TMVA analysis of void characteristics

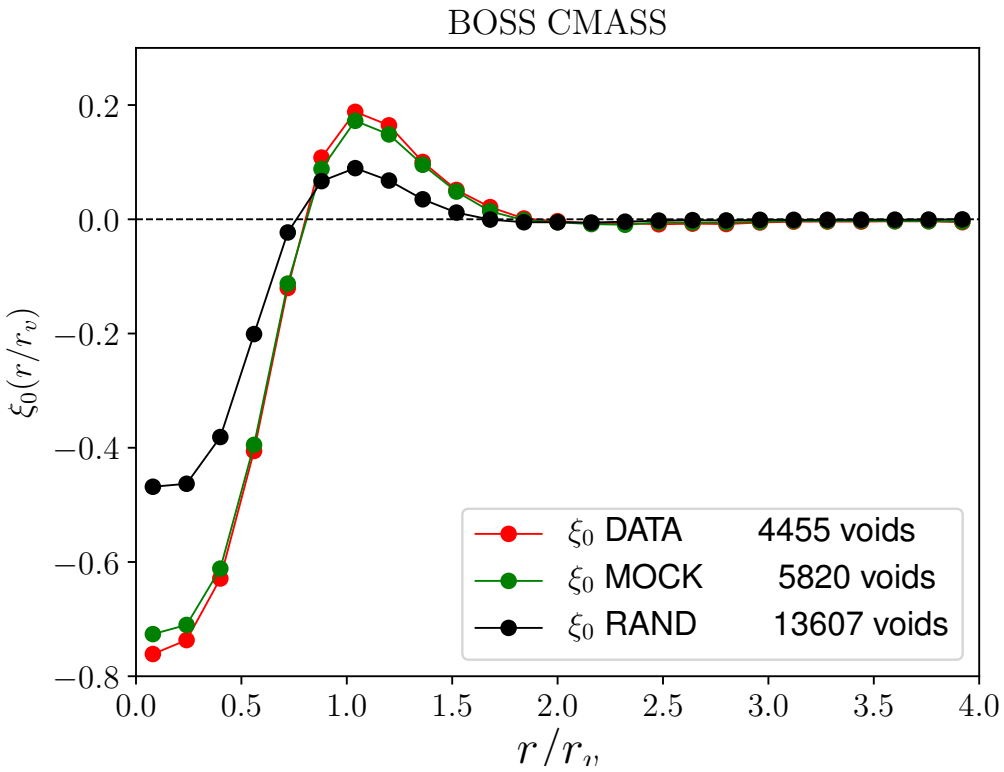
arXiv 1805.07181

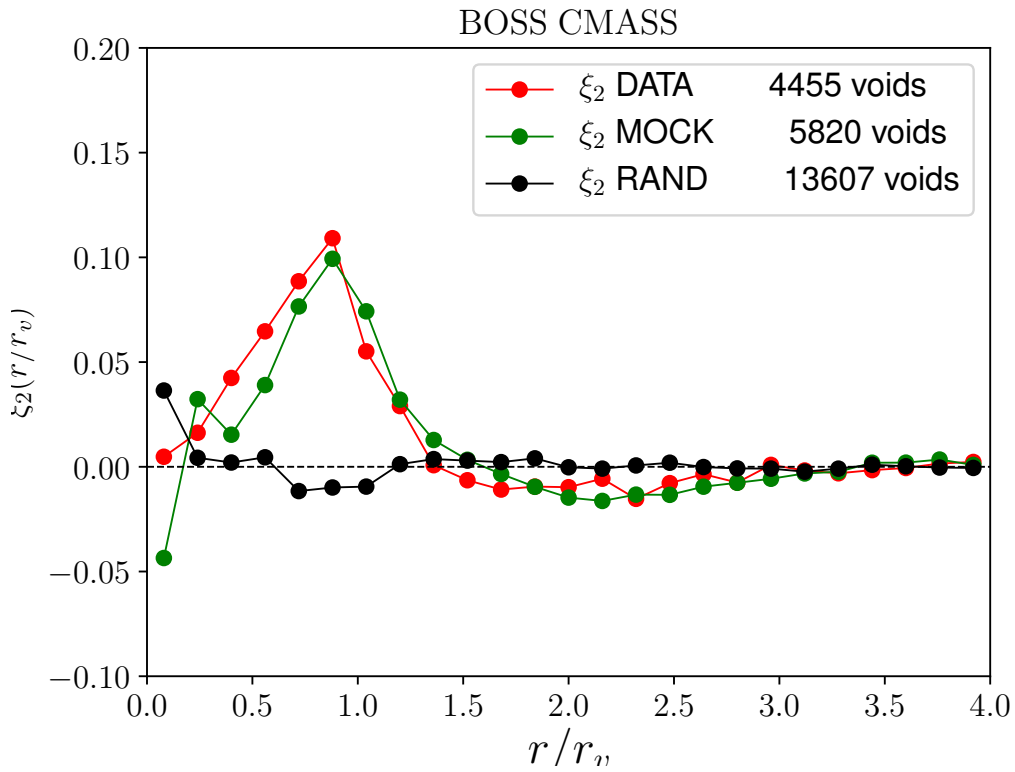
Void finder ZOBOV - VIDE

Effect of an possible contamination in a sample of voids obtained by running VIDE on a galaxy catalog by Poisson noise voids on an analysis

Often cuts on variables describing the size of the voids are applied to remove Poisson noise voids. For example a cut on the radius of the void : Rv > 2* mean particle separation .

Comparison of void-galaxy multipole distributions: voids from galaxy catalog, mock catalog, random catalog





Mock001 multidark patchy 1.356.442 galaxies 5820 voids
Random multidark patchy 1.356.084 randoms 13607 voids
DR12v5 Data 849.637 galaxies 4455 voids
using the « untrimmed_centers_central »
and « untrimmed_voidDesc_central » files

Variables : V_{norm}, $\rho_{contrast}$, $N_{particules}$, Z_{red}, prob(void), coredens

Two Methods tested: BDT and MLP using the variables: V_{norm}, $\rho_{contrast}$, $N_{particules}$, Z_{red}, prob(void), coredens

- Signal : 5820 voids from mock
- Background : 13607 voids from random catalog

```
TCut mycut = "volnorm < 500. && rho_cont < 5. && centrho < 50. && numpart < 1000.";
```

- Preparing the training and test samples

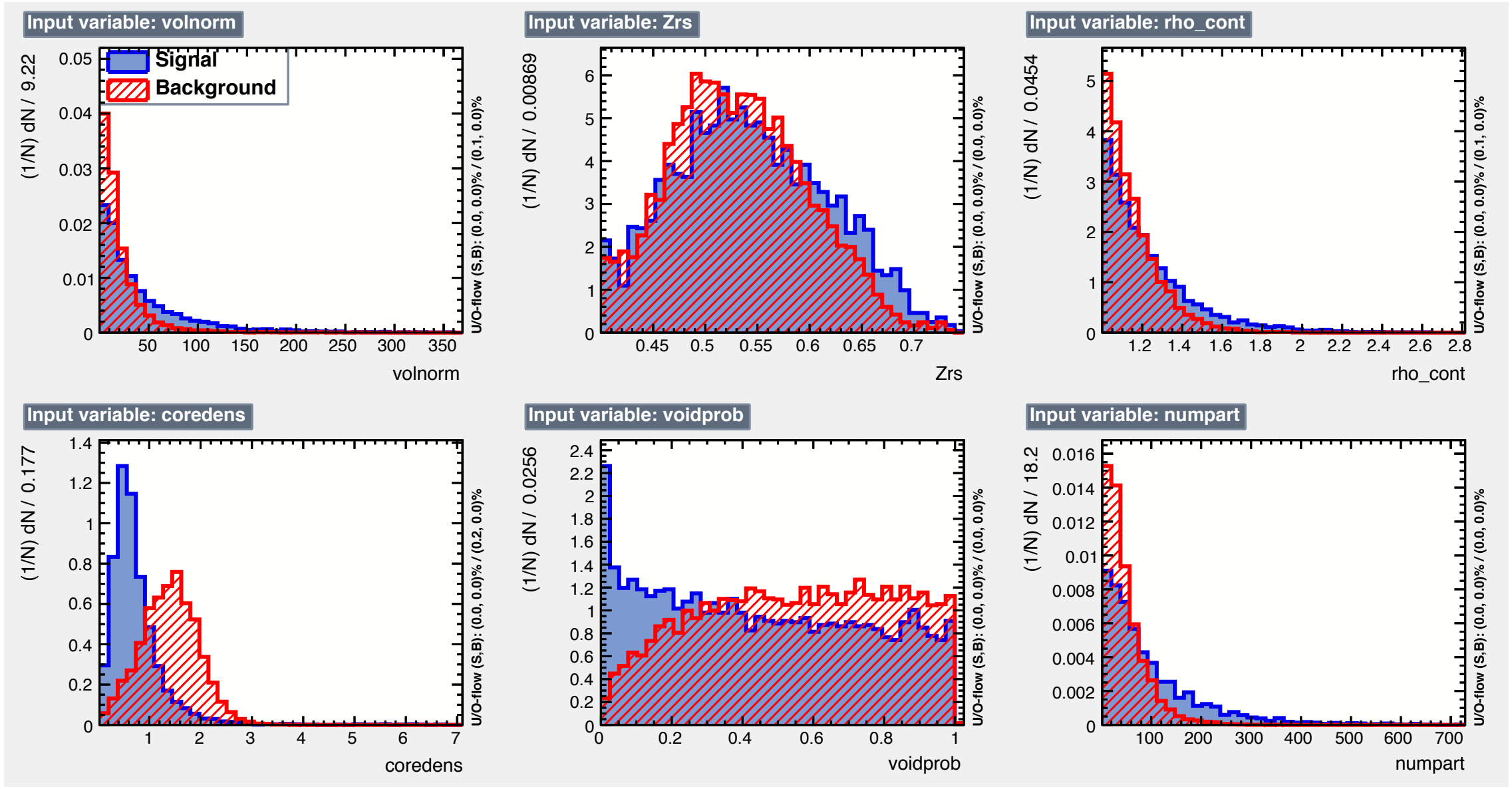
```
dataloader->PrepareTrainingAndTestTree(mycut,"NormMode=EqualNumEvents:SplitMode=Alternate");
```

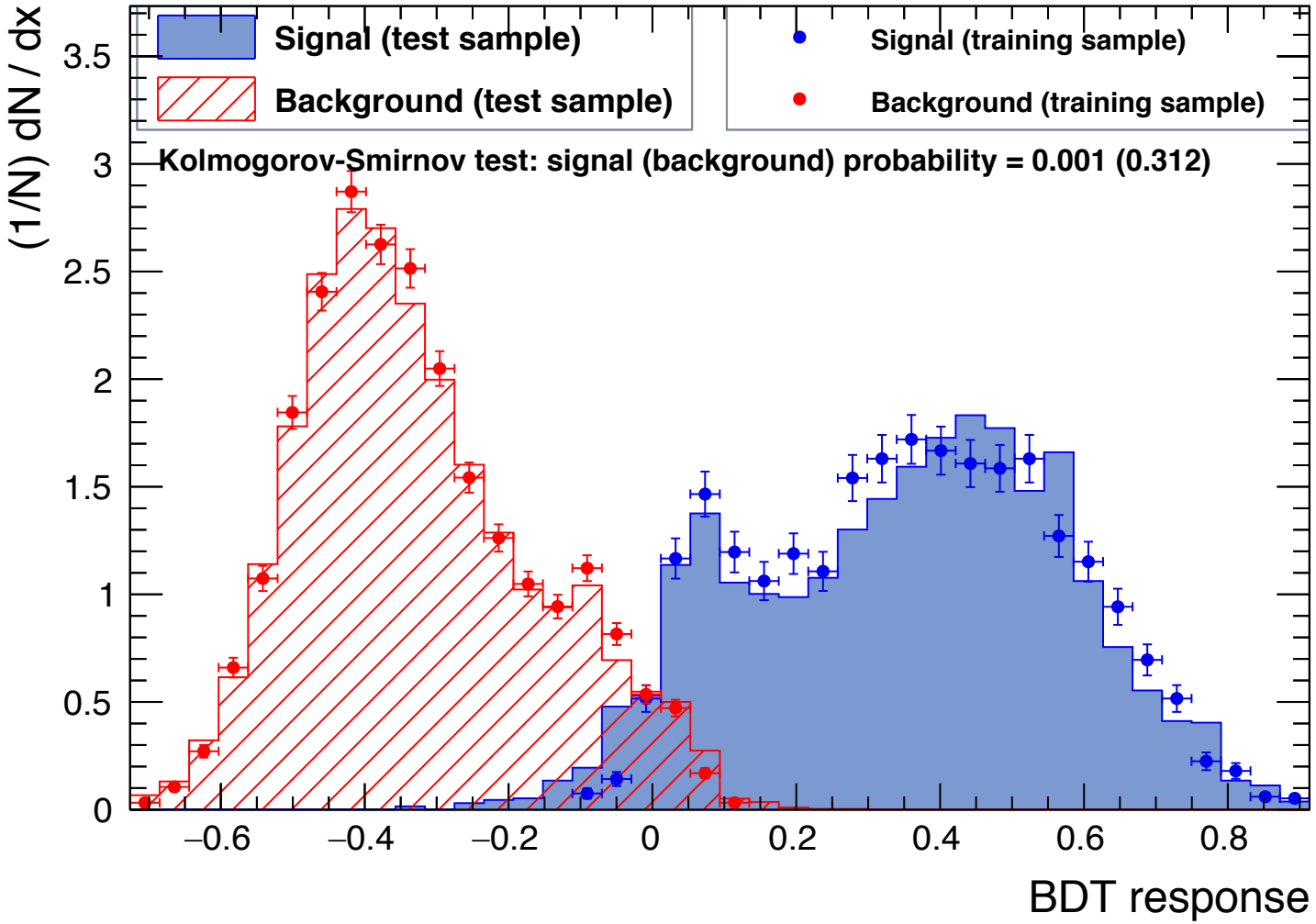
- Boost Decision Tree (BDT)

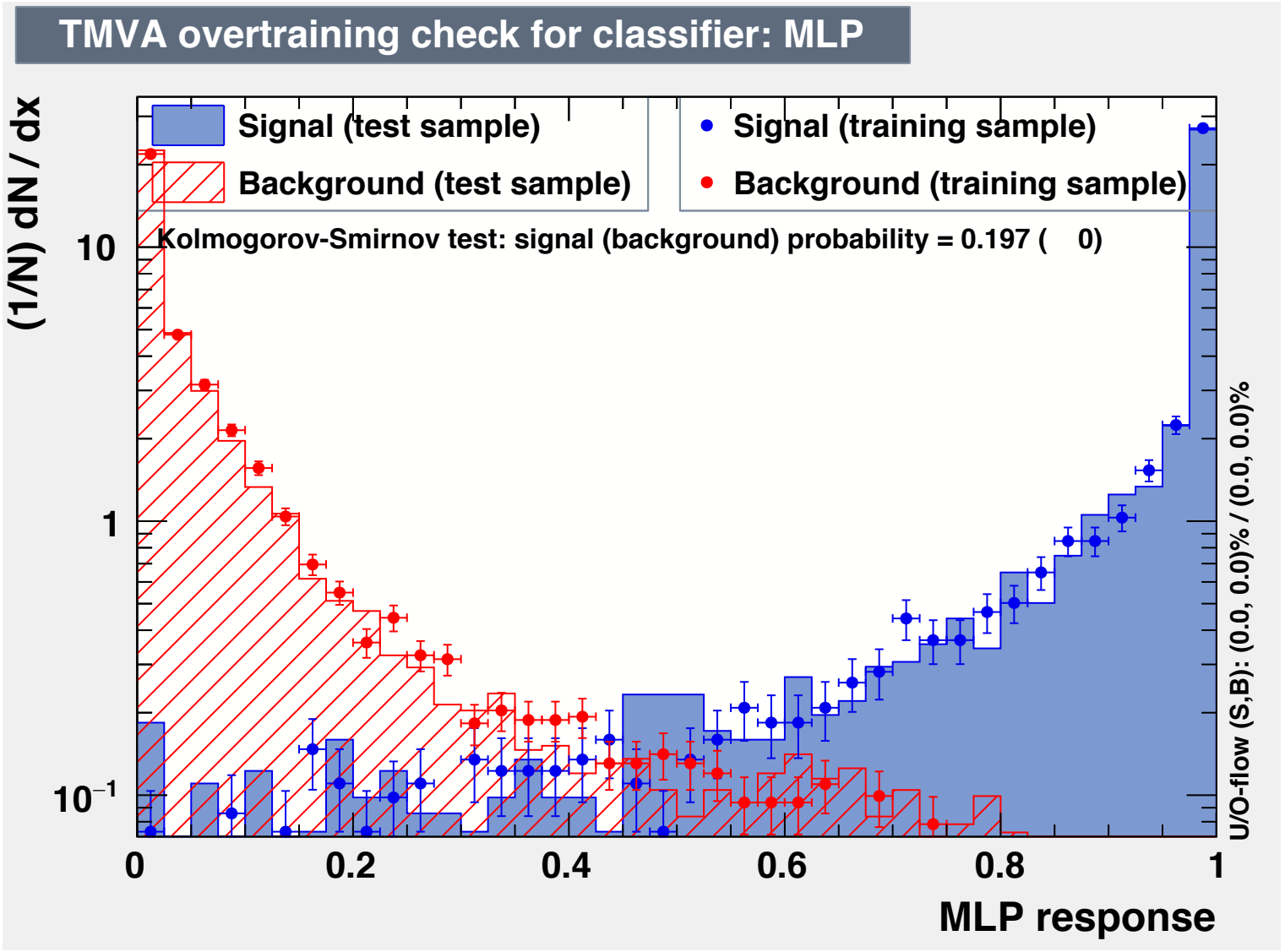
```
factory->BookMethod(dataloader,
TMVA::Types::kBDT,"BDT700","H:V:NTrees=700:MinNodeSize=4%:MaxDepth=5:BoostType=AdaBoost:AdaBoost
Beta=0.15:nCuts=100");
```

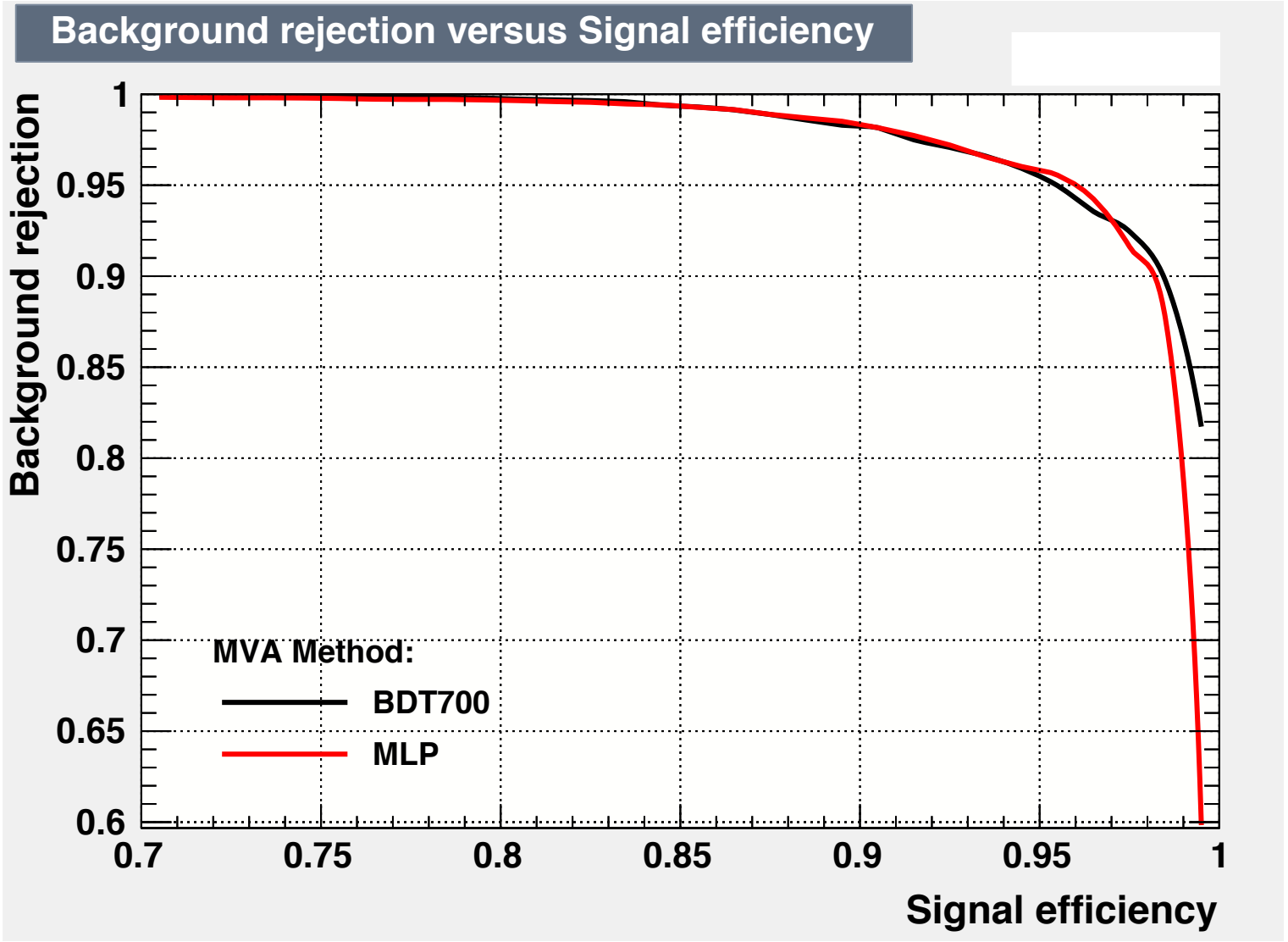
- Neural Network (MLP)

```
factory->BookMethod(dataloader, TMVA::Types::kMLP, "MLP",
"H:V:NeuronType=tanh:VarTransform=N:NCycles=600:HiddenLayers=N+5:TestRate=5:UseRegulator");
```

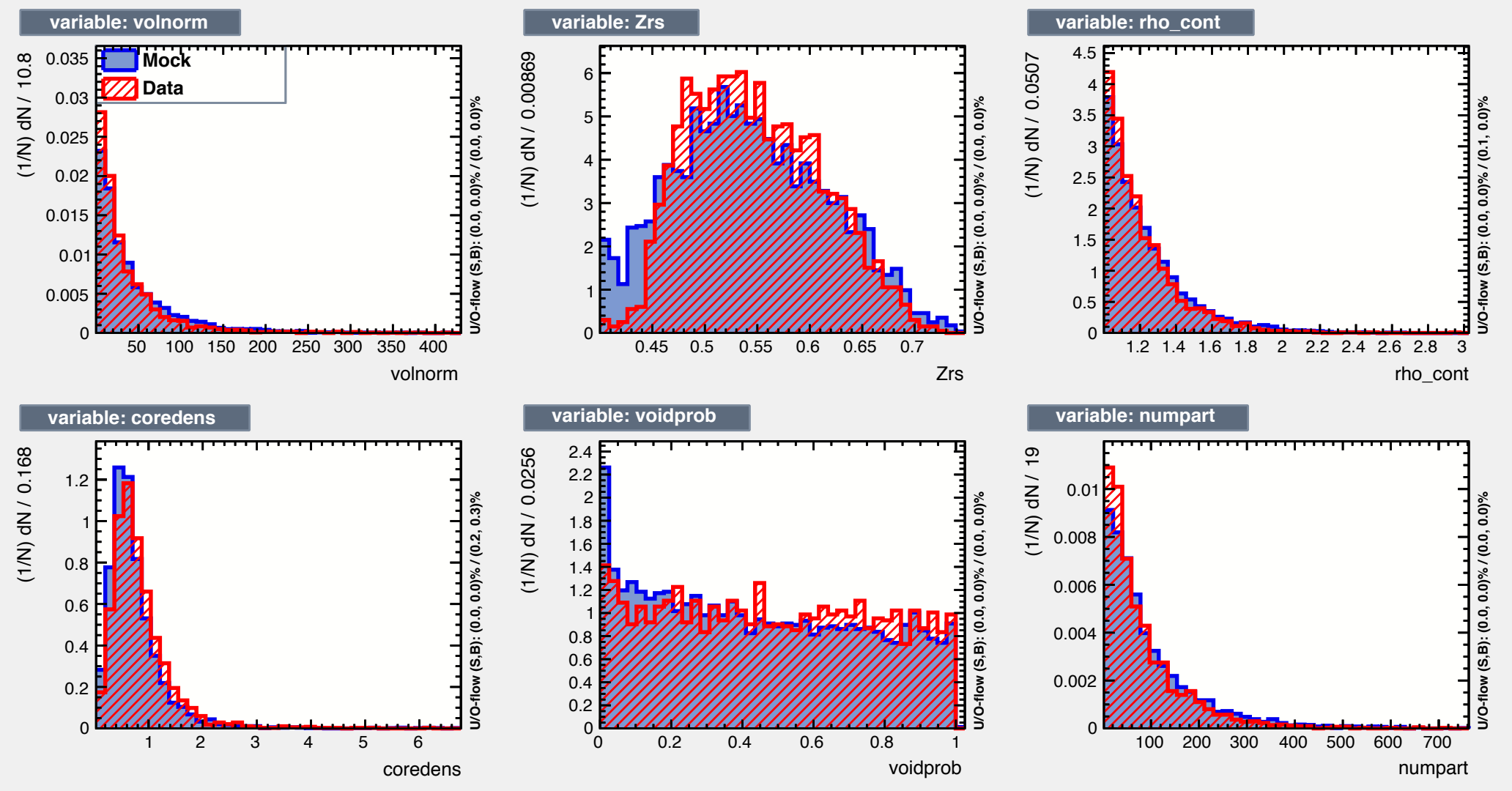

Input variable: volnorm
Signal
Background
(1/N) dN / 9.22
volnorm
U/O-flow (S,B): (0.0, 0.0)% / (0.1, 0.0)%
Input variable: Zrs
(1/N) dN / 0.00869
Zrs
U/O-flow (S,B): (0.0, 0.0)% / (0.0, 0.0)%
Input variable: rho_cont
(1/N) dN / 0.0454
rho_cont
U/O-flow (S,B): (0.0, 0.0)% / (0.1, 0.0)%
Input variable: coredens
(1/N) dN / 0.177
coredens
U/O-flow (S,B): (0.0, 0.0)% / (0.2, 0.0)%
Input variable: voidprob
(1/N) dN / 0.0256
voidprob
U/O-flow (S,B): (0.0, 0.0)% / (0.0, 0.0)%
Input variable: numpart
(1/N) dN / 18.2
numpart
U/O-flow (S,B): (0.0, 0.0)% / (0.0, 0.0)%


Signal (test sample)
Background (test sample)
Signal (training sample)
Background (training sample)
Kolmogorov-Smirnov test: signal (background) probability = 0.001 (0.312)
(1/N) dN / dx
BDT response
0
0.5
1
1.5
2
2.5
3
3.5
−0.6
−0.4
−0.2
0
0.2
0.4
0.6
0.8

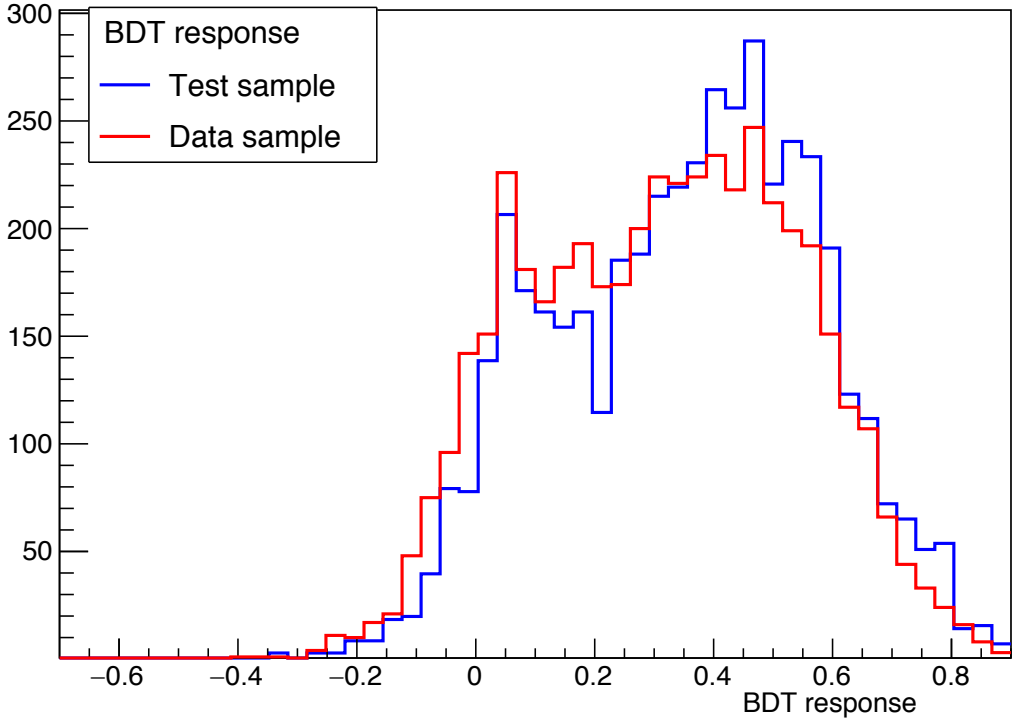

TMVA overtraining check for classifier: MLP
Signal (test sample)
Background (test sample)
Signal (training sample)
Background (training sample)
Kolmogorov-Smirnov test: signal (background) probability = 0.197 (0)
(1/N) dN / dx
10
1
10^{-1}
0
0.2
0.4
0.6
0.8
1
MLP response
U/O-flow (S,B): (0.0, 0.0)% / (0.0, 0.0)%

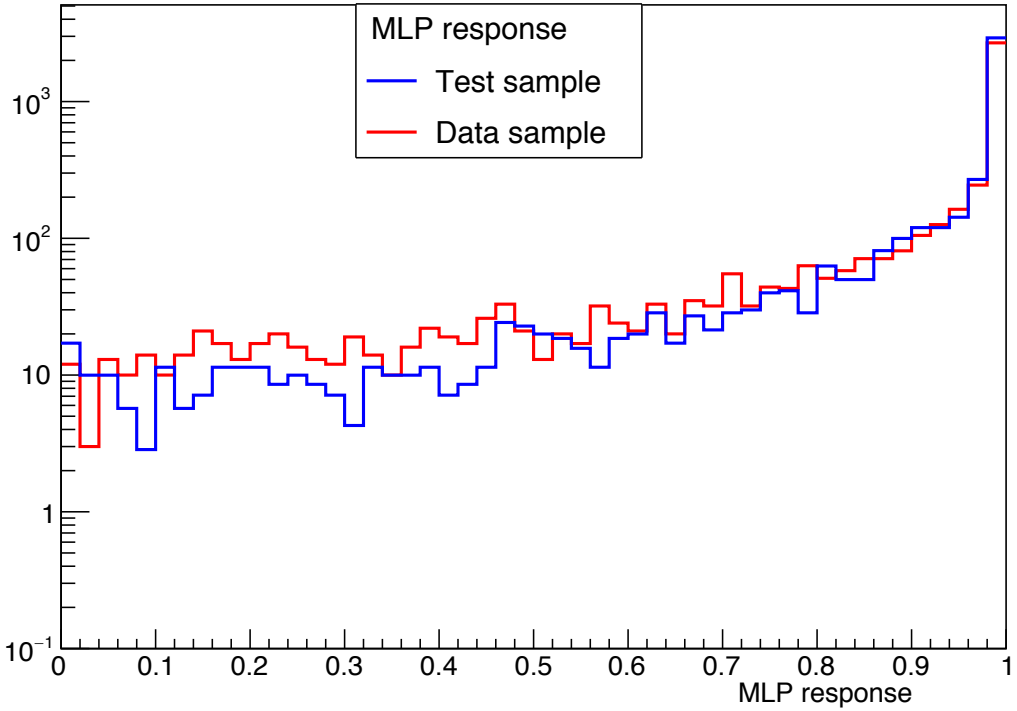

Background rejection versus Signal efficiency
Background rejection
1
0.95
0.9
0.85
0.8
0.75
0.7
0.65
0.6
0.7
0.75
0.8
0.85
0.9
0.95
1
Signal efficiency
MVA Method:
BDT700
MLP

Comparison of the variables of the mock and data samples

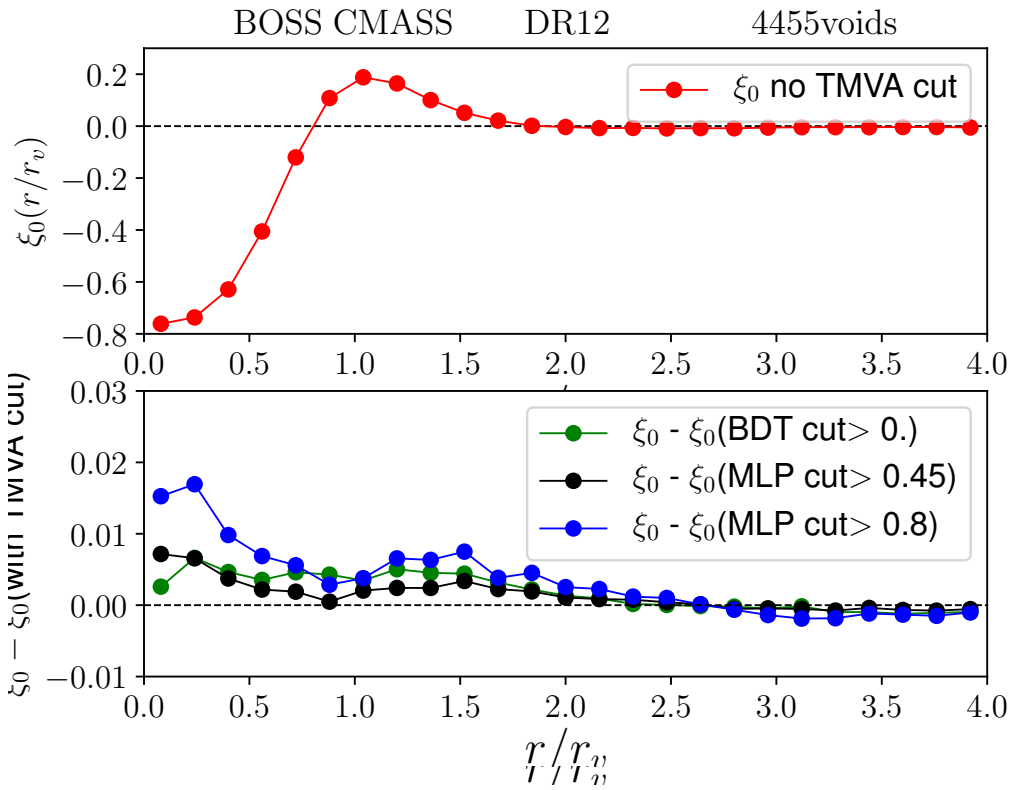


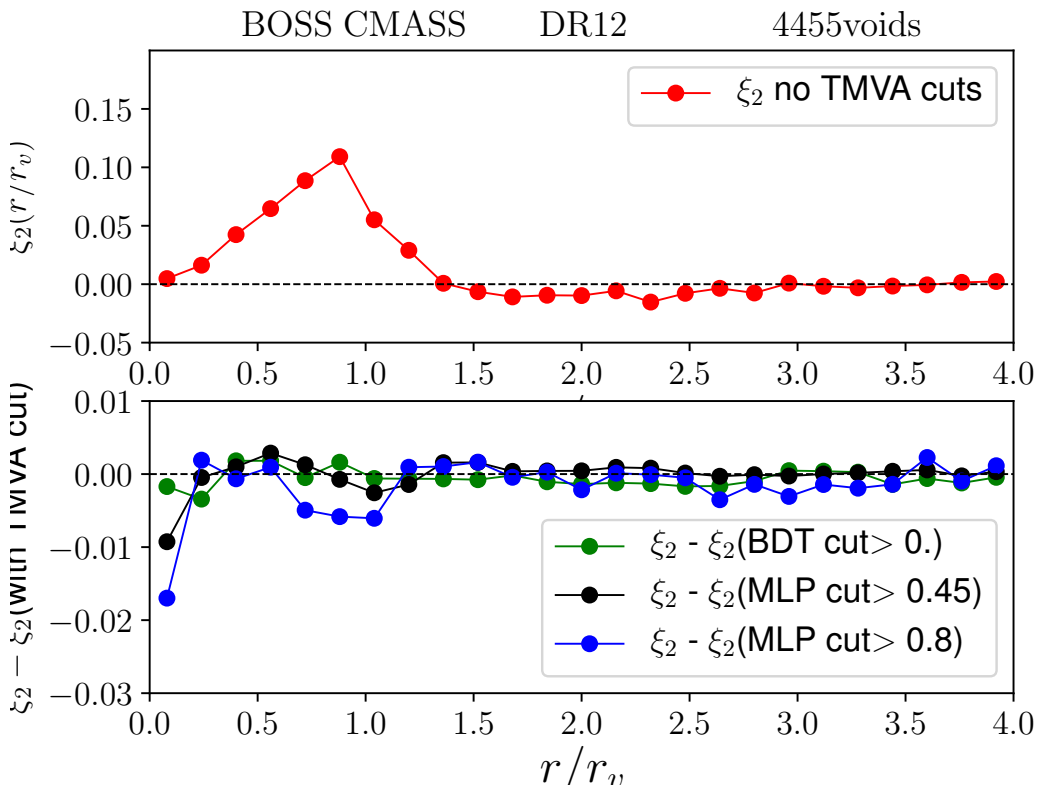
Response of the BDT and MLP analyses on data
(data and test samples renormalised to a same number of events)





Effect of the neural network cut on the void-galaxy multipole distributions

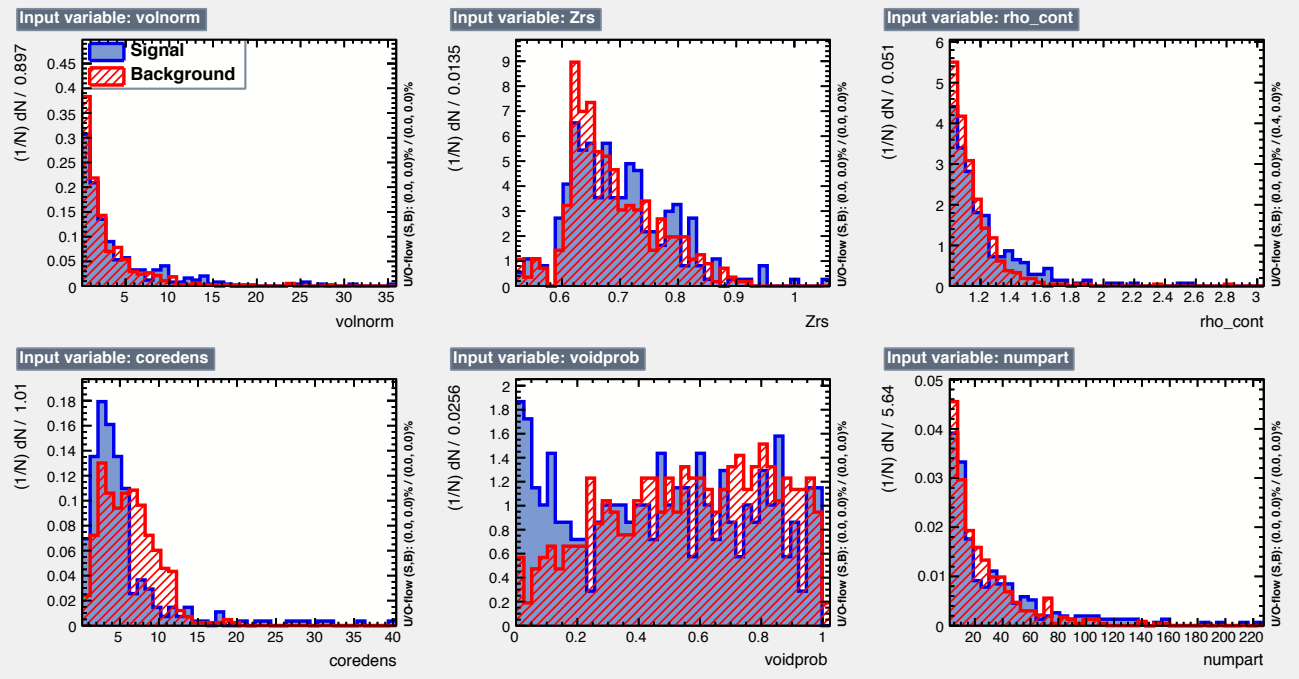


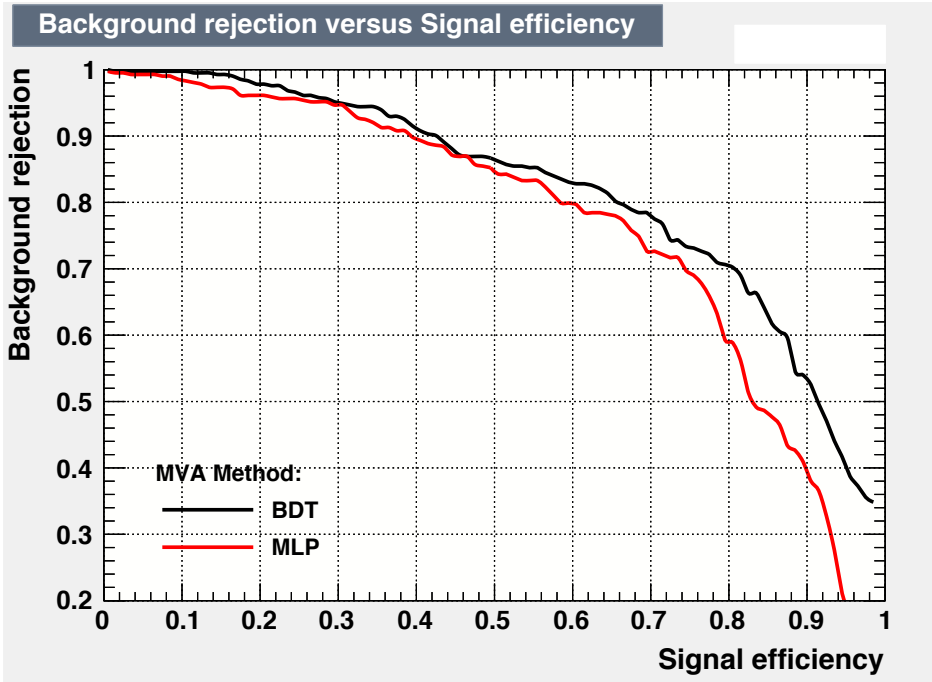


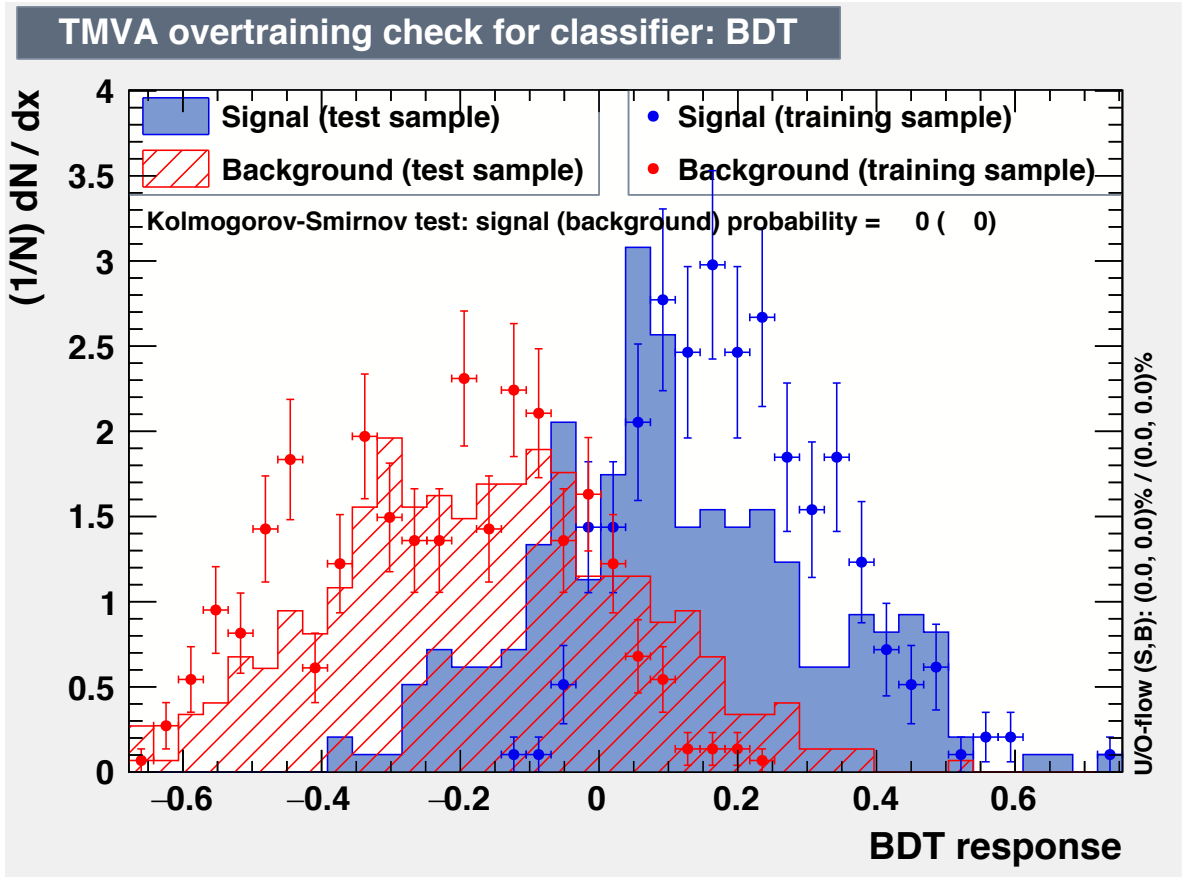
eBOSS LRG 138394 galaxies 572 voids from the LRG catalog
875 voids from the random catalog

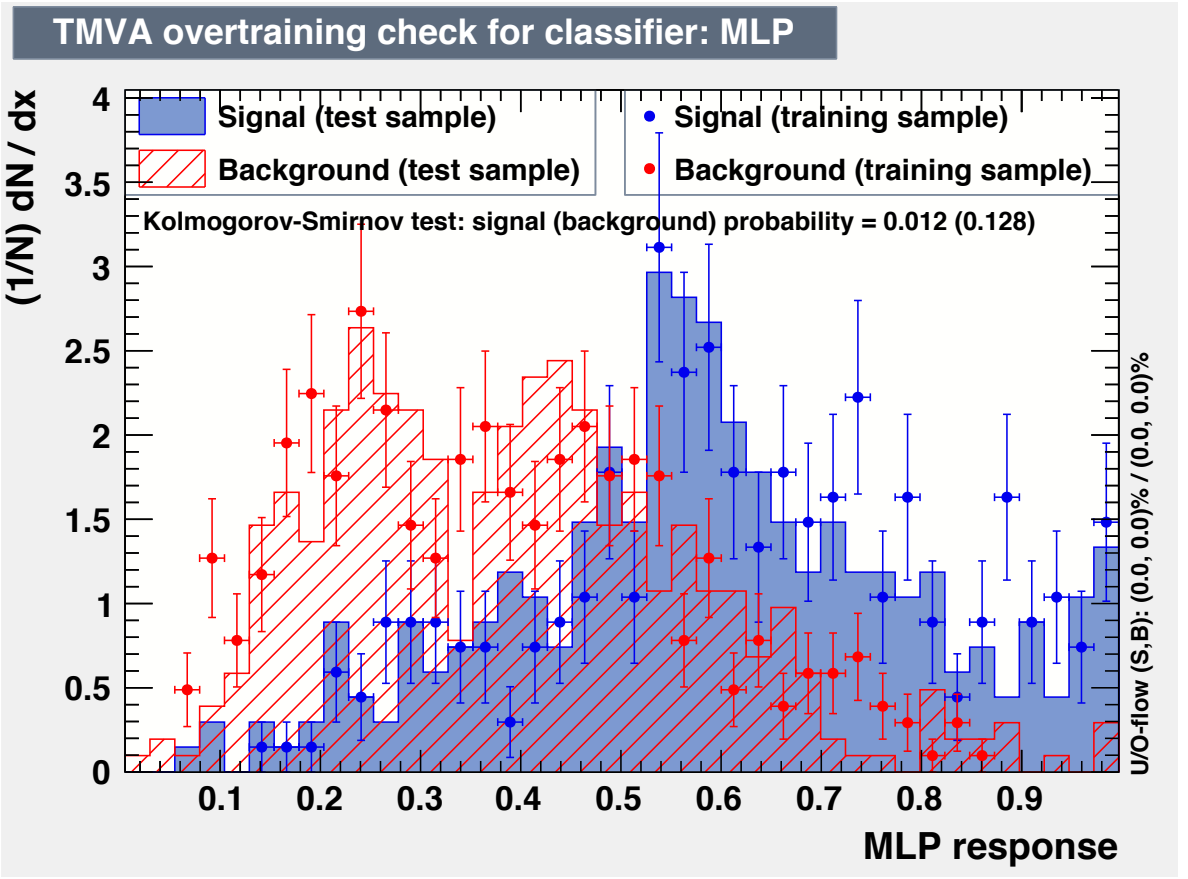
BOSS CMASS 849634 galaxies 4500 voids from the CMASS catalog
13000 voids from the random catalog

and $\langle\rho\rangle_{CMASS} \simeq 7 * \langle\rho\rangle_{LRG}$






TMVA overtraining check for classifier: BDT
Signal (test sample)
Signal (training sample)
Background (test sample)
Background (training sample)
Kolmogorov-Smirnov test: signal (background) probability = 0 (0)
(1/N) dN / dx
U/O-flow (S,B): (0.0, 0.0)% / (0.0, 0.0)%
BDT response


TMVA overtraining check for classifier: MLP
Signal (test sample)
Signal (training sample)
Background (test sample)
Background (training sample)
Kolmogorov-Smirnov test: signal (background) probability = 0.012 (0.128)
(1/N) dN / dx
U/O-flow (S,B): (0.0, 0.0)% / (0.0, 0.0)%
MLP response