

# Progress Report on Tau Final States of TTTT

Fabio Lemmi<sup>1</sup>   Huiling Hua<sup>1</sup>   Hongbo Liao<sup>1</sup>   Hideki Okawa<sup>2</sup>  
Yu Zhang<sup>2</sup>

<sup>1</sup>IHEP

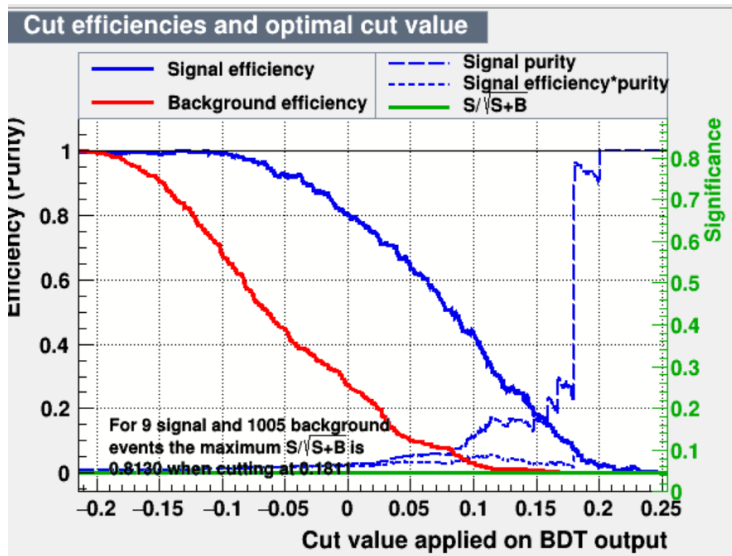
<sup>2</sup>Fudan University

IHEP Group Meeting, 2020

# Outline

- 1 Problem from last time
- 2 Event Yield
- 3 1Tau0L
  - step 1
  - step 2

## Problems



# TMVA Problems Solved

- Modified root5 TMVA plotting macro to be used in root6 TMVA
- Spotted a small bug that caused the flat significance line.
- the original code doesn't check the sanity of significance before the fill, which is totally fine for positive weighted events. but our sample has significant amount of negative weighted events, which could cause some of the significance value to be NaN, which in return makes the draw of histogram into a flat line
- So I added the sanity check before the fill and problem solved.
- Lorenzo in ROOT forum replied to my solution and guess the fix would be integrated in future ROOT TMVA



moneta 

Hi,

The fixes makes sense to me. I will fix also in the ROOT repository

Thank you and best regards

Lorenzo

# EventSelection

- MET fillters
- HLT requirements
  - `HLT_PFHT450_SixJet40_BTagCSV_p056==1,`  
`HLT_PFHT400_SixJet30_DoubleBTagCSV_p056==1`
- loose preselection
  - `tausL.size()>0, jets.size()>3, bjetsL.size()>1`
- Subchannel requirements

# MC reweighting

- genWeight
- prefireWeight
- PileUp reweighting

## Event Yield

```

1Tau0L
Raw entries:
TTTT      = 237389
TT        = 46656
TTX       = 37788
single top = 3839
TX        = 7785
DYJets    = 0
VV        = 8
VVV       = 98
QCD       = 305
Total BKG = 7777.06

Weighted:
TTTT      = 1729.23
TT        = 1.30564e+07
TTX       = 55880.3
single top = 183.319
TX        = 7709.81
DYJets    = 0
VV        = 714.399
VVV       = 10.5086
QCD       = 268.521
Total BKG = 13936.3

Event yield:
TTTT      = 9.47371
TT        = 6276.17
TTX       = 212.41
single top = 120.653
TX        = 13.0152
DYJets    = 0
VV        = 0.103319
VVV       = 0.982638
QCD       = 7285.38
Total BKG = 13936.3

```

```

1Tau1E
Raw entries:
TTTT      = 80344
TT        = 16515
TTX       = 10415
single top = 3361
TX        = 2598
DYJets    = 0
VV        = 3
VVV       = 20
QCD       = 2
Total BKG = 93.7667

Weighted:
TTTT      = 623.886
TT        = 2.18092e+06
TTX       = 13611.5
single top = 147.115
TX        = 2591.97
DYJets    = 0
VV        = 415.464
VVV       = 1.81758
QCD       = 2.17447
Total BKG = 1178.13

Event yield:
TTTT      = 3.418
TT        = 1092.39
TTX       = 48.6264
single top = 21.8069
TX        = 3.4963
DYJets    = 0
VV        = 0.0537144
VVV       = 0.249651
QCD       = 2.35911
Total BKG = 1178.13

```

```

1Tau1Mu
Raw entries:
TTTT      = 99667
TT        = 19225
TTX       = 11729
single top = 2529
TX        = 3055
DYJets    = 0
VV        = 6
VVV       = 12
QCD       = 0
Total BKG = 92.2959

Weighted:
TTTT      = 780.561
TT        = 2.46514e+06
TTX       = 15187.2
single top = 139.593
TX        = 2971.26
DYJets    = 0
VV        = 465.544
VVV       = 0.622918
QCD       = 0
Total BKG = 1324.63

Event yield:
TTTT      = 4.27635
TT        = 1236.92
TTX       = 54.2688
single top = 18.7832
TX        = 4.01954
DYJets    = 0
VV        = 0.0786659
VVV       = 0.0812971
QCD       = 0
Total BKG = 1324.63

```

## Event Yield

```

1Tau20S
Raw entries:
TTTT      = 28250
TT        = 1265
TTX       = 4718
single top = 3510
TX        = 789
DYJets    = 0
VV        = 1
VVV       = 5
QCD       = 0
Total BKG = 28.0118

```

```

Weighted:
TTTT      = 221.854
TT        = 92102.1
TTX       = 4935.03
single top = 199.663
TX        = 778.825
DYJets    = 0
VV        = 1.02674
VVV       = 0.650436
QCD       = 0
Total BKG = 78.0374

```

```

Event yield:
TTTT      = 1.21544
TT        = 58.7553
TTX       = 14.9337
single top = 0.508183
TX        = 0.962345
DYJets    = 0
VV        = 0.0375234
VVV       = 0.0953384
QCD       = 0
Total BKG = 78.0374

```

```

1Tau2SS
Raw entries:
TTTT      = 14204
TT        = 63
TTX       = 1321
single top = 520
TX        = 385
DYJets    = 0
VV        = 0
VVV       = 3
QCD       = 0
Total BKG = 7.59988

```

```

Weighted:
TTTT      = 114.525
TT        = 6095.78
TTX       = 1501.54
single top = 18.4616
TX        = 386.36
DYJets    = 0
VV        = 0
VVV       = 0.421751
QCD       = 0
Total BKG = 10.1519

```

```

Event yield:
TTTT      = 0.627434
TT        = 3.35655
TTX       = 4.78014
single top = 0.0469887
TX        = 0.498109
DYJets    = 0
VV        = 0
VVV       = 0.0625488
QCD       = 0
Total BKG = 10.1519

```

```

1Tau3L
Raw entries:
TTTT      = 2848
TT        = 0
TTX       = 355
single top = 26
TX        = 27
DYJets    = 0
VV        = 0
VVV       = 1
QCD       = 0
Total BKG = 1.4098

```

```

Weighted:
TTTT      = 23.2214
TT        = 0
TTX       = 358.464
single top = 2.24994
TX        = 25.843
DYJets    = 0
VV        = 0
VVV       = 0.0168772
QCD       = 0
Total BKG = 1.21668

```

```

Event yield:
TTTT      = 0.12722
TT        = 0
TTX       = 1.02004
single top = 0.00572658
TX        = 0.0294773
DYJets    = 0
VV        = 0
VVV       = 0.00242049
QCD       = 0
Total BKG = 1.21668

```



## Event Yield

```

2Tau0L
Raw entries:
TTTT      = 11422
TT        = 2848
TTX       = 4288
single top = 2796
TX        = 1575
DYJets    = 0
VV        = 1
VVV       = 4
QCD       = 2
Total BKG = 37.7907

```

```

Weighted:
TTTT      = 92.454
TT        = 500241
TTX       = 4980.38
single top = 158.423
TX        = 1562.55
DYJets    = 0
VV        = 1.14961
VVV       = 0.414473
QCD       = 1.95001
Total BKG = 278.15

```

```

Event yield:
TTTT      = 0.506514
TT        = 246.776
TTX       = 16.2812
single top = 6.77039
TX        = 2.56187
DYJets    = 0
VV        = 0.042014
VVV       = 0.0348188
QCD       = 1.41789
Total BKG = 278.15

```

```

2Tau1E
Raw entries:
TTTT      = 2569
TT        = 180
TTX       = 1007
single top = 459
TX        = 378
DYJets    = 0
VV        = 0
VVV       = 0
QCD       = 0
Total BKG = 5.83133

```

```

Weighted:
TTTT      = 21.8009
TT        = 22297.4
TTX       = 1020.17
single top = 29.586
TX        = 372.931
DYJets    = 0
VV        = 0
VVV       = 0
QCD       = 0
Total BKG = 16.2041

```

```

Event yield:
TTTT      = 0.119437
TT        = 11.5052
TTX       = 2.98042
single top = 0.0753027
TX        = 0.495236
DYJets    = 0
VV        = 0
VVV       = 0
QCD       = 0
Total BKG = 16.2041

```

```

2Tau1Mu
Raw entries:
TTTT      = 3264
TT        = 190
TTX       = 1021
single top = 442
TX        = 406
DYJets    = 0
VV        = 0
VVV       = 0
QCD       = 0
Total BKG = 5.98391

```

```

Weighted:
TTTT      = 25.5952
TT        = 21044
TTX       = 1128.94
single top = 24.0248
TX        = 394.063
DYJets    = 0
VV        = 0
VVV       = 0
QCD       = 0
Total BKG = 16.4386

```

```

Event yield:
TTTT      = 0.140225
TT        = 11.0659
TTX       = 3.52228
single top = 0.0611482
TX        = 0.530434
DYJets    = 0
VV        = 0
VVV       = 0
QCD       = 0
Total BKG = 16.4386

```

## Event Yield

```

2Tau20S
Raw entries:
TTTT      = 449
TT        = 1
TTX       = 111
single top = 8
TX        = 28
DYJets    = 0
VV        = 0
VVV       = 1
QCD       = 0
Total BKG = 0.59826

Weighted:
TTTT      = 3.23368
TT        = 66.1142
TTX       = 110.585
single top = 1.52484
TX        = 30.1004
DYJets    = 0
VV        = 0
VVV       = 0.0187778
QCD       = 0
Total BKG = 0.484251

Event yield:
TTTT      = 0.0177159
TT        = 0.042666
TTX       = 0.31452
single top = 0.00388105
TX        = 0.031774
DYJets    = 0
VV        = 0
VVV       = 0.00269306
QCD       = 0
Total BKG = 0.484251

```

```

2Tau2SS
Raw entries:
TTTT      = 190
TT        = 0
TTX       = 13
single top = 1
TX        = 4
DYJets    = 0
VV        = 0
VVV       = 0
QCD       = 0
Total BKG = 0.0644176

Weighted:
TTTT      = 1.75733
TT        = 0
TTX       = 11.9666
single top = -0.00338024
TX        = 2.99069
DYJets    = 0
VV        = 0
VVV       = 0
QCD       = 0
Total BKG = 0.0578978

Event yield:
TTTT      = 0.00962764
TT        = 0
TTX       = 0.0344717
single top = -8.60342e-06
TX        = 0.00315698
DYJets    = 0
VV        = 0
VVV       = 0
QCD       = 0
Total BKG = 0.0578978

```

# MVA Stragdy

- put all 20 most discriminate variables into BDT, see the performance
- remove the correlation  $> 80$  variables, then see the performance
- add some variables

# 1tau0l

```
TCut mycuts = "tausT_number==1 leptonsMVAT_number==0  
jets_number>=8 bjetsM_num>=2" 1tau0l
```

# Outline

- 1 Problem from last time
- 2 Event Yield
- 3 1Tau0L
  - step 1
  - step 2

# TMVA Setup

- Signal: TTTT\_TuneCP5;
- Background: all bg except H and HH and minor ones
- 70% goes to training and 30% goes to testing
- Global weight and event weight same in event yield calculation
- have added more interesting variables:
  - sphericity, aplanality
  - added some nonjets variables
- Dealing with negative weight events
- use the 20 most powerful variables as input for BDT( not including the angular variables)

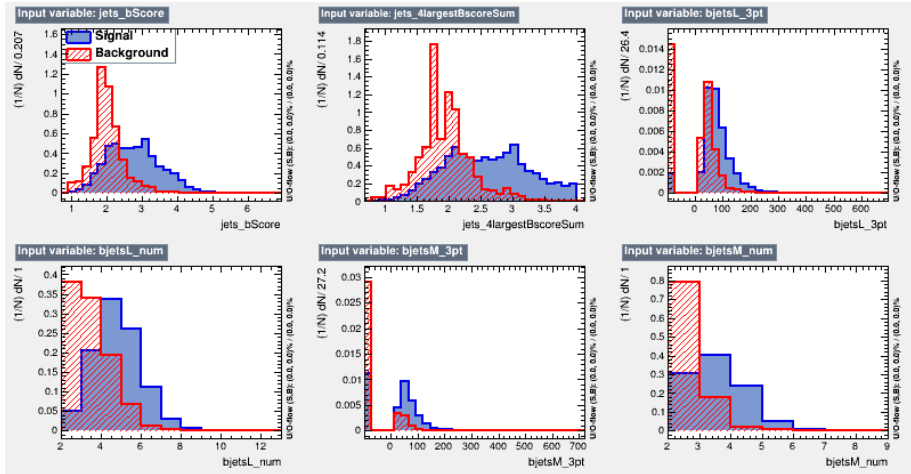
# Variables Separation Power Ranking

Rank	Variable	Separation
1	jets_bScore	: 3.249e-01
2	jets_4largestBscoreSum	: 3.064e-01
3	bjetsL_3pt	: 2.814e-01
4	bjetsL_num	: 2.729e-01
5	bjetsM_3pt	: 2.716e-01
6	bjetsM_num	: 2.646e-01
7	bjetsL_4pt	: 2.579e-01
8	bjetsL_invariantMass	: 2.512e-01
9	bjetsM_invariantMass	: 2.497e-01
10	bjetsM_3phi	: 2.488e-01
11	bjetsL_HT	: 2.447e-01
12	bjetsM_3eta	: 2.444e-01
13	bjetsL_transMass	: 2.292e-01
14	bjetsL_4phi	: 2.278e-01
15	toptagger_transMass	: 2.258e-01
16	toptagger_HT	: 2.191e-01
17	bjetsL_4eta	: 2.181e-01
18	jets_7pt	: 2.137e-01
19	bjetsM_HT	: 2.118e-01
20	jets_6pt	: 2.111e-01
21	toptagger_invariantMass	: 1.999e-01
22	bjetsM_transMass	: 1.964e-01
23	bjetsL_2pt	: 1.942e-01
24	jets_8pt	: 1.929e-01
25	jets_transMass	: 1.882e-01
26	bjetsT_invariantMass	: 1.788e-01
27	jets_number	: 1.758e-01
28	jets_9pt	: 1.752e-01
29	bjetsT_num	: 1.748e-01
30	jets_HT	: 1.738e-01
31	jets_5pt	: 1.693e-01
32	bjetsM_2pt	: 1.688e-01
33	bjetsT_HT	: 1.687e-01
34	jets_4pt	: 1.686e-01
35	bjetsL_3phi	: 1.679e-01
36	bjetsL_3eta	: 1.662e-01

37	bjetsT_3pt	: 1.609e-01
38	toptagger_2pt	: 1.598e-01
39	bjetsT_transMass	: 1.588e-01
40	bjetsT_3eta	: 1.578e-01
41	bjetsT_3phi	: 1.578e-01
42	Met_pt	: 1.538e-01
43	bjetsL_1pt	: 1.461e-01
44	jets_9eta	: 1.444e-01
45	jets_9phi	: 1.441e-01
46	bjetsM_1pt	: 1.393e-01
47	jets_3pt	: 1.393e-01
48	toptagger_minDeltaR_v1	: 1.380e-01
49	bjetsM_4pt	: 1.327e-01
50	bjetsM_4phi	: 1.321e-01
51	bjetsM_4eta	: 1.319e-01
52	toptagger_3pt	: 1.304e-01
53	jets_10pt	: 1.300e-01
54	nonbjetsM_num	: 1.294e-01
55	toptagger_num	: 1.291e-01
56	bjetsT_1pt	: 1.276e-01
57	bjetsT_2pt	: 1.270e-01
58	jets_10phi	: 1.267e-01
59	jets_10eta	: 1.265e-01
60	toptagger_3phi	: 1.218e-01
61	toptagger_3eta	: 1.207e-01
62	jets_MHT	: 1.200e-01
63	bjetsL_minDeltaR	: 1.174e-01
64	jets_invariantMass	: 1.160e-01
65	bjetsL_tausF_minDeltaR	: 1.113e-01
66	nonbjetsT_num	: 1.013e-01
67	jets_1pt	: 9.924e-02
68	jets_2pt	: 9.656e-02
69	bjetsM_minDeltaR	: 9.451e-02
70	tausL_HT	: 9.246e-02
71	tausL_1eta	: 8.874e-02
72	bjetsT_2phi	: 8.718e-02
73	tausF_HT	: 8.650e-02

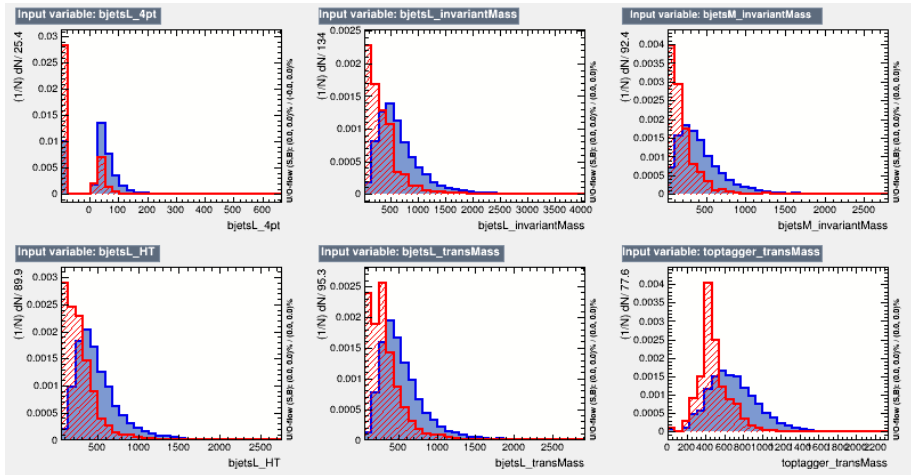
- how to take correlation into consideration when doing the ranking (mentioned in the SS of TTTT AN)

# Input Variable Distribution

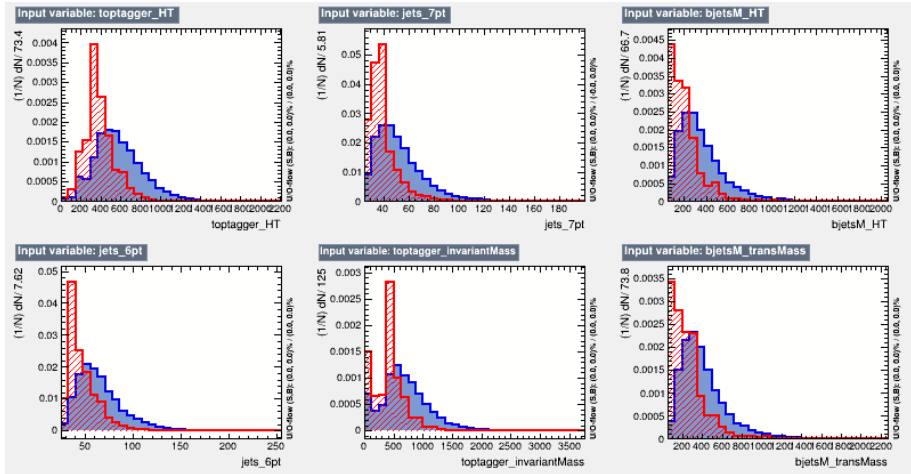




# Input Variable Distribution

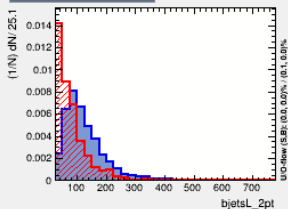


# Input Variable Distribution

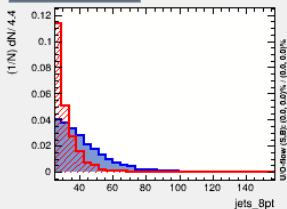


# Input Variable Distribution

Input variable: bjetsL\_2pt

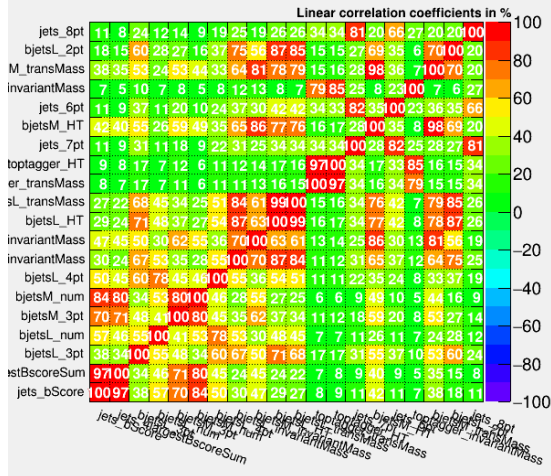


Input variable: jets\_8pt

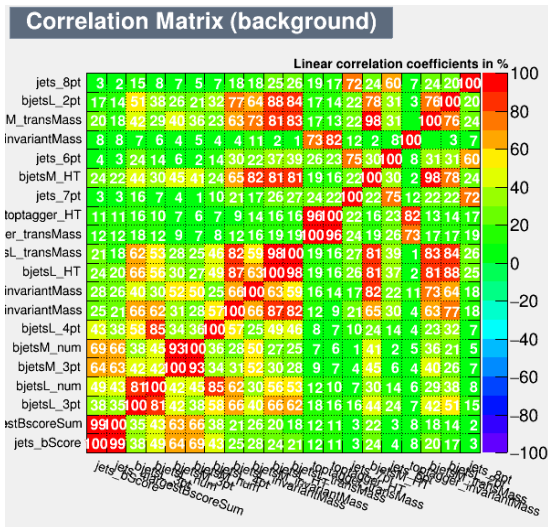


# Correlation

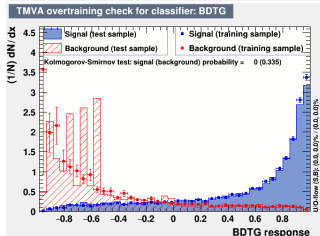
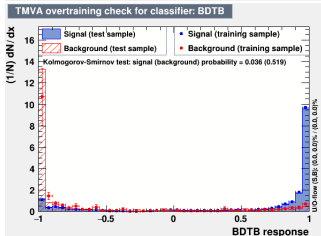
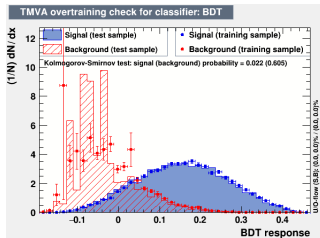
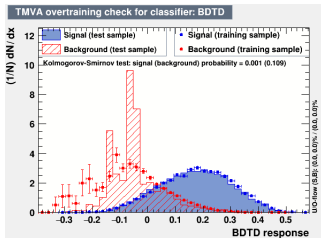
## Correlation Matrix (signal)



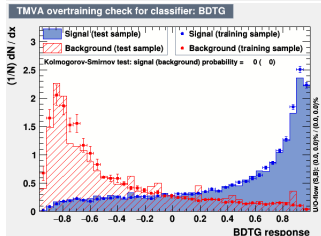
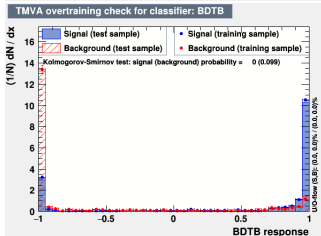
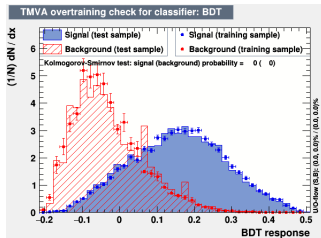
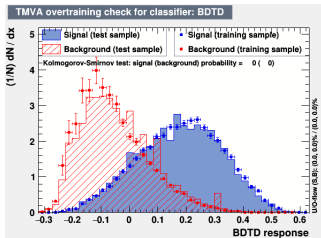
# Correlation



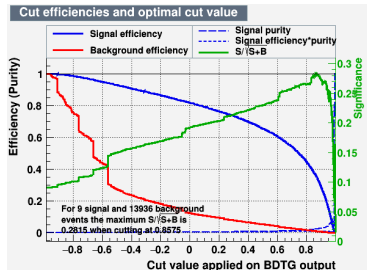
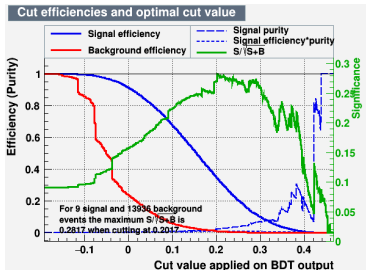
# OverTraining



# OverTraining



# Significance





# Outline

- 1 Problem from last time
- 2 Event Yield
- 3 1Tau0L
  - step 1
  - step 2

## Step 2

- Background:  $TT\text{To}2L2\text{Nu}$ ,  $TT\text{To}Hadronic$ ,  $TT\text{To}SemiLeptonic$ ,  $TTX(ttG, TTW, TTZ, TTH)$ ,  $QCD700\text{to}Inf$
- Here we contains only major background, because some low selection efficiency and big cross section can cause training to behave not as what we want

# Variables Separation Power Ranking

```

: Rank : Variable                               : Separation
:-----:-----:-----
: 1 : jets_bScore                               : 2.810e-01
: 2 : jets_4largestBscoreSum                   : 2.646e-01
: 3 : bjetsM_3pt                               : 2.448e-01
: 4 : bjetsM_num                               : 2.402e-01
: 5 : bjetsM_3eta                              : 2.205e-01
: 6 : bjetsM_3phi                              : 2.202e-01
: 7 : bjetsL_num                               : 1.987e-01
: 8 : bjetsL_4pt                               : 1.971e-01
: 9 : bjetsL_3pt                               : 1.957e-01
: 10 : bjetsM_invariantMass                     : 1.790e-01
: 11 : bjetsL_invariantMass                     : 1.744e-01
: 12 : toptagger_HT                             : 1.646e-01
: 13 : toptagger_transMass                       : 1.639e-01
: 14 : bjetsL_HT                                : 1.627e-01
: 15 : bjetsL_4eta                              : 1.587e-01
: 16 : bjetsL_4phi                              : 1.581e-01
: 17 : bjetsT_num                               : 1.534e-01
: 18 : jets_7pt                                 : 1.499e-01
: 19 : jets_8pt                                 : 1.455e-01
: 20 : bjetsT_3pt                               : 1.448e-01
: 21 : bjetsL_transMass                         : 1.426e-01
: 22 : bjetsT_3phi                              : 1.413e-01
: 23 : bjetsT_3eta                              : 1.412e-01
: 24 : bjetsM_HT                                : 1.387e-01
: 25 : toptagger_invariantMass                   : 1.382e-01
: 26 : jets_9pt                                 : 1.331e-01
: 27 : toptagger_num                             : 1.323e-01
: 28 : jets_6pt                                 : 1.306e-01
: 29 : jets_number                              : 1.296e-01
: 30 : bjetsT_invariantMass                     : 1.286e-01
: 31 : bjetsM_4pt                               : 1.212e-01
: 32 : bjetsL_2pt                               : 1.207e-01
: 33 : bjetsM_4phi                              : 1.202e-01

```

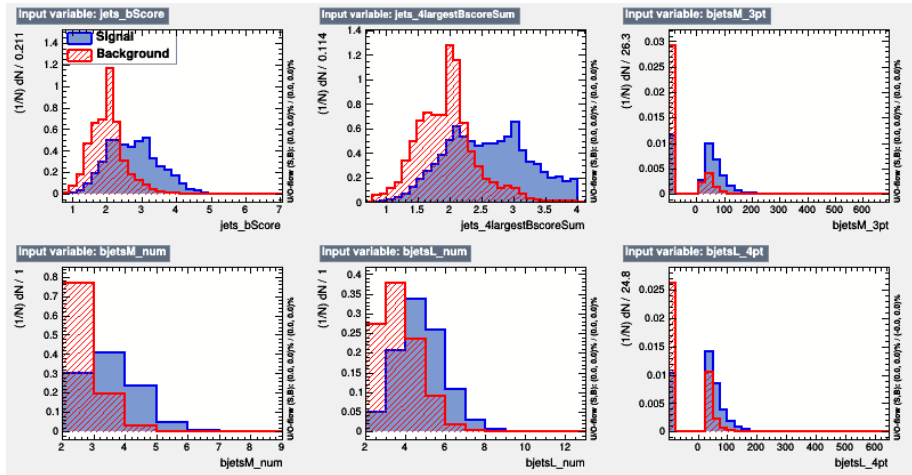
```

: 34 : bjetsM_4eta                              : 1.202e-01
: 35 : toptagger_3pt                            : 1.184e-01
: 36 : bjetsM_transMass                         : 1.179e-01
: 37 : toptagger_2pt                            : 1.152e-01
: 38 : toptagger_3phi                           : 1.127e-01
: 39 : toptagger_3eta                            : 1.127e-01
: 40 : bjetsT_HT                                : 1.113e-01
: 41 : bjetsM_2pt                               : 1.064e-01
: 42 : jets_5pt                                 : 1.010e-01
: 43 : jets_9eta                                : 1.007e-01
: 44 : bjetsT_transMass                         : 1.004e-01
: 45 : jets_9phi                                : 1.002e-01
: 46 : jets_10pt                                : 9.982e-02
: 47 : nonbjetsM_num                           : 9.922e-02
: 48 : jets_transMass                           : 9.559e-02
: 49 : jets_10eta                              : 9.379e-02
: 50 : jets_10phi                              : 9.377e-02
: 51 : bjetsT_3eta                              : 9.369e-02
: 52 : bjetsL_3phi                              : 9.261e-02
: 53 : jets_HT                                  : 8.820e-02
: 54 : Met_pt                                   : 8.740e-02
: 55 : toptagger_minDeltaR_v1                   : 8.671e-02
: 56 : bjetsT_2pt                               : 8.649e-02
: 57 : jets_MHT                                 : 7.888e-02
: 58 : jets_4pt                                 : 7.845e-02
: 59 : bjetsT_minDeltaR                         : 6.896e-02
: 60 : bjetsL_minDeltaR                         : 6.552e-02
: 61 : bjetsT_1pt                               : 6.073e-02
: 62 : bjetsL_1pt                               : 6.032e-02
: 63 : nonbjetsT_num                           : 5.954e-02
: 64 : bjetsM_1pt                               : 5.667e-02
: 65 : bjetsM_minDeltaR                         : 5.631e-02
: 66 : jets_11pt                                : 5.366e-02
: 67 : jets_3pt                                 : 5.349e-02

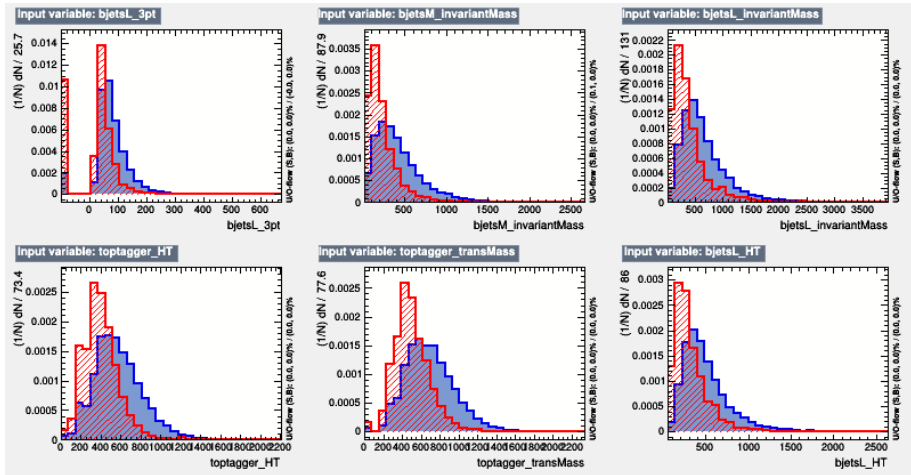
```

- how to take correlation into consideration when doing the ranking (mentioned in the SS of TTTT AN)

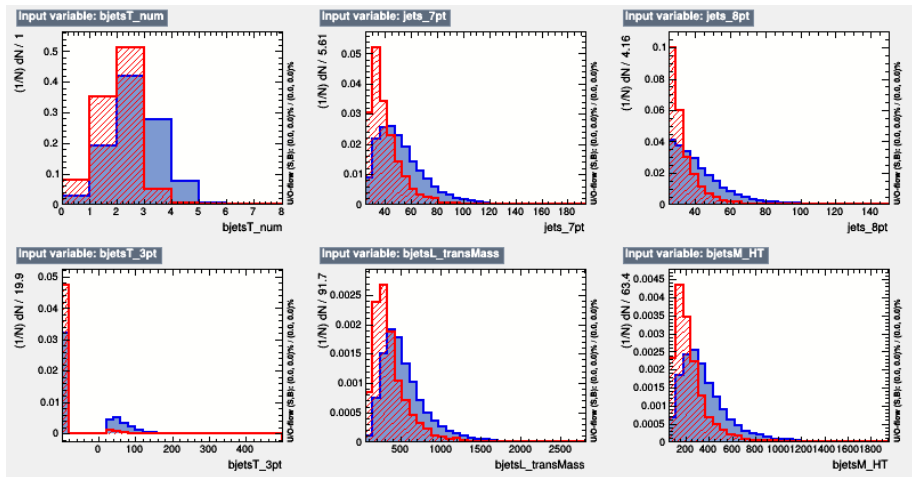
# Input Variable Distribution



# Input Variable Distribution

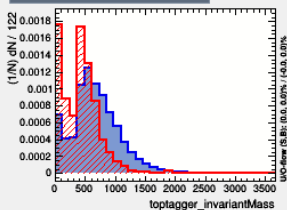


# Input Variable Distribution

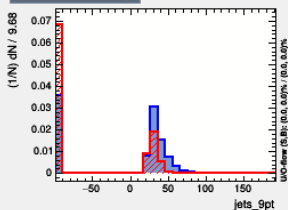


# Input Variable Distribution

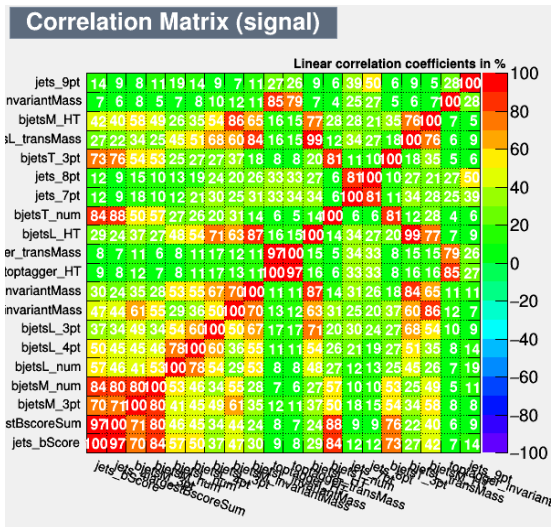
input variable: toptagger\_invariantMass



input variable: jets\_9pt



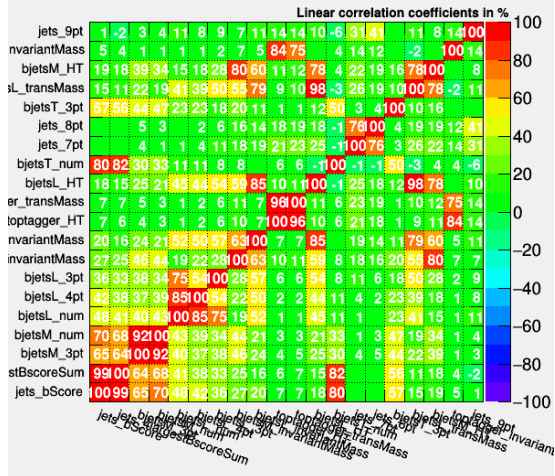
## Correlation



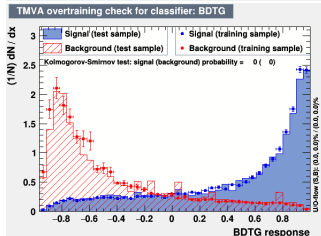
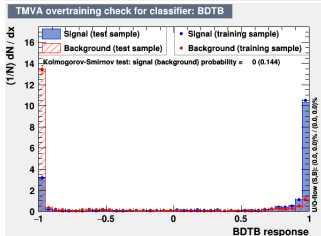
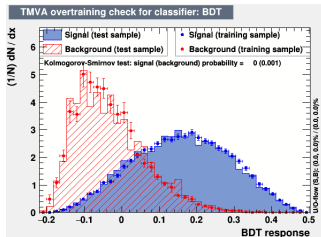
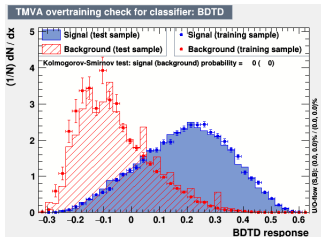


# Correlation

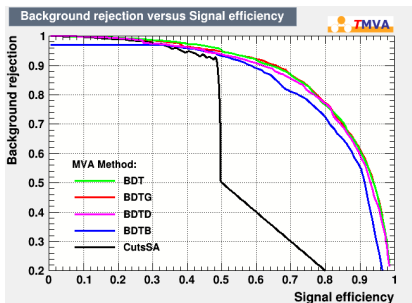
## Correlation Matrix (background)



# OverTraining

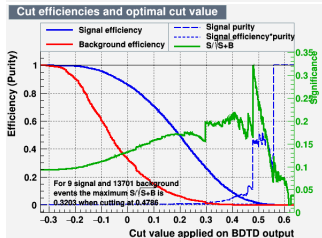
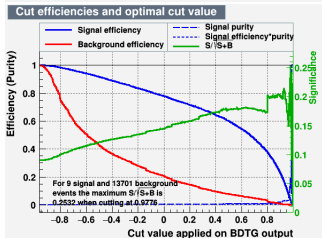
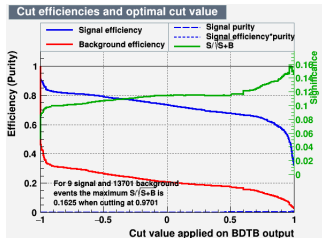
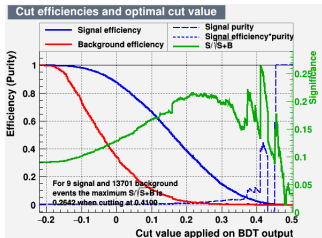


# ROC



DataSet	MVA	
Name:	Method:	ROC-integ
dataset	BDT	: 0.870
dataset	BDTG	: 0.870
dataset	BDTD	: 0.861
dataset	BDTB	: 0.836
dataset	CutsSA	: 0.607

# Significance



# questions

- why correlation matter for BDT input variables?
- how does per event weight and global event weight impact each MVA method?
- why keep small set of input variables?
  - Theoretically, with more variables, we have more information about the event and the BDT should gain greater separating power. However, in practice we want to reduce the number of variables used to train the BDT while retaining its performance as much as possible. This is mainly due to the fact that we use Monte Carlo simulated data to train these Classification BDTs. As Monte Carlo simulations are not always perfectly reliable for all variables, each variable used in the training needs to be checked for validity and bias towards different Monte Carlo generators. With less variables used, less human effort is needed on these checks and systematic error can be potentially reduced
  - Minimising the number of training variables also has the potential benefit of removing information redundancy and reducing training time. Moreover, by looking at which variables are ranked as the

# Questions

- in the training of BDT, for each node, what is the number of signal and bg events? raw entries or weighted(per event weight and global weight)?  $\delta I$  is sensitive to this, right?