

TMVA Study With Combine Results with Nano Part1 (v3)

Progress Report on Tau Final States of TTTT

Duncan Leggat² Fabio Lemmi¹ **Huiling Hua**¹ Hongbo Liao¹
Hideki Okawa² Yu Zhang²

¹IHEP

²Fudan University

Weekly Meeting

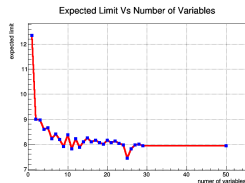
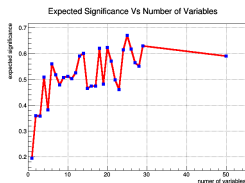
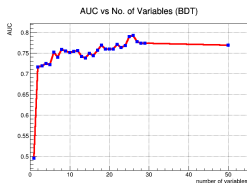
Outline

- 1 $1\tau 2L$ (BDT)
- 2 $2\tau 0l$ (BDT)
- 3 $2\tau 1l$ (HT)
- 4 $1\tau 0l$ (HT)
- 5 Combination of channels

Outline

- 1 1Tau2L(BDT)
- 2 2tau0l(BDT)
- 3 2tau1l(HT)
- 4 1tau0l(HT)
- 5 Combination of channels

Training performance Vs number of input variables

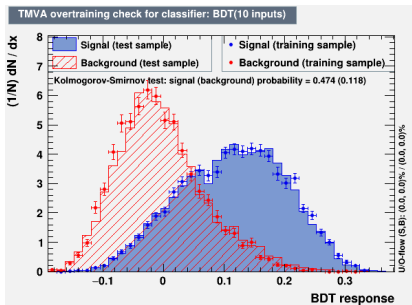


- Reached plateau of **0.5** significance at around **8 input variables**
- Reaching plateaus too soon, we need more good input variables to enhance performance. for example SUSY Tops

Final Input Variable Set

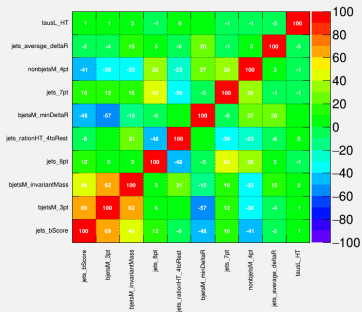
1tau2l	
1: jets_bScoreMultiply	1.366
2: jets_4largestBscoreSum	0.349
3: tausF_leptonsT_invariantMass0.1242	
4: tausF_leptonsTMVA_minDeltaR 0.1237	
5: nonbjetsM_1pt	0.1123
6: jets_centrality	0.1099
7: jets_5pt	0.09966
8: jets_tausF_minDeltaR	0.08138

10 input variables training performance

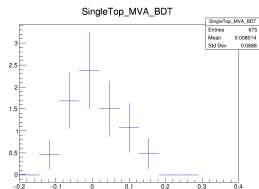
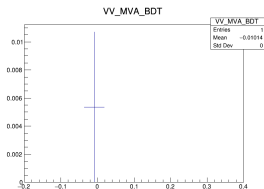
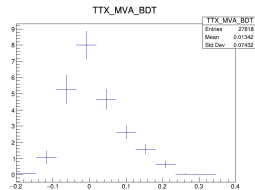
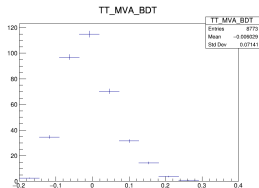
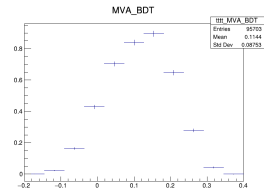


Correlation Matrix (signal)

(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

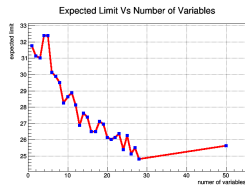
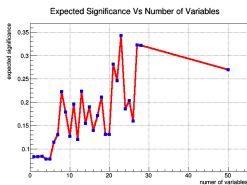
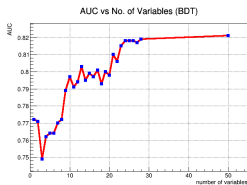
Outline

- 1 1Tau2L(BDT)
- 2 2tau0l(BDT)**
- 3 2tau1l(HT)
- 4 1tau0l(HT)
- 5 Combination of channels

2tau0l

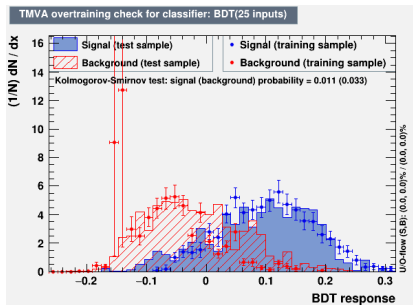
- it seems with new channel characterization BDT training in 2tau0l is possible
- But we have less than 1 signal events, how to interpret this?

Training performance Vs number of input variables



- Reached plateau of **0.25** significance at around **25 input variables**
- Reaching plateaus too soon, we need more good input variables to enhance performance. for example SUSY Tops
- it seems from 10 variables to 20 we are not gaining performance. maybe we need a better removal method like iterative method

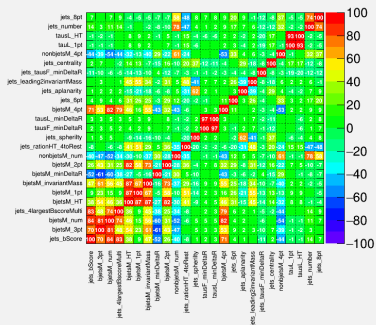
25 input variables training performance



- Spikes need to be understood

Correlation Matrix (signal)

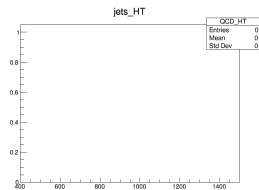
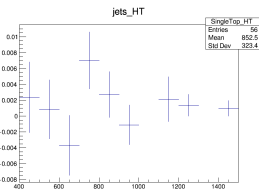
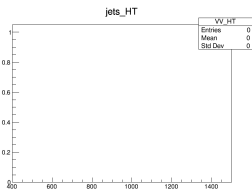
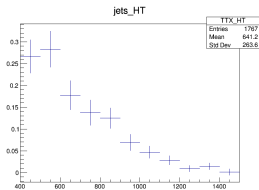
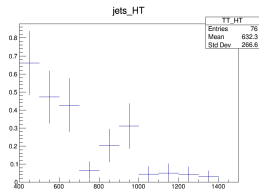
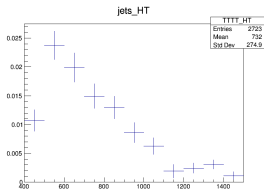
(25 input variables)



Outline

- 1 1Tau2L(BDT)
- 2 2tau0l(BDT)
- 3 2tau1l(HT)**
- 4 1tau0l(HT)
- 5 Combination of channels

2tau1l



- HT all come from MC

Fitting results of 2tau1l

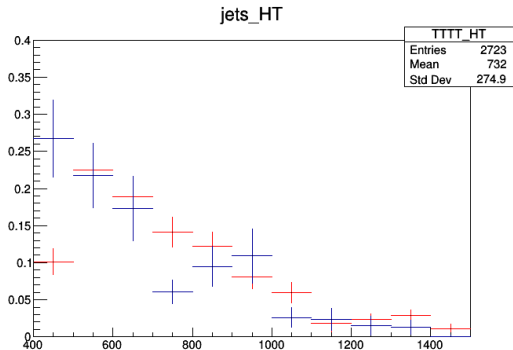


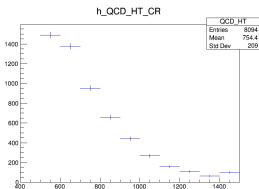
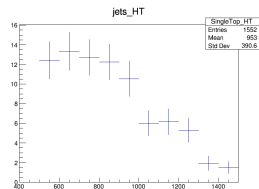
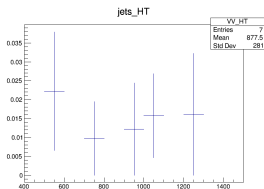
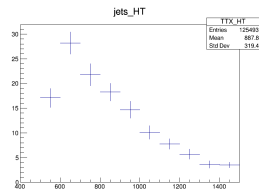
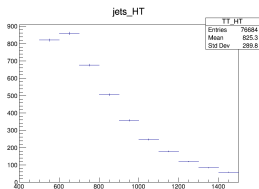
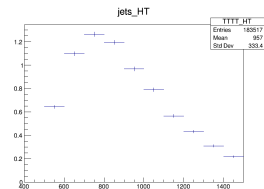
Figure: HT shape of TTTT and bg

- expected significance 0.0660922
- expected limit 45.2500
- need better variable for the fitting

Outline

- 1 1Tau2L(BDT)
- 2 2tau0l(BDT)
- 3 2tau1l(HT)
- 4 1tau0l(HT)**
- 5 Combination of channels

1tau0l HT distribution



- 1tau0l use HT as templates for combine
- HT of QCD from data driven method
- HT of other processes from MC

Fitting results of 1tau0l

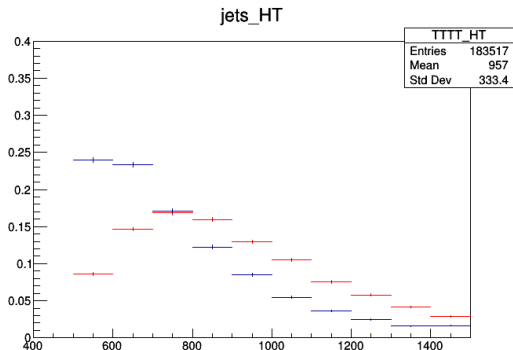


Figure: HT shape of TTTT and bg

- expected significance **0.0719726**
- expected limit **27.3750**
- might need better variable for the fitting

Outline

- 1 1τ 2L(BDT)
- 2 2τ 0l(BDT)
- 3 2τ 1l(HT)
- 4 1τ 0l(HT)
- 5 Combination of channels**

Combination of all subchannels for 2016

- Using BDT score of BDT channels(1tau1l, 1tau2l, 2tau0l) and HT of nonBDT channels(1tau0l, 2tau1l)
- Write a datacard for combination
- expected significance 0.710121
- expected limit 3.6719

datacard for 2016 combination

- use combineCards.py to generate combination datacard
- need to check and remove some empty process for Subchannel
- /publicfs/cms/user/huahuil/tauOfTTTT_NanoAOD/TMVAoutput/2016/
v2Add2Variables_fromV9/combination/2016_datacard.txt

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: 2tau1l

raw entries:

TTTT = 2723.000000
 TT = 76.000000
 TTX = 1767.000000
 VV = 0.000000
 SingleTop = 56.000000
 QCD = 0.000000
 allBg = 1899.000000

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: 2tauXl

raw entries:

TTTT = 9096.000000
 TT = 823.000000
 TTX = 6616.000000
 VV = 0.000000
 SingleTop = 313.000000
 QCD = 3.000000
 allBg = 7755.000000

2016: 2tau1l

scaled to LUMI:

TTTT = 0.114396
 TT = 3.200461
 TTX = 1.386142
 VV = 0.000000
 SingleTop = 0.013795
 QCD = 0.000000
 allBg = 4.600398

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

2016: 2tauXl

scaled to LUMI:

TTTT = 0.392067
 TT = 38.755373
 TTX = 5.714652
 VV = 0.000000
 SingleTop = 0.713523
 QCD = 9.271531
 allBg = 54.455079

Event Yield

2016: 1tau2os

raw entries:

TTTT = 8920.000000

TT = 88.000000

TTX = 2070.000000

VV = 0.000000

SingleTop = 88.000000

QCD = 0.000000

allBg = 2246.000000

2016: 1tau2ss

raw entries:

TTTT = 4904.000000

TT = 24.000000

TTX = 671.000000

VV = 0.000000

SingleTop = 15.000000

QCD = 0.000000

allBg = 710.000000

2016: 1tau2l

raw entries:

TTTT = 13824.000000

TT = 112.000000

TTX = 2741.000000

VV = 0.000000

SingleTop = 103.000000

QCD = 0.000000

allBg = 2956.000000

2016: 1tau2os

scaled to LUMI:

TTTT = 0.357489

TT = 3.850505

TTX = 1.880743

VV = 0.000000

SingleTop = 0.244627

QCD = 0.000000

allBg = 5.975875

2016: 1tau2ss

scaled to LUMI:

TTTT = 0.190199

TT = 0.952928

TTX = 0.450209

VV = 0.000000

SingleTop = 0.000547

QCD = 0.000000

allBg = 1.403684

2016: 1tau2l

scaled to LUMI:

TTTT = 0.547687

TT = 4.803433

TTX = 2.330952

VV = 0.000000

SingleTop = 0.245174

QCD = 0.000000

allBg = 7.379558