

TMVA Study With Combine Results with Nano Part1 (v3)

Progress Report on Tau Final States of TTTT

Duncan Leggat²

Fabio Lemmi¹

Huiling Hua¹

Hongbo Liao¹

Hideki Okawa²

Yu Zhang²

¹IHEP

²Fudan University

Weekly Meeting

Outline

- 1 1Tau1L
 - 1tau1l

Bugs for previous spiky and absurdly high expected significance

- For high value significance, forgot to **change the channel selection** in the training and application code
 - ▶ With highly automated code comes with hidden detail.
 - ▶ I have updated my auto workflow code to output log for critical information
- Switched the BDT range to **[-0.2, 0.4]** with 11 bins
 - ▶ For BDT signal and bg range are actually within range of [-0.2, 0.4]
 - ▶ Previously I put the default range of [-0.8, 0.8] with 11 bins. this is causing too few bins in the sensitive range which means we are losing information

New Channel Definition

- Preselection: $HT > 500$ GeV; 6th jet pt > 40 GeV; jets number ≥ 6
- Tighter selection for each channels

channel	tight tau	tight lepton	jet	medium b jet
1tau0l	1	0	8	2
1tau1l	1	1	7	2
1tau2l	1	2	6	2
1tau3l	1	3	6	2
2tau0l	2	0	7	2
2tau1l	2	1	6	2
2tau2l	2	2	4	2

Table: Subchannel definition

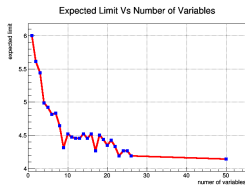
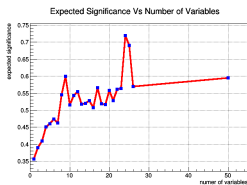
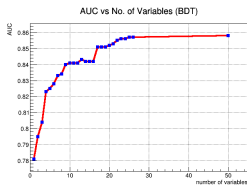
Recap of MVA optimization strategy

- Correlation removal method
 - ▶ Start with 50 most powerful variables
 - ▶ Generate 50 variable list by remove one variable at a time according to Correlation
 - ▶ Run BDT training with the 50 lists for 50 times
 - ▶ For each BDT training calculate AUC and expected significance and expected limit
- Use Combine fitting results as indicator
 - ▶ After each training we do application to get the BDT distribution
 - ▶ Feed the distribution to Combine to get **expected significance and limit**

Outline

- 1 1Tau1L
 - 1tau1l

Training performance Vs number of input variables

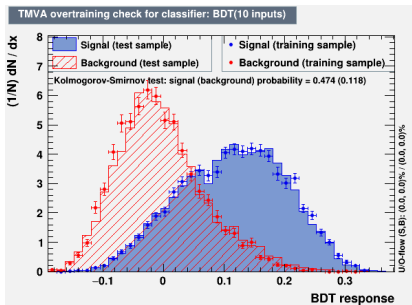


- Reached plateau of **0.55** significance at around **10 input variables**
- Reaching plateaus too soon, we need more good input variables to enhance performance. for example SUSY Tops
- we have 2 spiky points in significance plot, not pleasing but possible

Final Input Variable Set

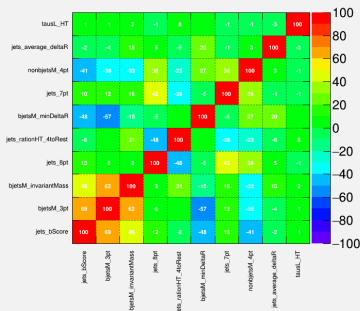
1tau1l	
1: jets_bScore	0.2421
2: bjetsM_3pt	0.2254
3: bjetsM_invariantMass	0.1215
4: jets_8pt	0.08294
5: jets_rationHT_4toRest	0.06637
6: bjetsM_minDeltaR	0.06056
7: jets_7pt	0.06053
8: nonbjetsM_4pt	0.0416
9: jets_average_deltaR	0.03739
10: tausL_HT	0.03368

10 input variables training performance

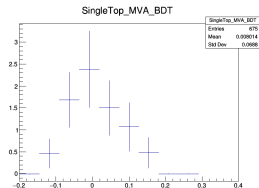
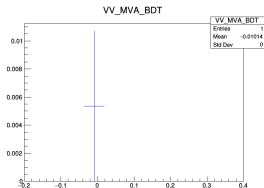
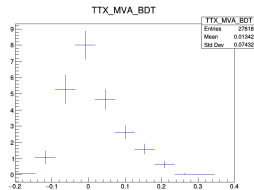
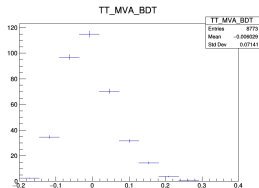
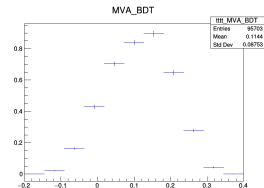


Correlation Matrix (signal)

(10 input variables)



Templates for Combine(10 inputs)



Datacard(10 inputs)

```
imax *
jmax *
kmax *
-----
shapes * * /publicfs/cms/user/huahuil/tauOfTTTT_NanoAOD/TMVAoutput/2016/v2Add2Variables_fromV9/1t
au1l_v2/AppResults_changeBDTrange_11bins/TMVAApp_1tau1l_10var_forCombine.root $PROCESS_MVA_BDT
bin          SR_1tau1l
observation  -1
-----
bin          SR_1tau1l          SR_1tau1l          SR_1tau1l          SR_1tau1l          SR_1tau1l
process      tttt              TT              TTX              VV              SingleTop
process      0                1                2                3                4
rate         -1                -1                -1                -1                -1
-----
SR_1tau1l   autoMCStats  10
```

- Excluded empty process QCD in datacard because combine won't accept empty histograms

Backup

back up

EventSelection

- MET fillters
- HLT requirements
 - ▶ `HLT_PFHT450_SixJet40_BTagCSV_p056==1,`
`HLT_PFHT400_SixJet30_DoubleBTagCSV_p056==1`
 - ▶ `HLT_PFJet450`
- loose preselection
 - ▶ `tausL.size()>0, jets.size()>=6, bjetsL.size()>1`
 - ▶ `HT>500`
- Subchannel requirements

MC reweighting

- genWeight
- prefireWeight
- PileUp reweighting

Definition of Backgrounds

Event Yield

2016: 1tau1e

raw entries:

TTTT = 41174.000000
 TT = 3894.000000
 TTX = 12788.000000
 VV = 1.000000
 SingleTop = 405.000000
 QCD = 0.000000
 allBg = 17088.000000

2016: 1tau1m

raw entries:

TTTT = 54529.000000
 TT = 4879.000000
 TTX = 15030.000000
 VV = 0.000000
 SingleTop = 270.000000
 QCD = 0.000000
 allBg = 20179.000000

2016: 1tau1l

raw entries:

TTTT = 95703.000000
 TT = 8773.000000
 TTX = 27818.000000
 VV = 1.000000
 SingleTop = 675.000000
 QCD = 0.000000
 allBg = 37267.000000

2016: 1tau1e

scaled to LUMI:

TTTT = 1.752807
 TT = 163.272679
 TTX = 11.033890
 VV = 0.005344
 SingleTop = 5.555625
 QCD = 0.000000
 allBg = 179.867538

2016: 1tau1m

scaled to LUMI:

TTTT = 2.268438
 TT = 206.278997
 TTX = 12.897172
 VV = 0.000000
 SingleTop = 2.006958
 QCD = 0.000000
 allBg = 221.183126

2016: 1tau1l

scaled to LUMI:

TTTT = 4.021245
 TT = 369.551676
 TTX = 23.931062
 VV = 0.005344
 SingleTop = 7.562583
 QCD = 0.000000
 allBg = 401.050665

Event Yield

2016: 1tau0l

raw entries:

TTTT = 183517.000000

TT = 76684.000000

TTX = 125493.000000

VV = 7.000000

SingleTop = 1552.000000

QCD = 331.000000

allBg = 204067.000000

2016: 2tau0l

raw entries:

TTTT = 6259.000000

TT = 747.000000

TTX = 4819.000000

VV = 0.000000

SingleTop = 257.000000

QCD = 3.000000

allBg = 5826.000000

2016: 1tau0l

scaled to LUMI:

TTTT = 8.021262

TT = 4045.997619

TTX = 137.165365

VV = 0.075979

SingleTop = 90.983223

QCD = 2044.378702

allBg = 6318.600888

2016: 2tau0l

scaled to LUMI:

TTTT = 0.275864

TT = 35.554913

TTX = 4.286863

VV = 0.000000

SingleTop = 0.699728

QCD = 9.271531

allBg = 49.813034

Input lists

1tau1l	
1: jets_bScore	0.2421
2: bjetsM_3pt	0.2254
3: bjetsM_invariantMass	0.1215
4: jets_8pt	0.08294
5: jets_ratioHT_4toRest	0.06637
6: bjetsM_minDeltaR	0.06056
7: jets_7pt	0.06053
8: nonbjetsM_4pt	0.0416
9: jets_average_deltaR	0.03739
10: tausL_HT	0.03368

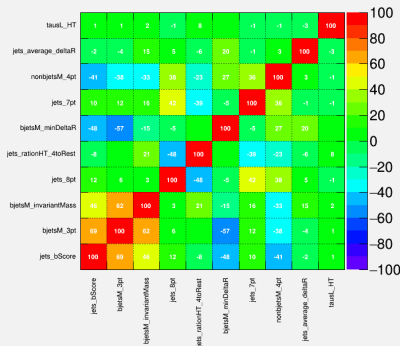
1tau1l	
1: jets_bScore	0.2421
2: bjetsM_3pt	0.2254
3: bjetsM_invariantMass	0.1215
4: jets_8pt	0.08294
5: jets_ratioHT_4toRest	0.06637
6: bjetsM_minDeltaR	0.06056
7: jets_7pt	0.06053
8: nonbjetsM_4pt	0.0416
9: jets_6pt	0.04093
10: jets_average_deltaR	0.03739
11: tausL_HT	0.03368

1tau1l	
1: jets_bScore	0.2421
2: bjetsM_3pt	0.2254
3: bjetsM_invariantMass	0.1215
4: jets_8pt	0.08294
5: jets_ratioHT_4toRest	0.06637
6: bjetsM_minDeltaR	0.06056
7: jets_7pt	0.06053
8: bjetsM_2pt	0.05418
9: nonbjetsM_4pt	0.0416
10: jets_6pt	0.04093
11: jets_average_deltaR	0.03739
12: tausL_HT	0.03368

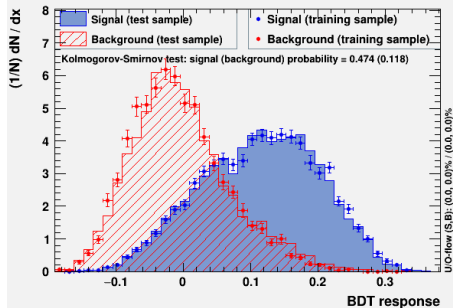
1tau1l	
1: jets_bScore	0.2421
2: bjetsM_3pt	0.2254
3: bjetsM_invariantMass	0.1215
4: jets_8pt	0.08294
5: jets_number	0.07828
6: jets_ratioHT_4toRest	0.06637
7: bjetsM_minDeltaR	0.06056
8: jets_7pt	0.06053
9: bjetsM_2pt	0.05418
10: nonbjetsM_4pt	0.0416
11: jets_6pt	0.04093
12: jets_average_deltaR	0.03739
13: tausL_HT	0.03368

Correlation Matrix (signal)

(10 input variables)

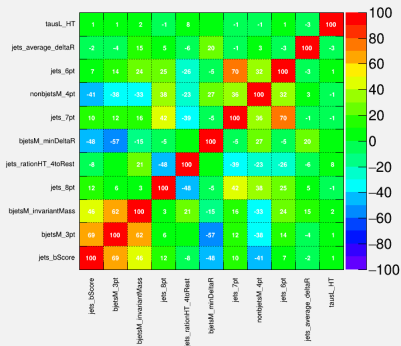


TMVA overtraining check for classifier: BDT(10 inputs)

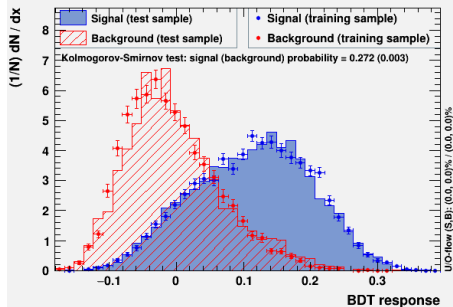


Correlation Matrix (signal)

(11 input variables)

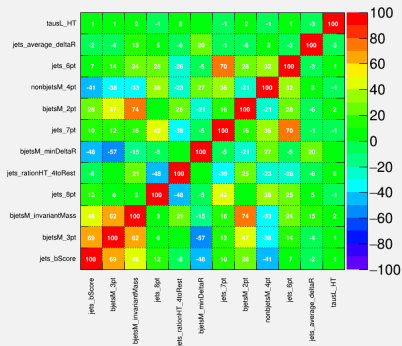


TMVA overtraining check for classifier: BDT(11 inputs)

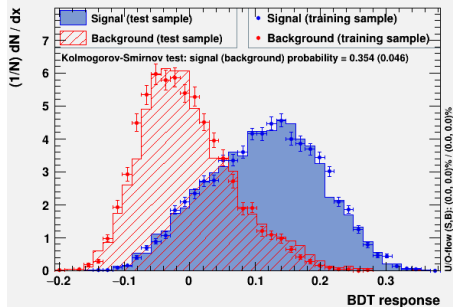


Correlation Matrix (signal)

(12 input variables)

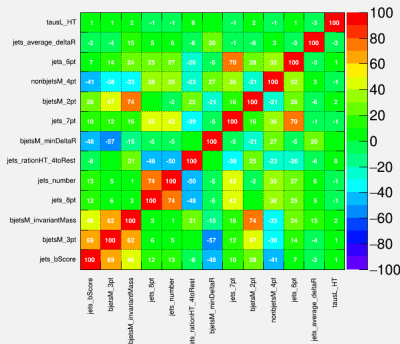


TMVA overtraining check for classifier: BDT(12 inputs)

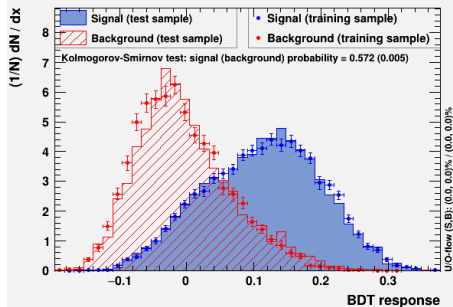


Correlation Matrix (signal)

(13 input variables)

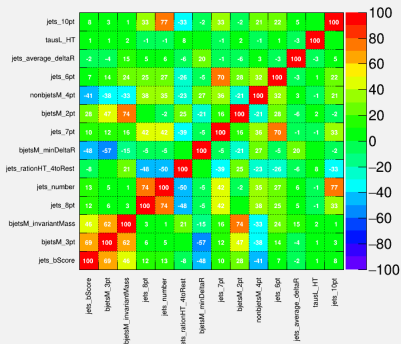


TMVA overtraining check for classifier: BDT(13 inputs)

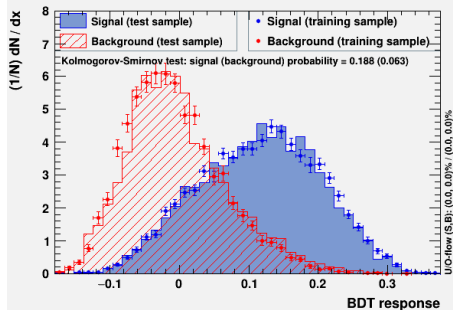


Correlation Matrix (signal)

(14 input variables)

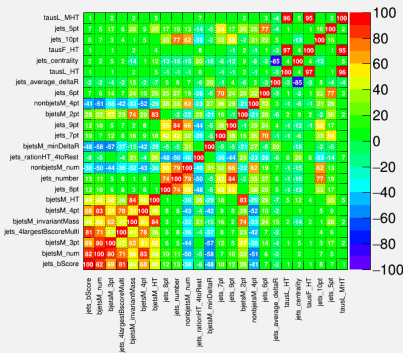


TMVA overtraining check for classifier: BDT(14 inputs)

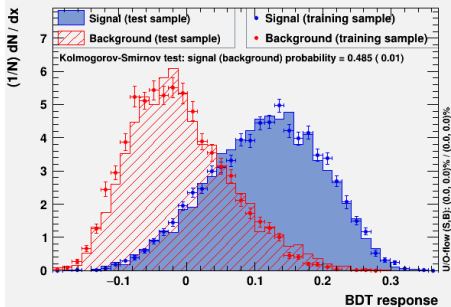


Correlation Matrix (signal)

(24 input variables)

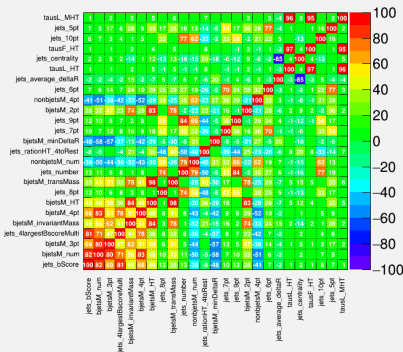


TMVA overtraining check for classifier: BDT(24 inputs)

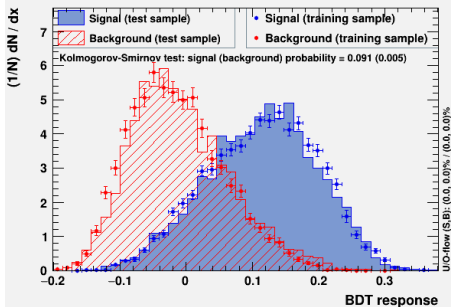


Correlation Matrix (signal)

(25 input variables)

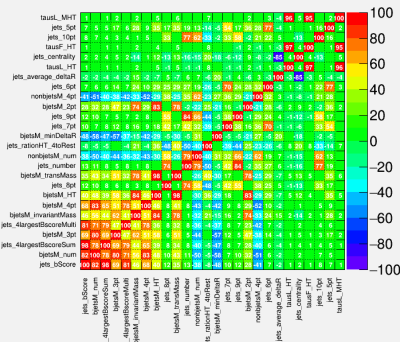


TMVA overtraining check for classifier: BDT(25 inputs)



Correlation Matrix (signal)

(26 input variables)



TMVA overtraining check for classifier: BDT(26 inputs)

