

TMVA Study With Combine Results(V3)

Progress Report on Tau Final States of TTTT

Fabio Lemmi¹ Huiling Hua¹ Hongbo Liao¹ Hideki Okawa²
Yu Zhang²

¹IHEP

²Fudan University

Group Meeting, 2021

Outline

- 1 1Tau1L
 - 100 bins
 - 11 bins
- 2 1Tau2L
- 3 2TauXL
- 4 Subchannels Combination
- 5 1Tau0L

V3 training setup

- Compared to v2 training in 0726, we have added the b tag weight and HLT weight to all MC samples
- The v2 training are in slides with date 20210726
- So the MC corrections we have considered so far are
 - prefiring weight
 - PU weight
 - gen weight
 - B tag efficiency weight
 - HLT efficiency weight

Outline

- 1 1Tau1L
 - 100 bins
 - 11 bins
- 2 1Tau2L
- 3 2TauXL
- 4 Subchannels Combination
- 5 1Tau0L

Input variable sets

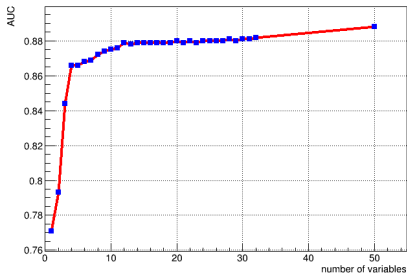
	1tau1l	
1:	jets_bScore	0.2525
2:	jets_largestBScoreSum	0.2377
3:	bjetsL_num	0.214
4:	bjetsM_num	0.2126
5:	bjetsM_3pt	0.212
6:	bjetsL_3pt	0.1963
7:	bjetsL_4pt	0.1817
8:	jets_7pt	0.176
9:	jets_number	0.1739
10:	toptagger_HT	0.1716
11:	toptagger_transMass	0.1654
12:	bjetsL_HT	0.1532
13:	bjetsL_invariantMass	0.1454
14:	jets_6pt	0.1407
15:	bjetsL_transMass	0.1397
16:	toptagger_invariantMass	0.139
17:	toptagger_num	0.1386
18:	jets_6pt	0.1382
19:	toptagger_2pt	0.1332
20:	bjetsM_invariantMass	0.1279
21:	bjetsL_num	0.1249
22:	toptagger_minDeltaR_v1	0.1232
23:	bjetsL_3pt	0.1213
24:	bjetsM_HT	0.1141
25:	jets_5pt	0.1138
26:	jets_rationHT_4toRest	0.1125
27:	jets_transMass	0.104
28:	bjetsM_transMass	0.09949
29:	nonbjetsM_num	0.09857
30:	jets_HT	0.09757
31:	bjetsL_minDeltaR	0.09183
32:	bjetsL_2pt	0.0878
33:	jets_4pt	0.08673
34:	bjetsL_invariantMass	0.0844
35:	bjetsM_4pt	0.08415
36:	jets_9pt	0.07987
37:	bjetsL_HT	0.07716
38:	nonbjetsL_num	0.0752
39:	nonbjetsM_4pt	0.07057
40:	bjetsL_transMass	0.06764
41:	bjetsM_2pt	0.06291
42:	nonbjetsL_4pt	0.06213
43:	bjetsM_minDeltaR	0.06138
44:	jets_3pt	0.0604
45:	toptagger_3pt	0.05585
46:	bjetsL_2pt	0.0520
47:	toptagger_MHT	0.05139
48:	bjetsL_minDeltaR	0.04791
49:	bjetsL_1pt	0.04692
50:	leptonsMVAL_number	0.04156

	1tau1l	
1:	jets_bScore	0.2525
2:	bjetsM_3pt	0.212
3:	jets_7pt	0.176
4:	jets_number	0.1739
5:	toptagger_HT	0.1716
6:	jets_6pt	0.1407
7:	bjetsM_invariantMass	0.1279
8:	jets_rationHT_4toRest	0.1125
9:	jets_transMass	0.104
10:	bjetsM_4pt	0.08415
11:	nonbjetsM_4pt	0.07057
12:	bjetsM_2pt	0.06291
13:	bjetsM_minDeltaR	0.06138
14:	toptagger_3pt	0.05585
15:	toptagger_MHT	0.05139
16:	leptonsMVAL_number	0.04156

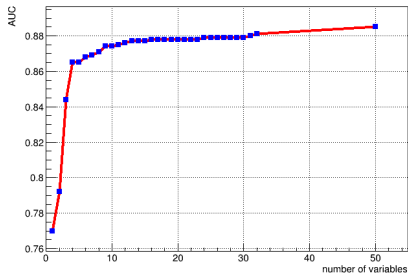
	1tau1l	
1:	jets_bScore	0.2525
2:	jets_7pt	0.176
3:	toptagger_HT	0.1716
4:	jets_6pt	0.1407
5:	bjetsM_invariantMass	0.1279
6:	jets_transMass	0.104
7:	nonbjetsM_4pt	0.07057
8:	bjetsM_minDeltaR	0.06138
9:	toptagger_3pt	0.05585
10:	toptagger_MHT	0.05139
11:	leptonsMVAL_number	0.04156

AUC results

AUC vs No. of Variables (BDT)

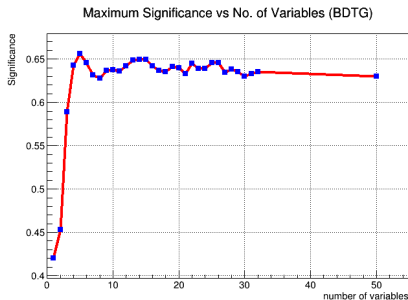
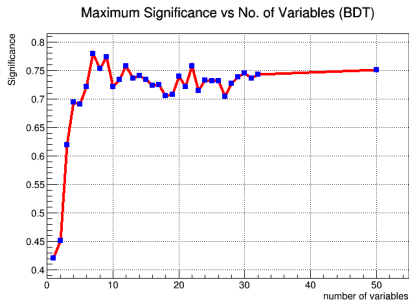


AUC vs No. of Variables (BDTG)



- AUC results similar with v2 training

Significance results

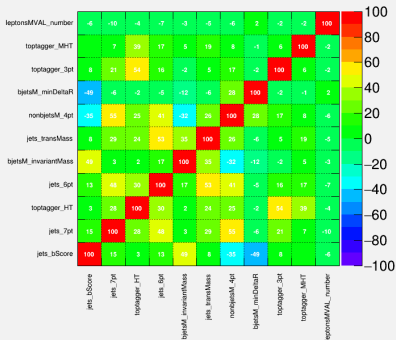


- similar with v_2 results
- because we lack statistic in the right end region of BDT score, to avoid fluctuation simply only consider the 1-30(40 in total) bin for BDT and 1-35 bin for BDTG

11 input variables set

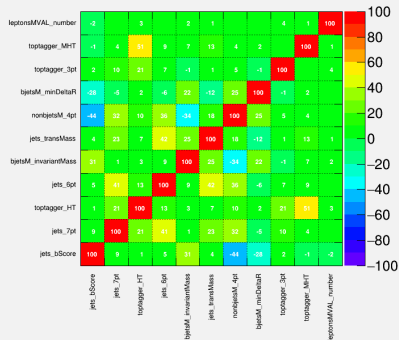
Correlation Matrix (signal)

(11 input variables)



Correlation Matrix (background)

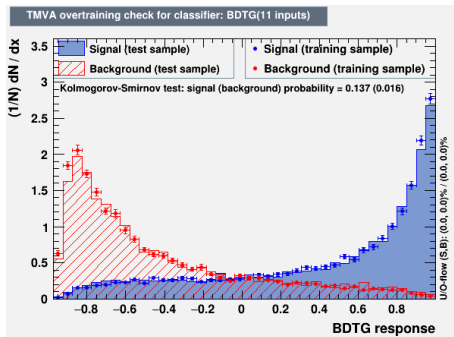
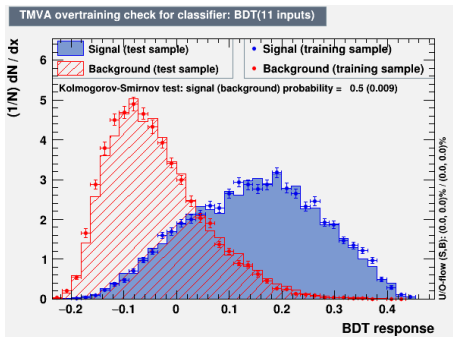
(11 input variables)



- in parentheses below the title is the number of input variable for training
- this is the correlation matrix for training with 11 input variables



11 input variables set



- number in the parenthese after the title indicates the number of input variables
- so this is the overtraining check plot for 11 input variables(which is the middle list in page 7)

Datacard for separate backgrounds

```

jhuang@lxslc704:publicfs/cms/user/huahuil/TauOfTTTT/2016v1/TMVAOutput/v46_v3addBtagHLTweights/1tau1_v1/AppResults/datacard/seperateDC
shaper ** /publicfs/cms/user/huahuil/TauOfTTTT/2016v1/TMVAOutput/v46_v3addBtagHLTweights/1tau1_v1/AppResults/TMVAApp_1tau1_11var_forCombine.root $PROCE
SS_MVA_BDT
bin SR_1tau1l
observation -1
-----
bin SR_1tau1l SR_1tau1l SR_1tau1l SR_1tau1l SR_1tau1l SR_1tau1l SR_1tau1l SR_1tau1l SR_1tau1l SR_1tau1l SR_1tau1l SR_1tau1l SR_1tau1l SR_1tau1l SR_1tau1l SR_1tau1l SR_1tau1l SR_1tau1l SR_1tau1l
SR_1tau1l SR_1tau1l SR_1tau1l SR_1tau1l SR_1tau1l SR_1tau1l SR_1tau1l SR_1tau1l SR_1tau1l SR_1tau1l SR_1tau1l SR_1tau1l SR_1tau1l SR_1tau1l SR_1tau1l SR_1tau1l SR_1tau1l SR_1tau1l SR_1tau1l
process TTTT TGJets TTTo2L2Nu TTToHadronic TTToSemiLeptonic TTGJets ttZJets ttWJets WZG ttH ttZq_ll
W ST_tW_antitop WGJets ST_tW_top TGJets ToLLG WW THW WZ THQ QCD_HT150to2000
process 0 9 18 10 2 11 3 12 4 13 5 14 6 15 7 16 8
rate 17 -1 18 -1 19 -1 20 -1 21 -1 22 -1 -1 -1 -1 -1 -1 -1 -1
1 -1 -1 -1 -1 -1 -1 -1 -1 -1 -1 -1 -1 -1 -1
SR_1tau1l autoMCSStats 0

```

Datacard for summed backgrounds

```
imax *
jmax *
kmax *

shapes ** /publicfs/cms/user/huahuil/Tau0FTTTT/2016v1/TMVA0Output/v46_v3addBtagHLTweights/1tau1_v1/AppResults/TMVAApp_1tau1_l1var_forCombine.root $PROCE
SS_MVA_BDT
bin SR_1tau1l
observation -1

bin SR_1tau1l SR_1tau1l SR_1tau1l SR_1tau1l SR_1tau1l SR_1tau1l SR_1tau1l SR_1tau1l
process TTTT TT TTX VV WV SingleTop TX QCD
process 0 1 2 3 4 5 6 7
rate -1 -1 -1 -1 -1 -1 -1 -1

SR_1tau1l autoMCStats 0
```

- sum the the histograms of seperate backgrounds
- feed the summed templates to combine

Expected significance of separate and summed templates

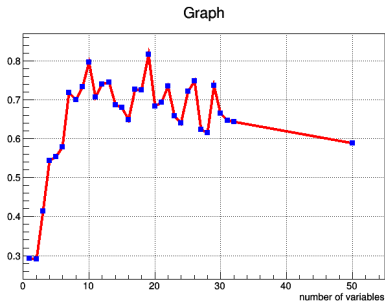


Figure: separate

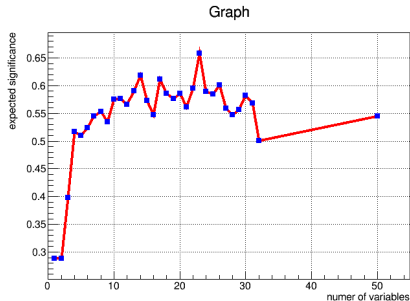


Figure: summed

- it seems the summed significance is worse than the separate
- I could not understand the reason for the difference for now
- need to check my code to make sure the difference is not stem from bugs
- must study the algorithm under the hood of the combine deeper to understand the difference

Expected limit for separate and summed templates

Graph

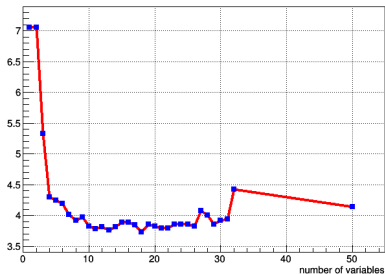


Figure: separate

Graph

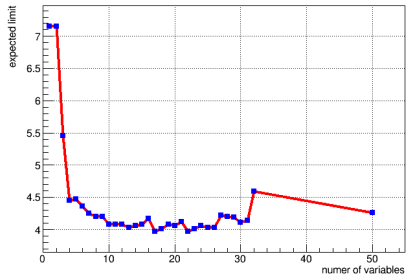


Figure: summed

- expected limit is also worse for summed templates

Combine results with different binning

- Use the same training weight file
- Rerun the application code to set the fill of BDT score histograms with 11 bins rather than 100
- Everything else the same

Expected significance of 100 bins and 11 bins templates(summed bg)

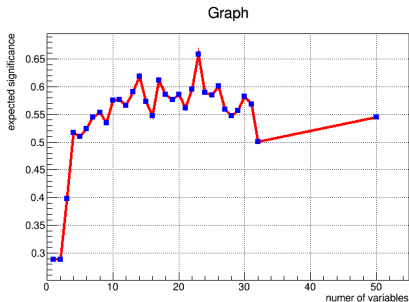


Figure: 100 bins

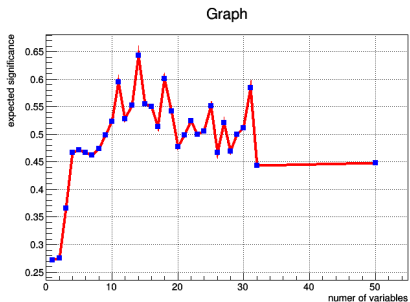


Figure: 11 bins

- why for 11 bins it seems fluctuation is bigger?

Expected limit for 100bins and 11 bins templates(summed bg)

Graph

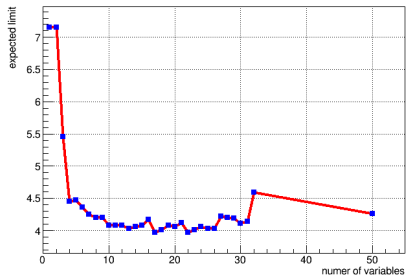


Figure: 100 bins

Graph

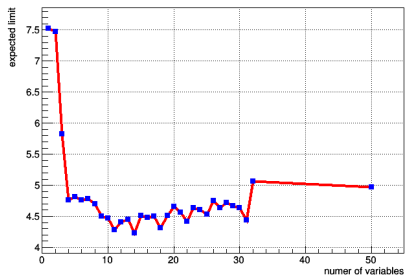


Figure: 11 bins

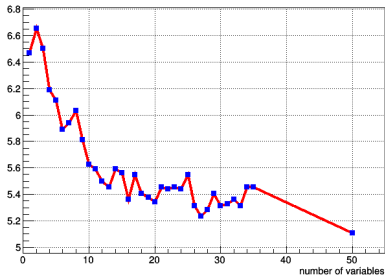
- still it fluctuates more for 11 bins

Outline

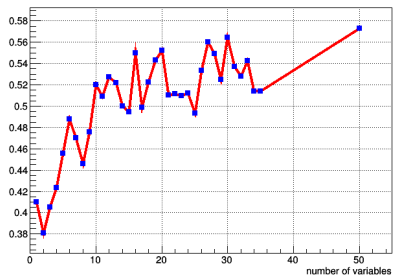
- 1 1Tau1L
 - 100 bins
 - 11 bins
- 2 1Tau2L
- 3 2TauXL
- 4 Subchannels Combination
- 5 1Tau0L

Results of summed 100 bins

Graph



Graph

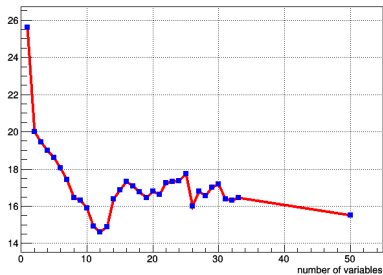


Outline

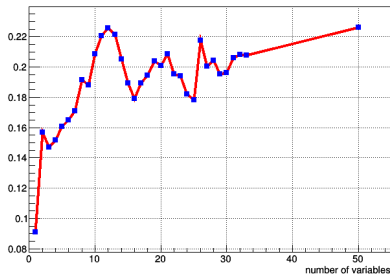
- 1 1Tau1L
 - 100 bins
 - 11 bins
- 2 1Tau2L
- 3 2TauXL**
- 4 Subchannels Combination
- 5 1Tau0L

Combine results of summed 100bins templates

Graph



Graph



Outline

- 1 1Tau1L
 - 100 bins
 - 11 bins
- 2 1Tau2L
- 3 2TauXL
- 4 Subchannels Combination**
- 5 1Tau0L

Combination of subchannels

- Tried to combine 1tau1l and 1tau2l and 2tauXl together
- summed histograms, 100 bins
- expected significance: 0.792603 , expected limit 2.9219

```

lmax 3 number of bins
lmax 7 number of processes minus 1
lmax 0 number of nuisance parameters
-----
shapes * ch1 /publicfs/cms/user/luhuhul/Tau0FTTTT/2016v1/TMVAOutput/v46_v3addBtagHLTweights/1tau2l_v1/AppResults/TMVAApp_1tau2l_15var_forCombine.root $P
PROCESS_MVA_BDT
shapes * ch2 /publicfs/cms/user/luhuhul/Tau0FTTTT/2016v1/TMVAOutput/v46_v3addBtagHLTweights/1tau1l_v1/AppResults/TMVAApp_1tau1l_11var_forCombine.root $P
PROCESS_MVA_BDT
shapes * ch3 /publicfs/cms/user/luhuhul/Tau0FTTTT/2016v1/TMVAOutput/v46_v3addBtagHLTweights/2tauXl_v1/AppResults/TMVAApp_2tauXl_12var_forCombine.root $P
PROCESS_MVA_BDT
-----
bin
observation ch1 ch2 ch3
             -1 -1 -1
-----
bin          ch1  ch1  ch1  ch1  ch1  ch1  ch2  ch2  ch2  ch2  ch2  ch2
ch2          ch3  ch3  ch3  ch3  ch3  ch3  ch3  ch2  ch2  ch2  ch2
process      TTTT  TTX  TX  VVV  TT  SingleTop  TTTT  TTX  TX  VVV  TT  VV  QCD
SingleTop    TTTT  TTX  TX  VVV  TT  VV  SingleTop  TTTT  TTX  TX  VVV  TT  VV  QCD
process      0    1    2    3    4    5    0    1    2    3    4    6    7
5            0    1    2    3    4    6    0    1    2    3    4    6    7
rate         -1   -1   -1   -1   -1   -1   -1   -1   -1   -1   -1   -1   -1
-1           -1   -1   -1   -1   -1   -1   -1   -1   -1   -1   -1   -1
-----
ch1 autoMCStats 0 0 1
ch2 autoMCStats 0 0 1
ch3 autoMCStats 0 0 1

```

Outline

- 1 1Tau1L
 - 100 bins
 - 11 bins
- 2 1Tau2L
- 3 2TauXL
- 4 Subchannels Combination
- 5 1Tau0L

1tau0l

- Corrected HT of QCD from Fabio(11 bins)

Backup

back up