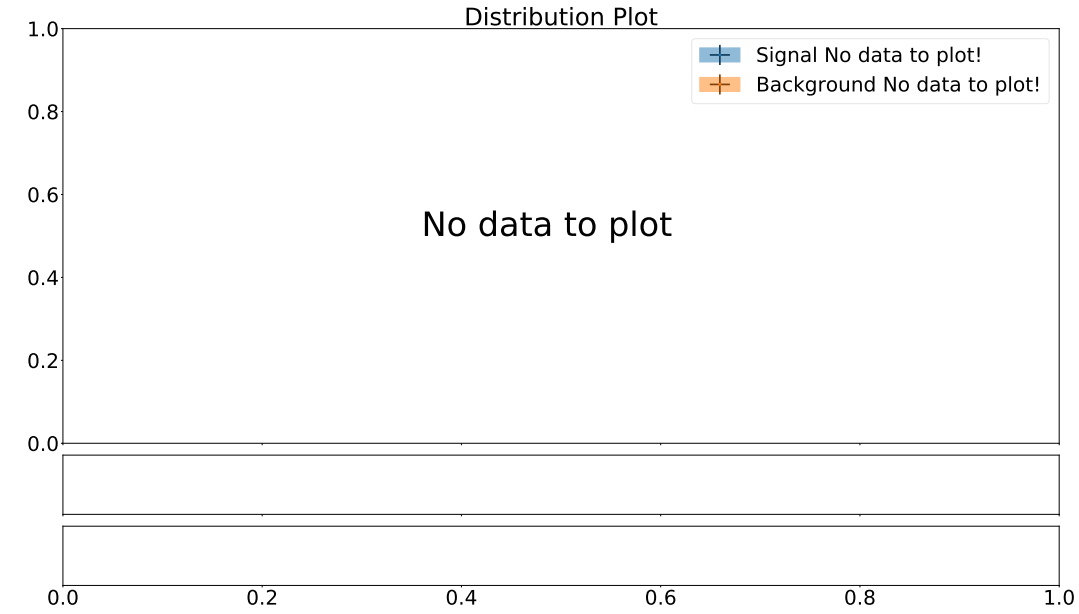
2.6 tripletFit_P_X



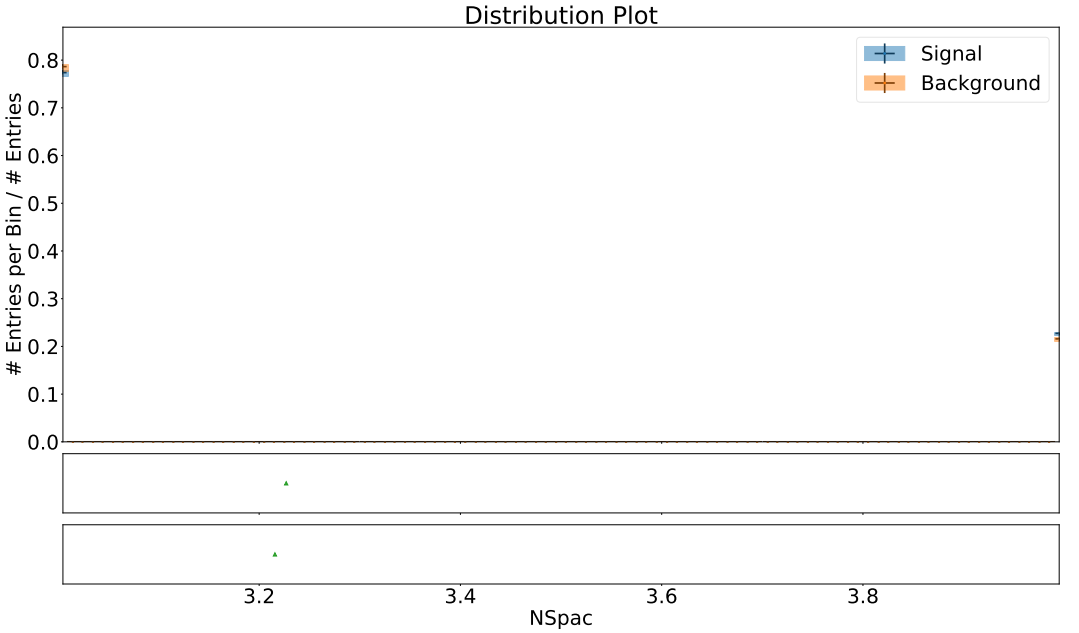
2.7 tripletFit_P_Y



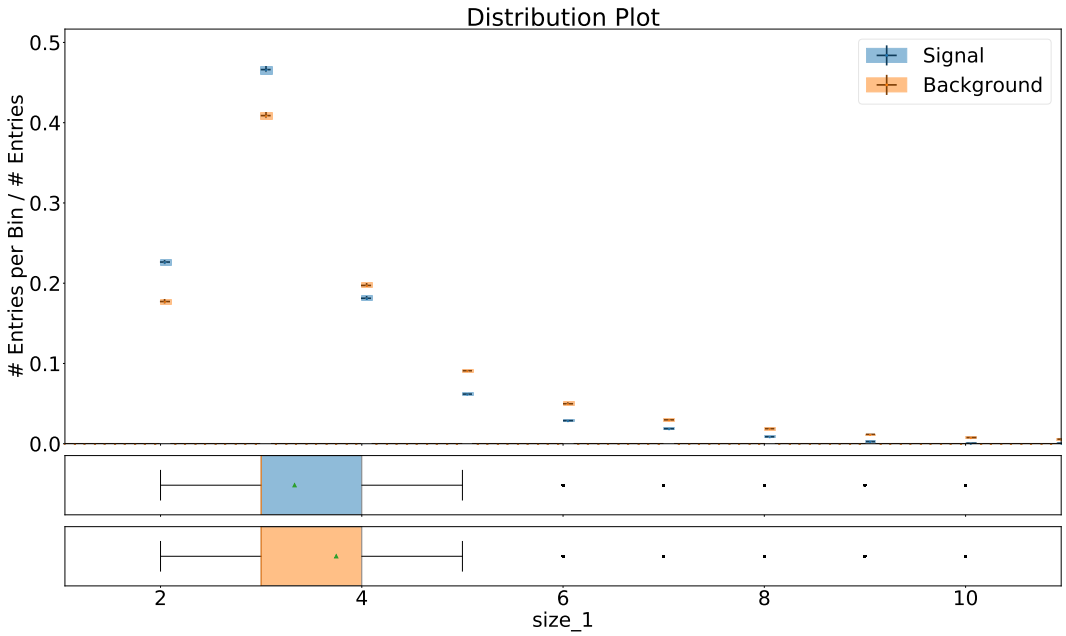
2.8 tripletFit_P_Z



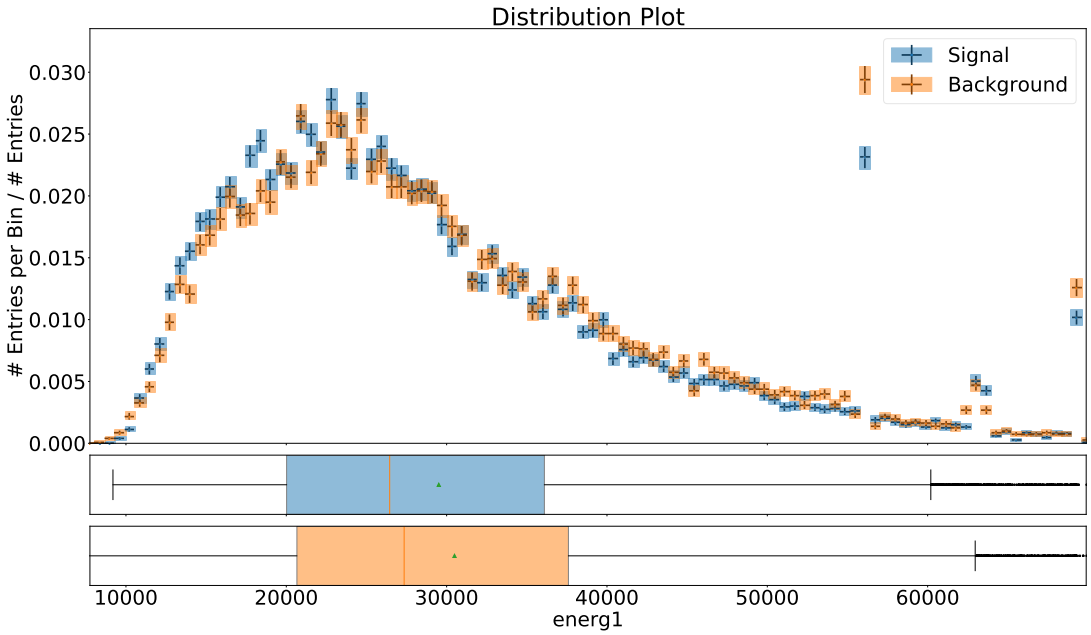
2.9 NSpacePoints



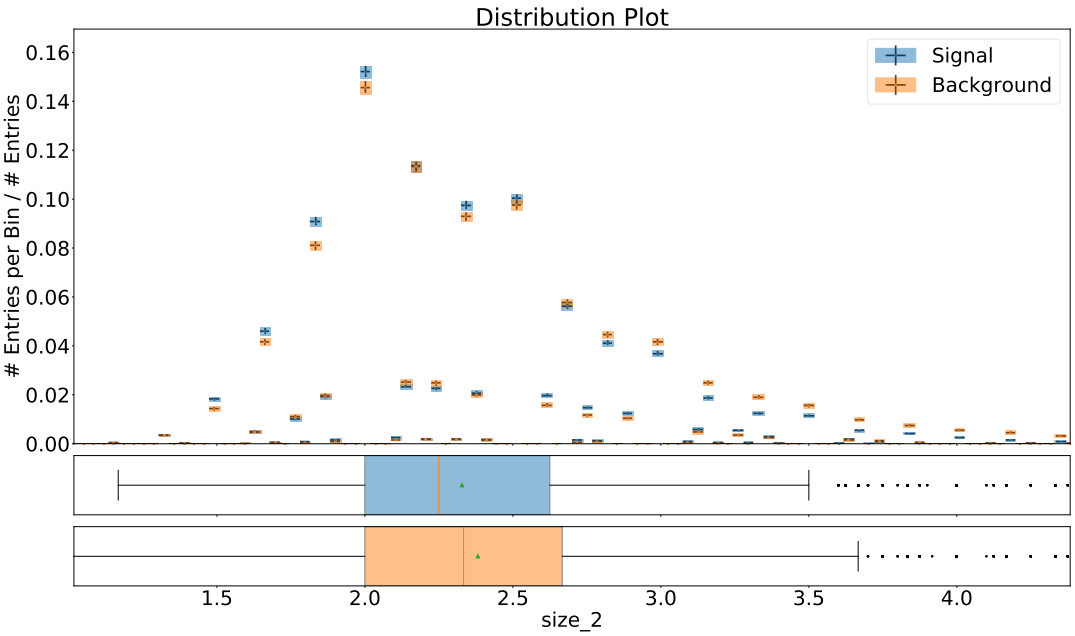
2.10 size__max



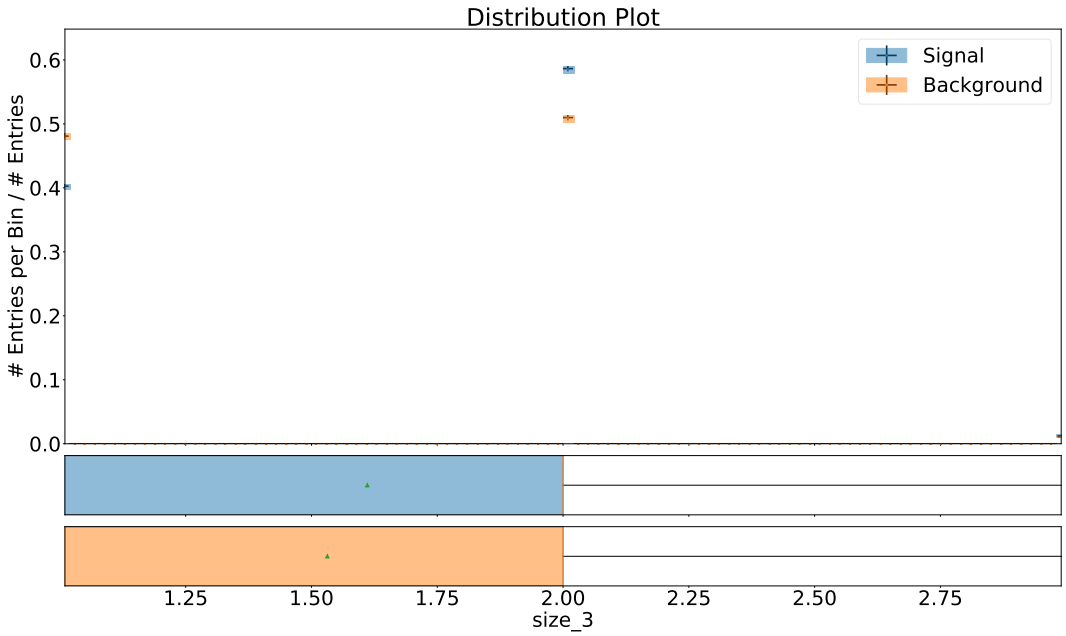
2.11 energyLoss__max



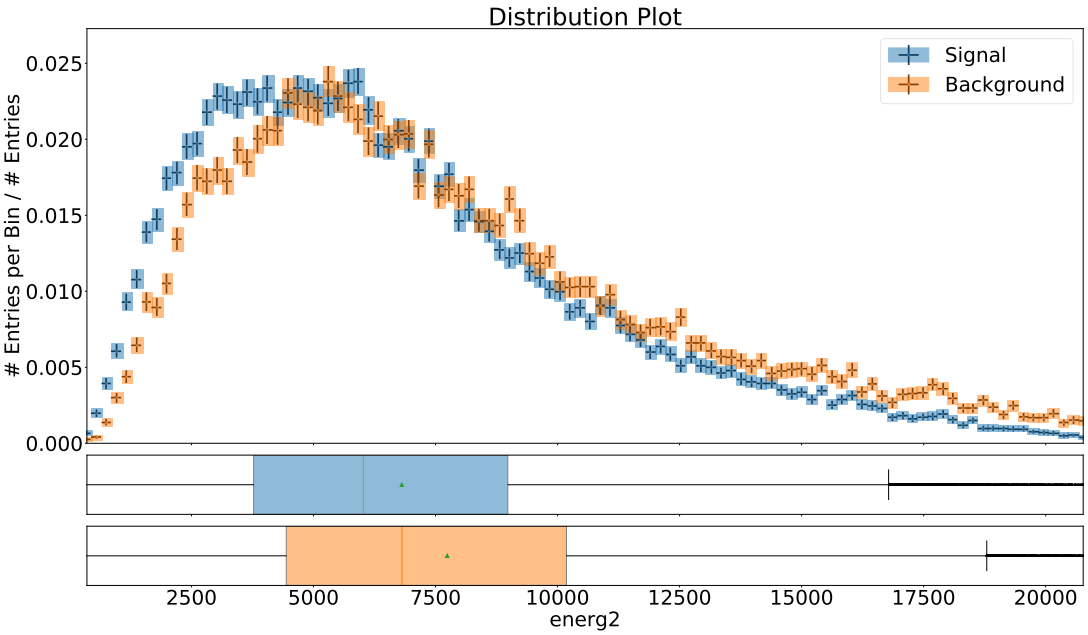
2.12 size__mean



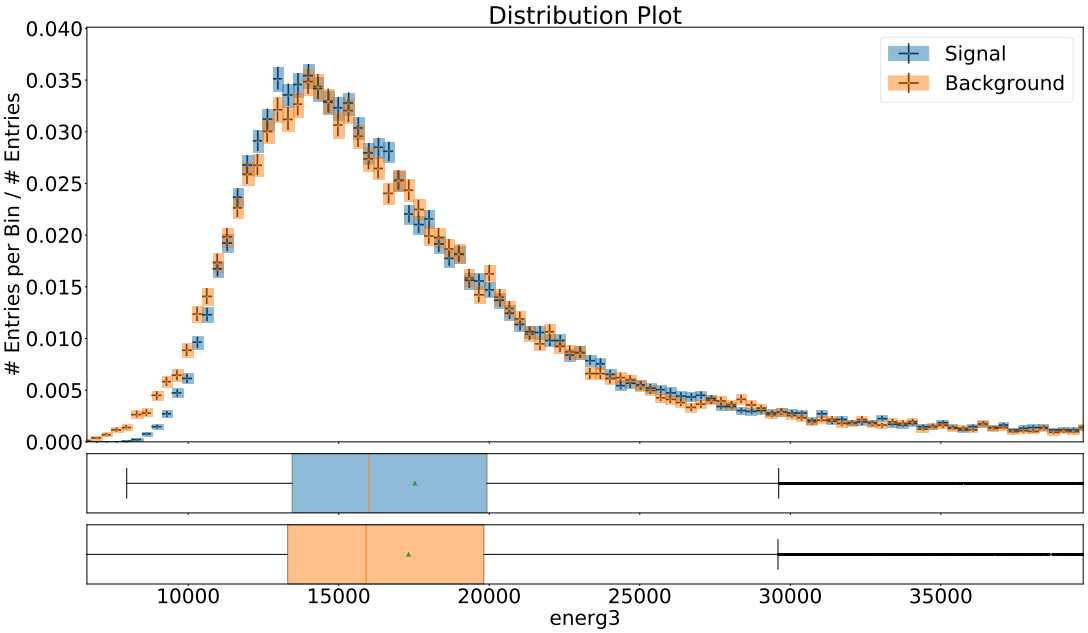
2.13 size__min



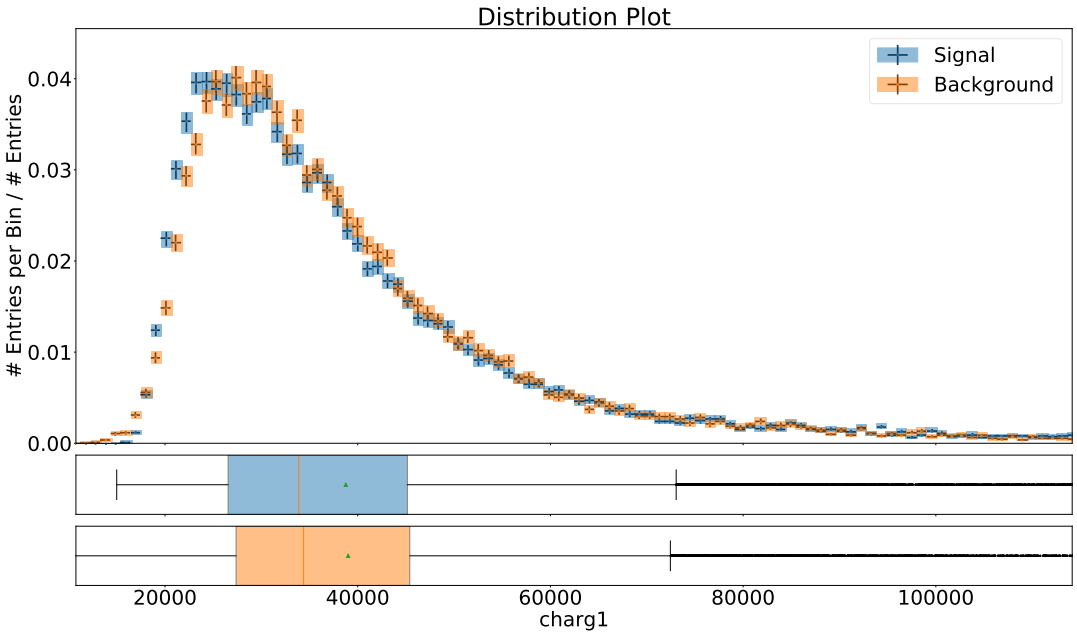
2.14 energyLoss_std



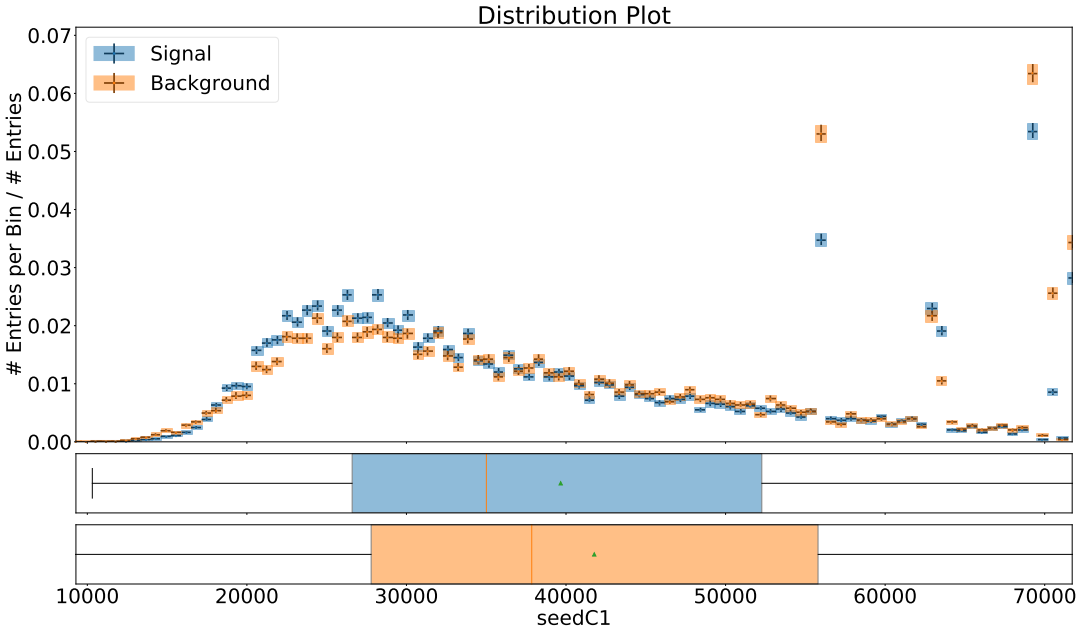
2.15 energyLoss_mean



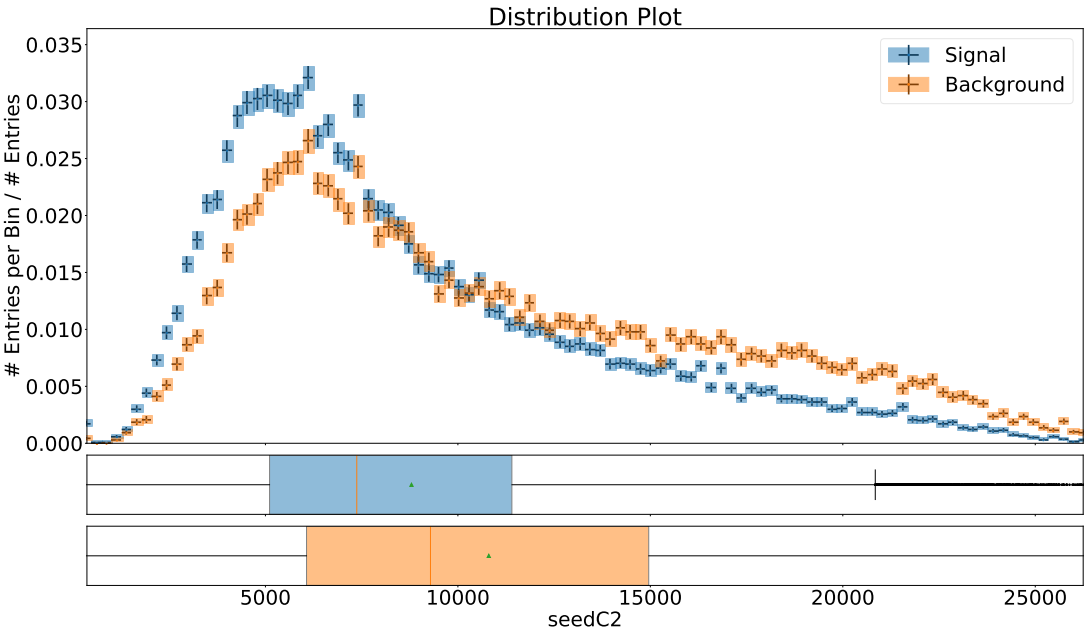
2.16 charge_mean



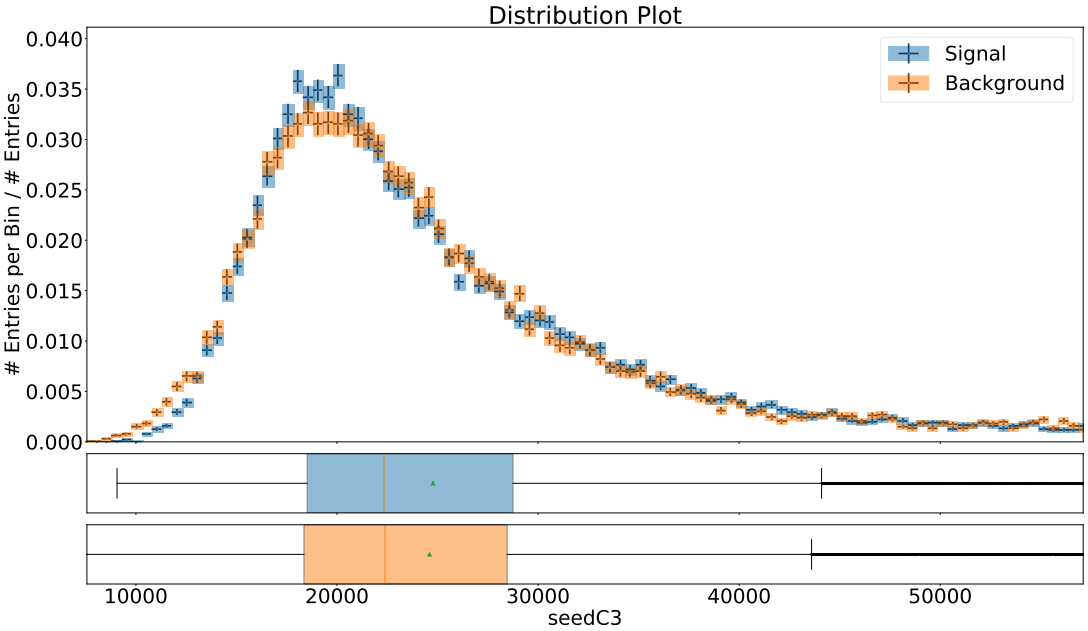
2.17 seedCharge_max



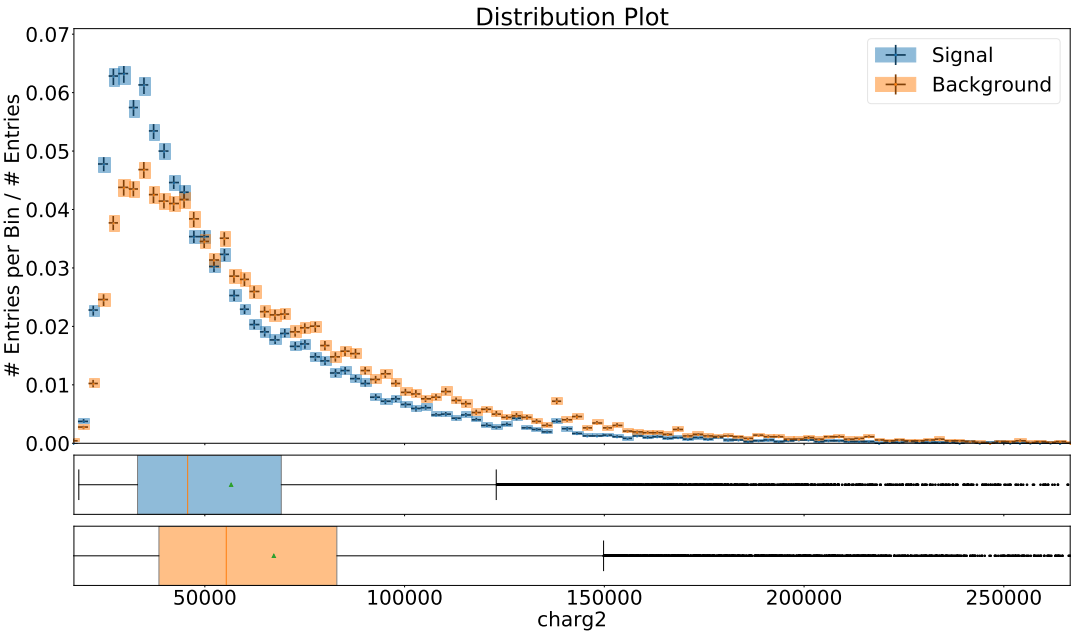
2.18 seedCharge_std



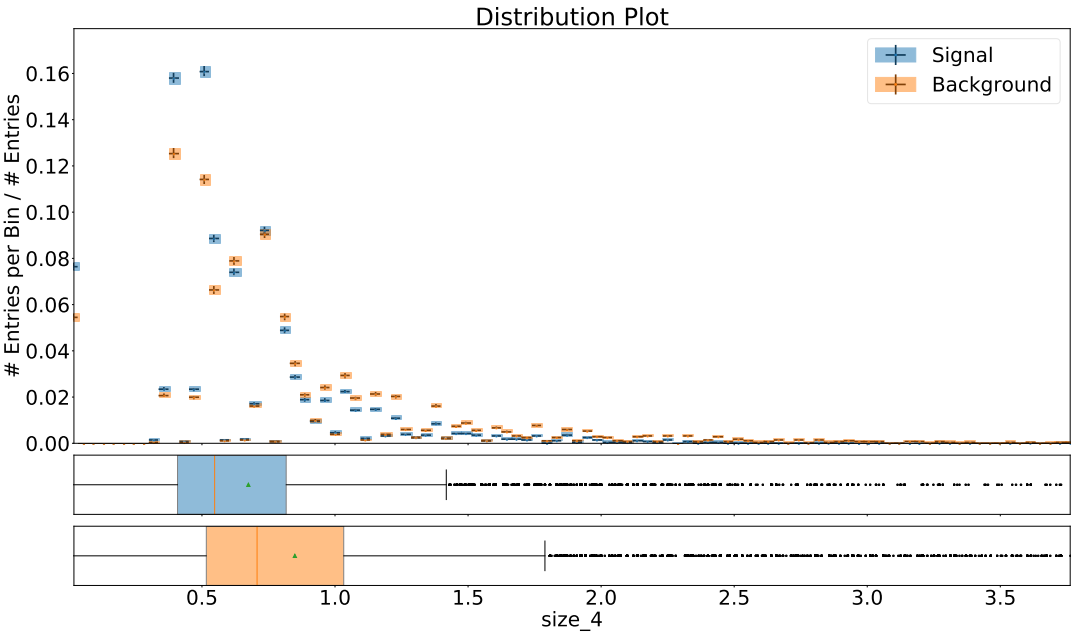
2.19 seedCharge_mean



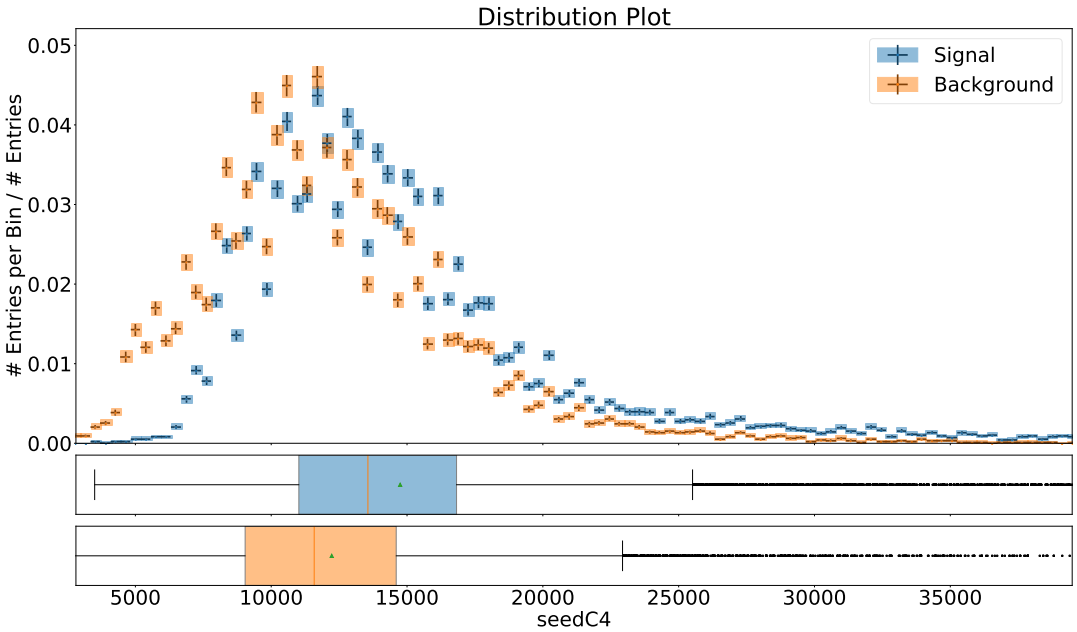
2.20 charge_max



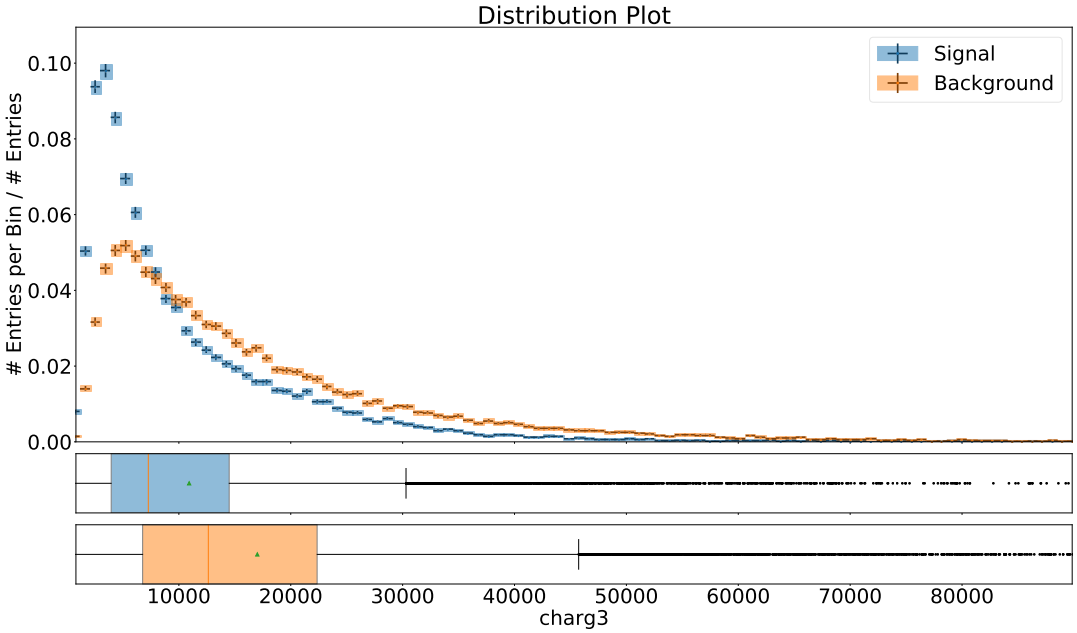
2.21 size_std



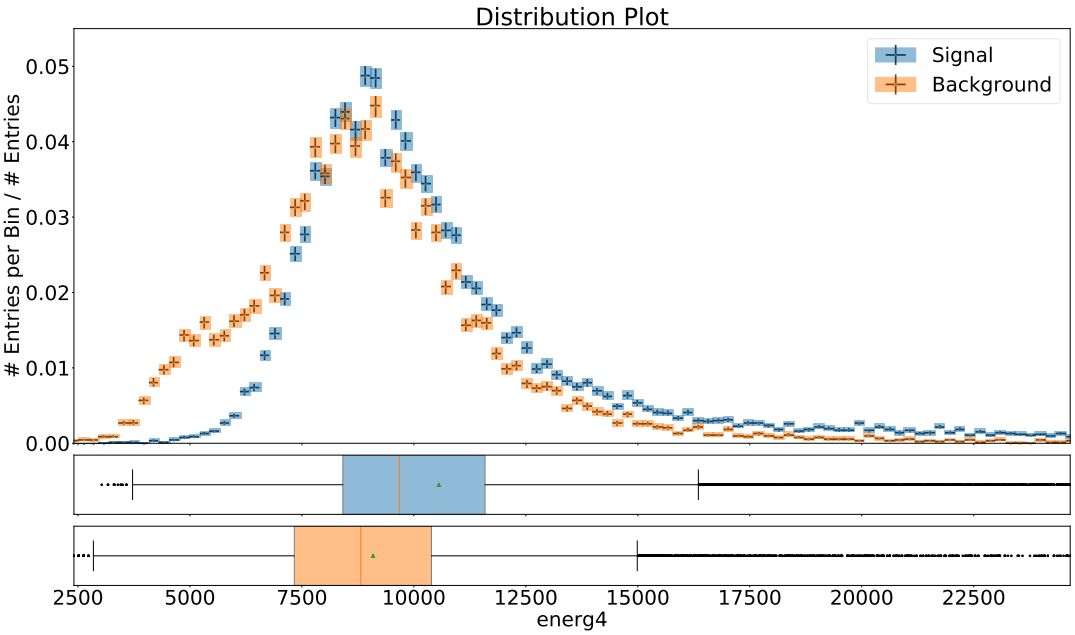
2.22 seedCharge_min



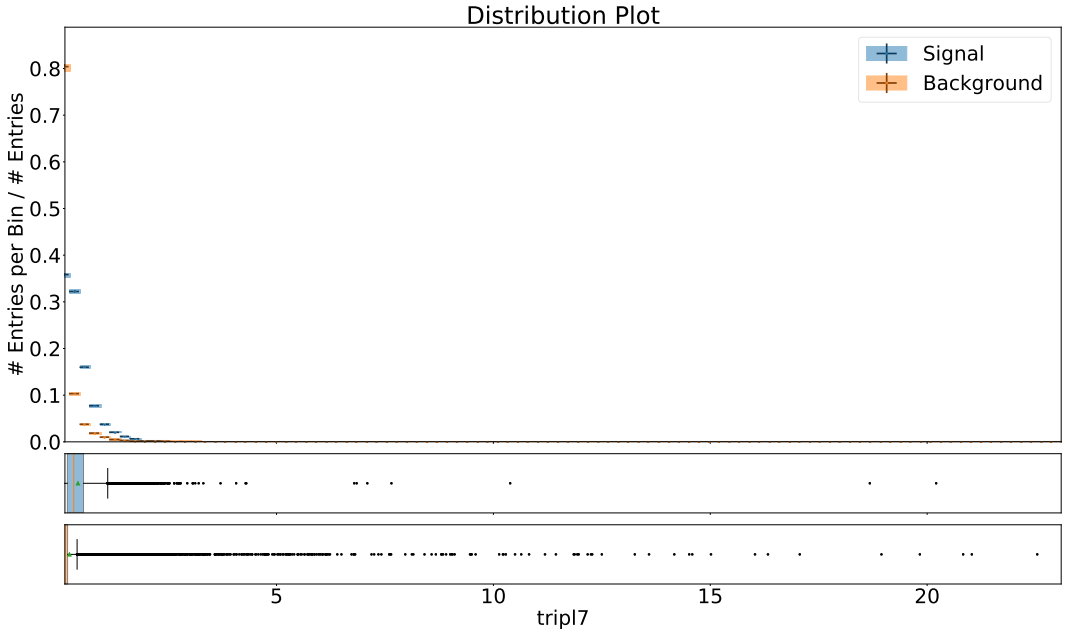
2.23 charge_std



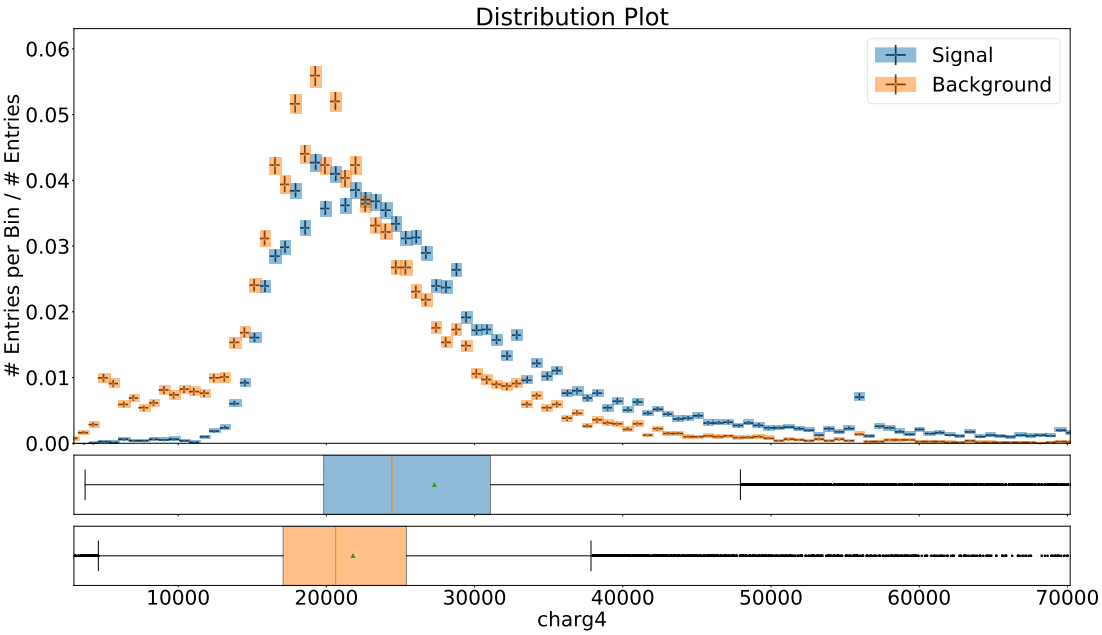
2.24 energyLoss_min



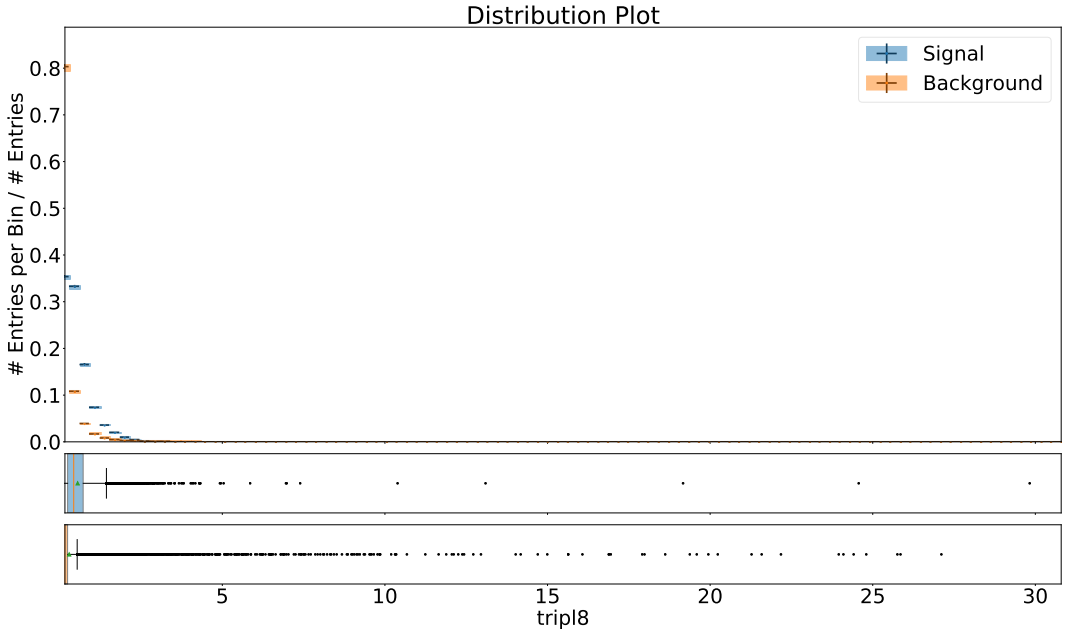
2.25 tripletFit_Pt



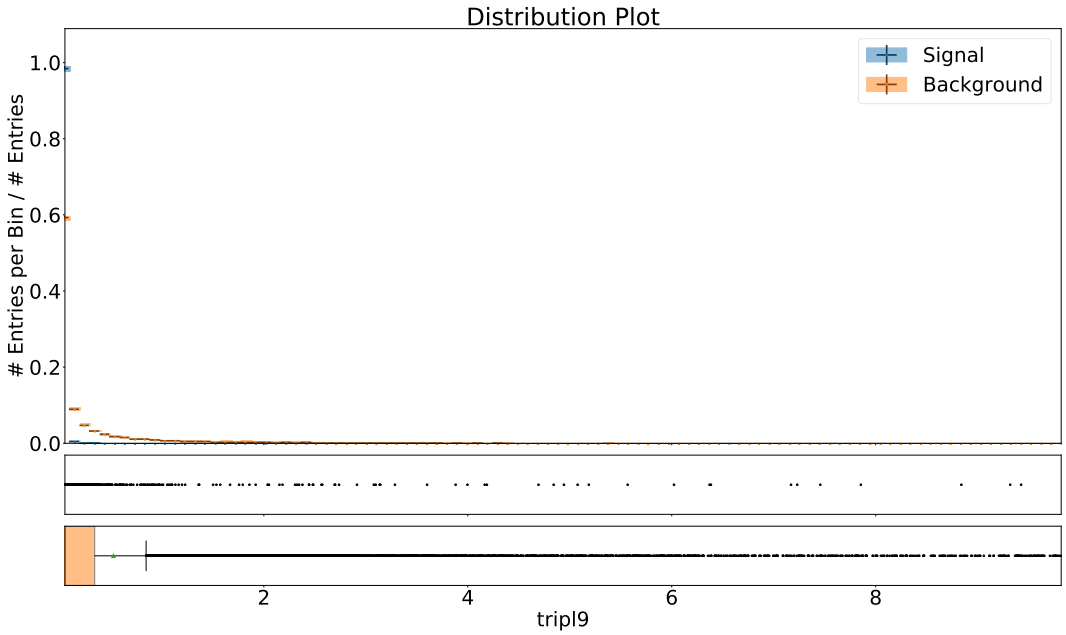
2.26 charge_min



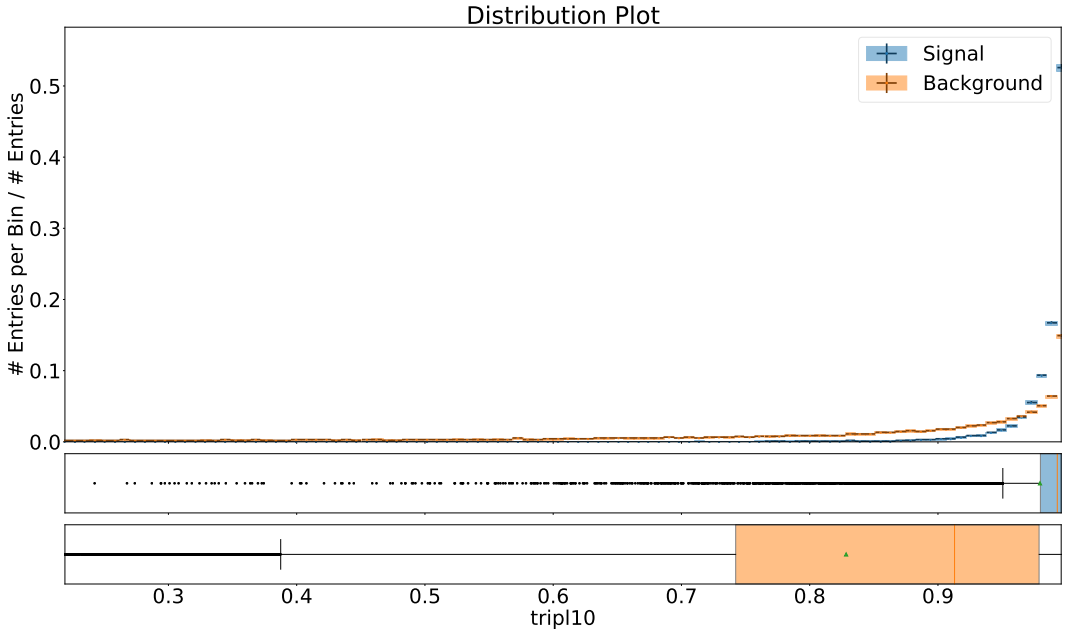
2.27 tripletFit_PMag



2.28 tripletFit_Chi2



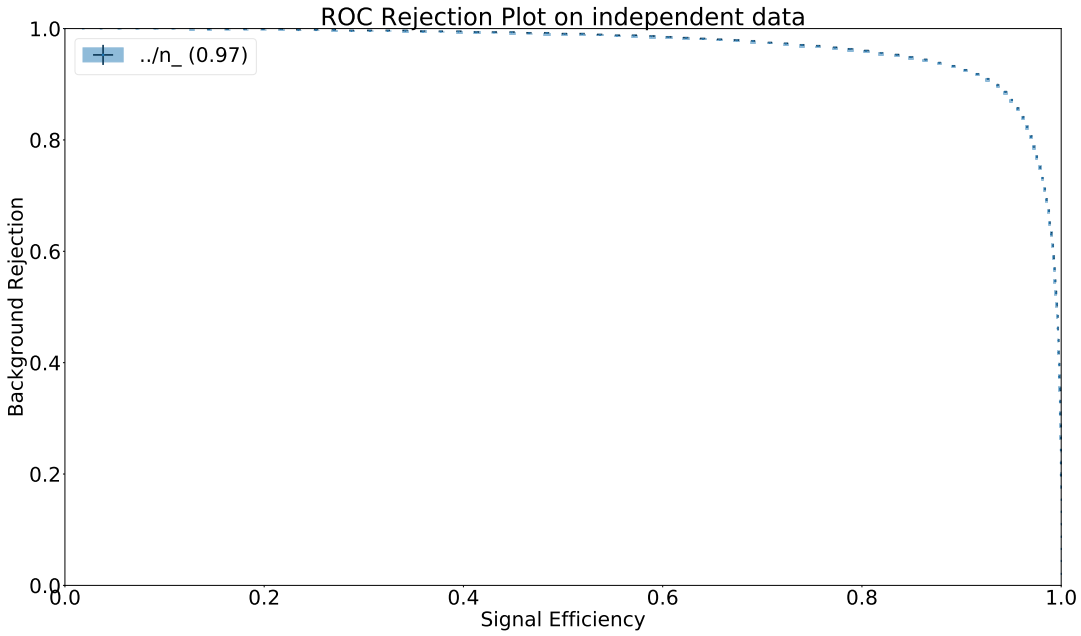
2.29 tripletFit_QI



3 Classifier Plot

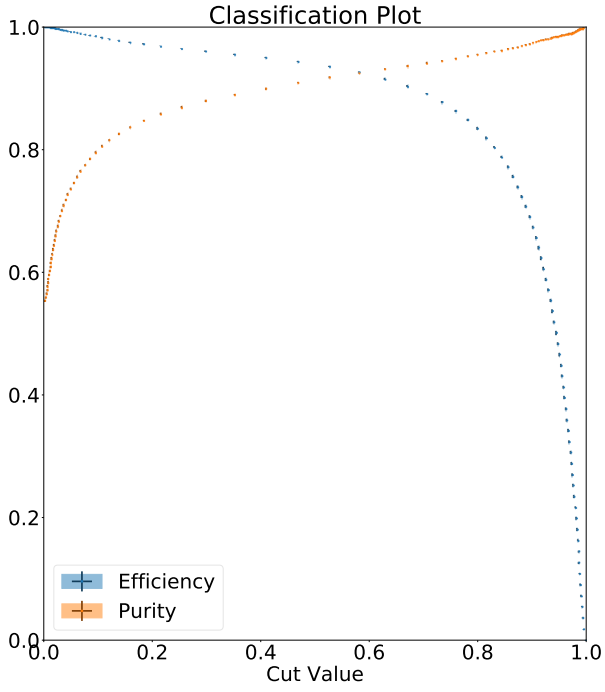
This section contains the receiver operating characteristics (ROC), purity projection, ...of the classifiers on training and independent data.The legend of each plot contains the shortened identifier and the area under the ROC curvein parenthesis.

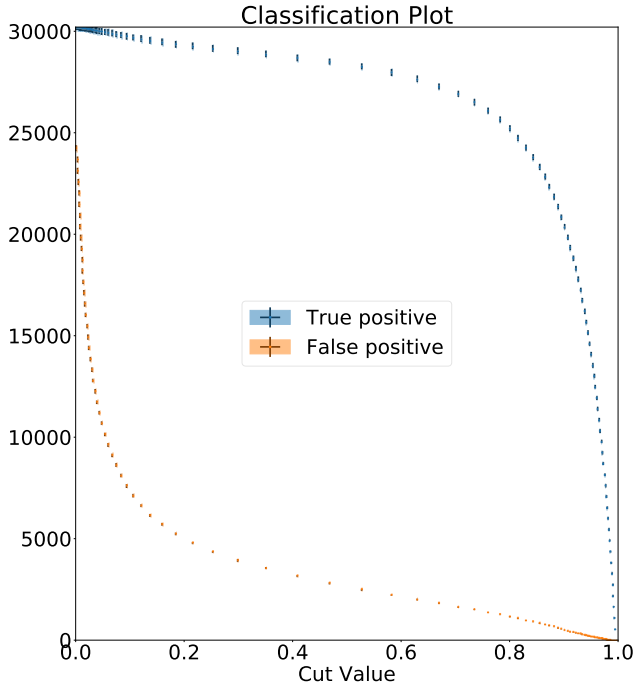
4 ROC Plot



5 Classification Results

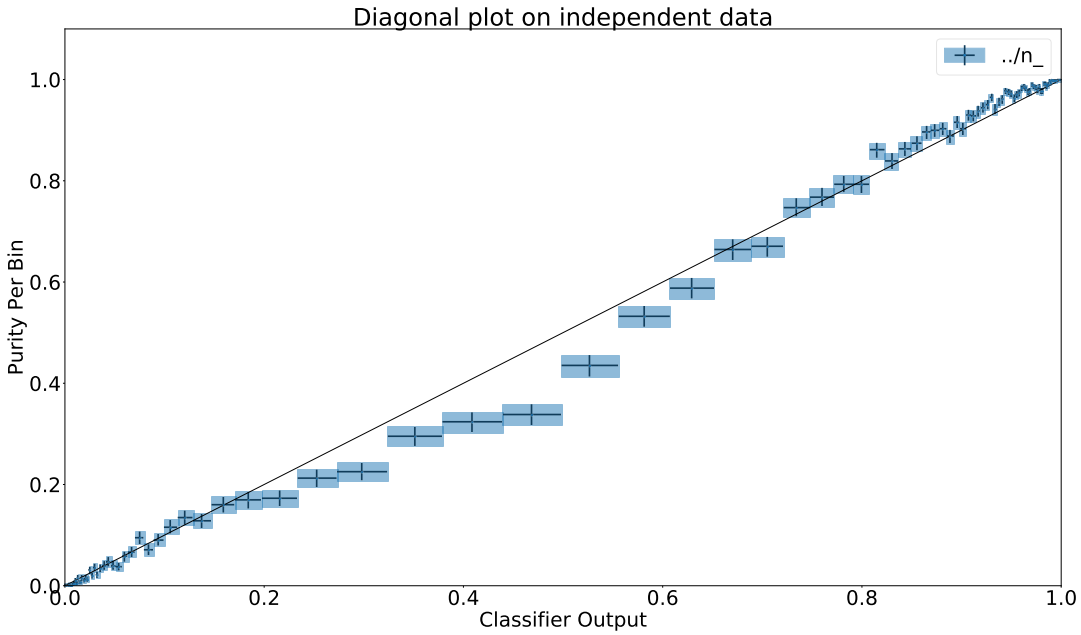
5.1 ../n_





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