Status Report of TTTT Study in Tau Channel

Huiling Hua[1] Fabio Iemmi[1] Duncan Leggat[2] Hongbo Liao[1]
Hideki Okawa[2] Yu Zhang[2]

[1]IHEP

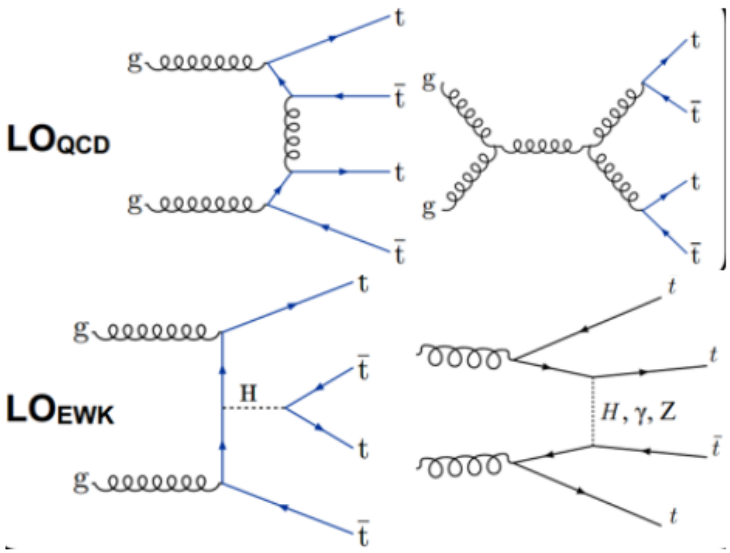
[2]Fudan University

TTX Meeting

Outline

Outline

Introduction and motivation



- Four tops process yet to be discovered
- Very rare(12 fb) SM process
- Sensitive to BSM
- Tau final states constitutes a significant decay mode
- We explore the TTTT from hardonic tau channel

Update to UL and NanoAOD dataset

- Major update: switched to UL NanoAOD(v9) dataset
  - preUL we used MINIAOD
- Updated analysis workflow to accomodate to NanoAOD
  - Have repeated previous study with 2016
- No SUSY top tagger for now
  - preUL data we implemented top tagger under cmssw with MINIAOD as input
- Tighter preselection and channel catergorization
  - To make trigger efficient in signal region

Channel catergerazation and analysis stradgy

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

subchannel	N_τ	N_l	N_{jet}	N_{bjet}	analysis stradgy	notes
1tau0l	1	0	8	2	HT	QCD estimation
1tau1l	1	1	7	2	BDT	
1tau2l	1	2	6	2	BDT	
2tau0l	2	0	7	2	HT	
2tau1l	2	1	6	2	HT	lack of statistics

Table: Subchannel definition

Outline

Trigger study recap

- Goal
  - Finding triggers highly efficient for TTTT
  - Calculating data/MC correction
- Attemps and strategy
  - Tau triggers inefficient for our signal
  - HT+mutijet+btag triggers turn out to be good(inspiration from TTTT hardonic channel)
  - Reference trigger method to measure efficiency of data

Trigger efficiency of data and MC(2016preVFP)

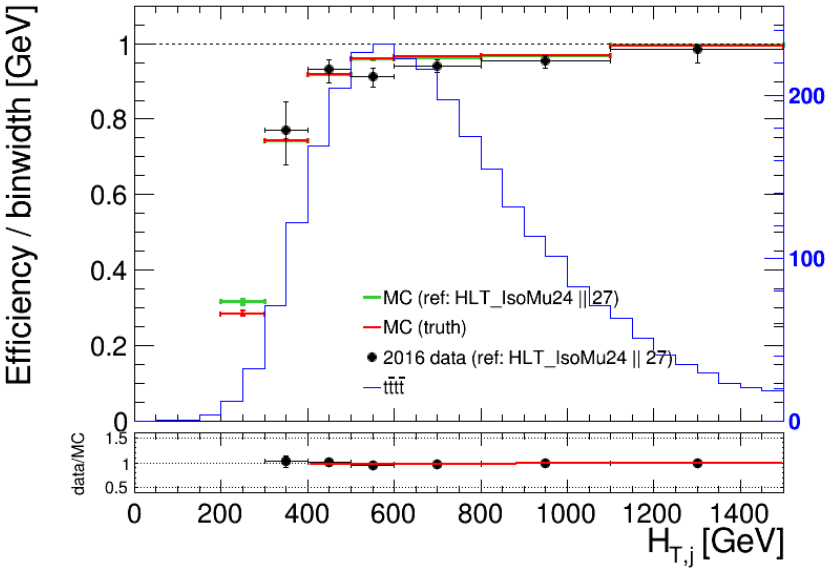


Figure: Trigger efficiency for 2016preVFP

- HTL trigger choice(OR)
  - HLT_PFHT450_SixJet40_BTagCSV_p056
  - HLT_PFHT400_SixJet30_DoubleBTagCSV_p056
  - HLT_PFJet450(recover data efficiency in hight HT)
- Reference trigger(unbiased)
  - HLT_IsoMu24
  - HLT_IsoMu27
- Selection
  - Preselection + 1 tight muon
  - Using Singlemuon dataset

Data/MC rate of trigger efficiency(2016preVFP)

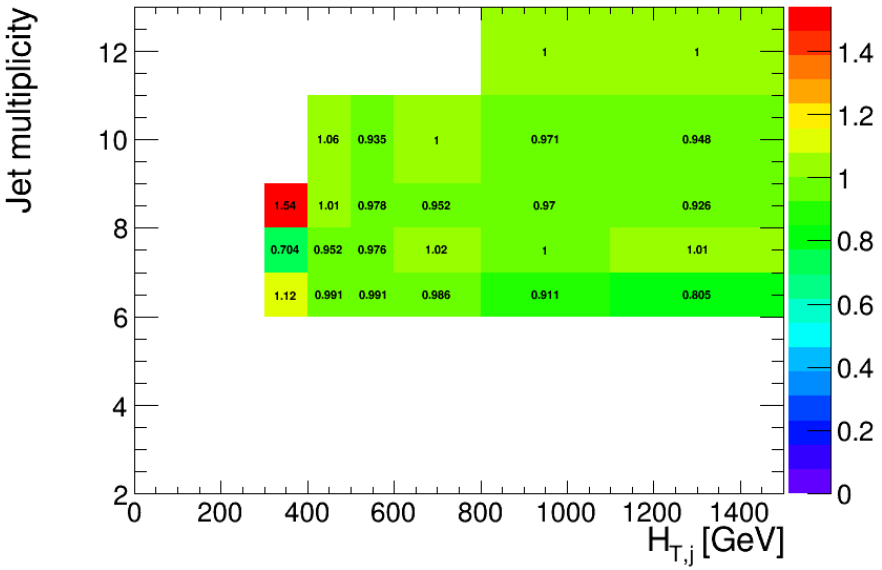


Figure: data/MC

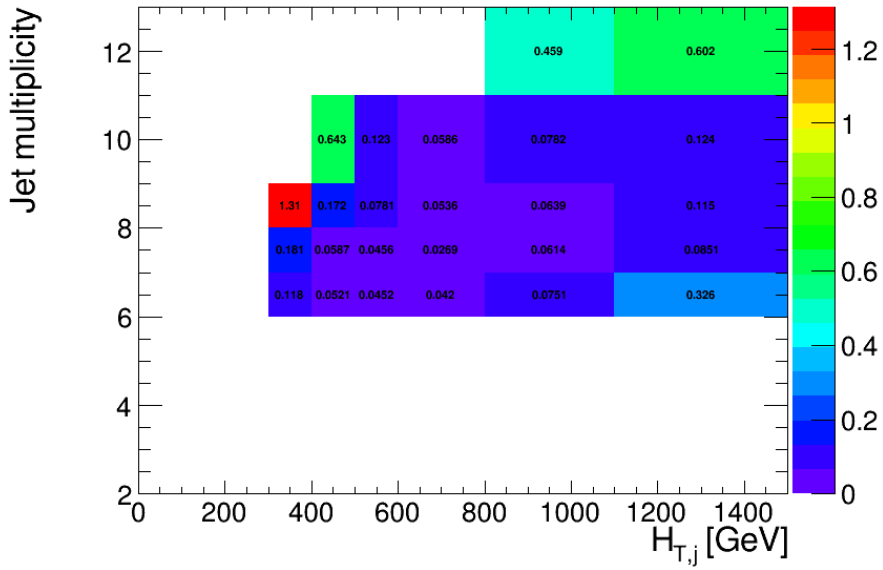


Figure: data/MC rate error

- Source for HLT efficiency correction for MC
- Results of 2016postVFP in backup

Outline

Estimation of QCD in 1tau0l

- Motivation
  - QCD constitutes a major background in 1tau0l
  - Low selection efficiency and statistics for QCD
  - Large theoretical uncertainties on cross sections and NLO corrections
- Data driven method for QCD
  - Fake rate method for yield
  - Extrapolate HT shape from control region using transition function

Fake rate method for yield of QCD

$$N_{FT\tau} = \sum_{pt,\eta} N_{FT\tau}(pt,\eta) = \sum_{pt,\eta}[N_{F\overline{T}}(pt,\eta) \times \frac{FR(pt,\eta)}{1 - FR(pt,\eta)}] \quad (1)$$

	seletion				MC event yield			
	N_τ	N_l	N_{jet}	N_{bjet}	TTTT	TT	QCD	TTX
SR(1tau0l)	1	0	8	2	8	4186	1903	120
CR	1	0	8	0	0	222	3567	5

Table: Region definition

- Calculating $FR(pt,\eta)$ from control region(enriched with fakeable τs)
- Apply FR in signal region(SR) and application region to get $N_{FT\tau}$

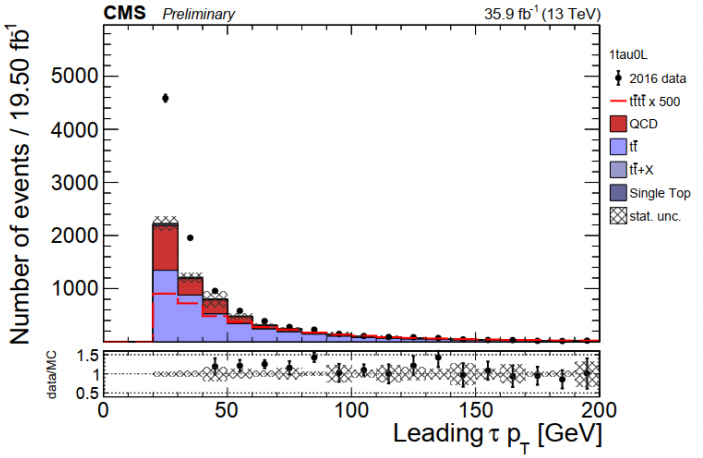
Closure test of fake rate method

- Apply the method in the same CR, compare with number of events in CR counting from MC
- Apply the method in the SR, compare with number of events in SR counting from MC
- Apply the method in CR, compare with number of events in CR counting from data

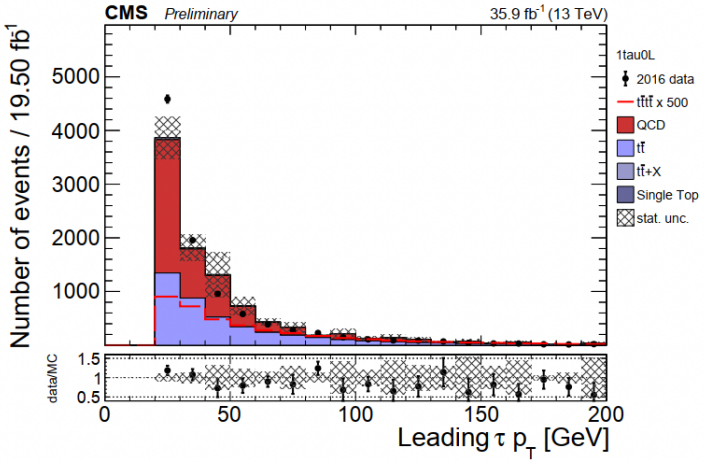
	CR(compare with MC)		SR(compare with MC)		CR(compare with data)	
	value	raw entries	value	raw entries	value	raw entries
Fake rate method	3384 ± 804	-	1834 ± 534	-	8084 ± 329	-
Counting	3264 ± 450	471	2223 ± 454	331	8084 ± 90	8084

Table: Closure test results

Event yield from fake rate method(2016)




```
======== taus_1pt 1tau0L ====================
Data events:  9789
signal events: 7.88975
ttbar events: 4186.48
QCD events: 1902.53
tt+X events:   120.151
single top events:   0.193993
total MC events: 6209.36
data/MC agreement: 57.6491%
```

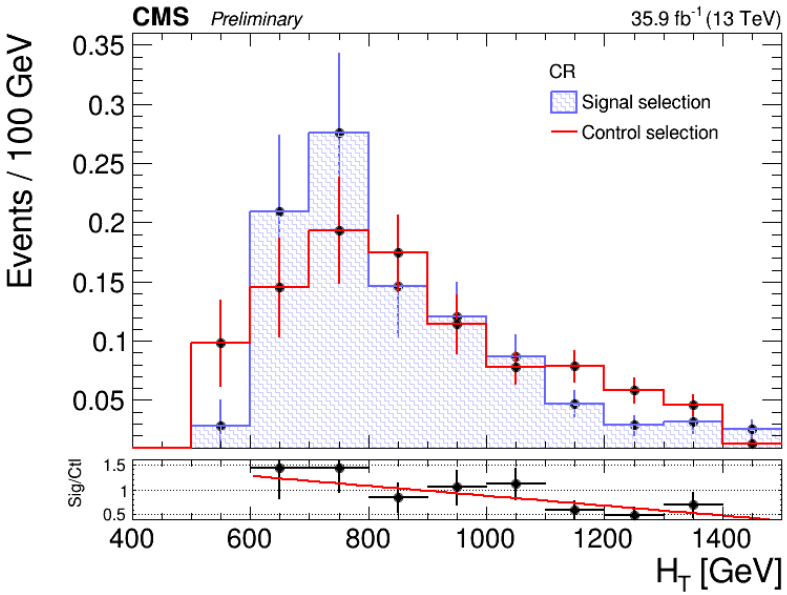


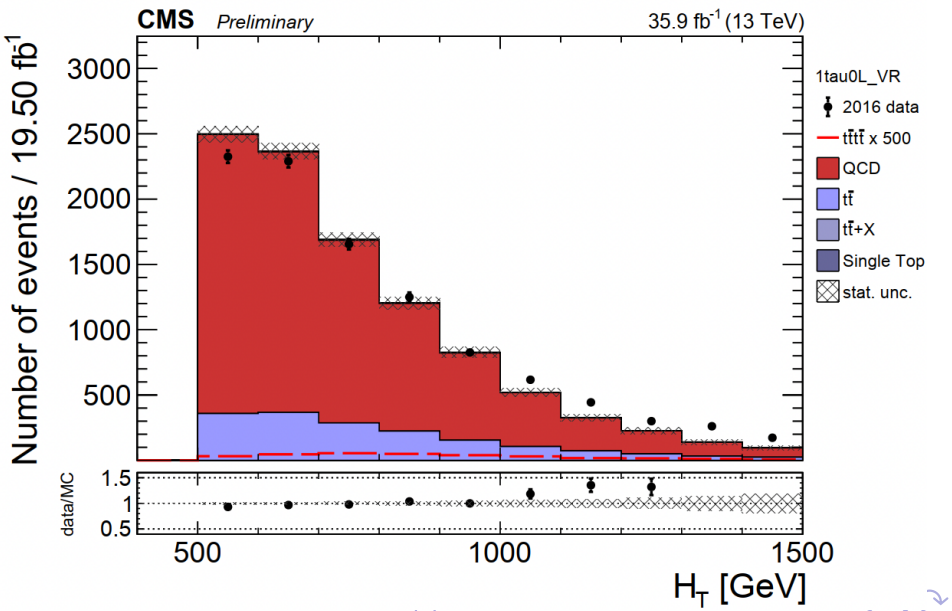

```
======== taus_1pt 1tau0L ====================
Data events:  9789
signal events: 7.88975
ttbar events: 4186.48
QCD events: 5617
tt+X events:   120.151
single top events:   0.193993
total MC events: 9923.83
data/MC agreement: -1.35864%
```

- For 2016 estimated QCD yield is 5617

QCD shape estimation

- Basic idea
  - Take QCD shape(data) in CR => correct the shape with transition function => get QCD shape in SR
- Transition function
  - Use HT shape(MC) in CR and region of interest
  - Fit the shape to get a transition function



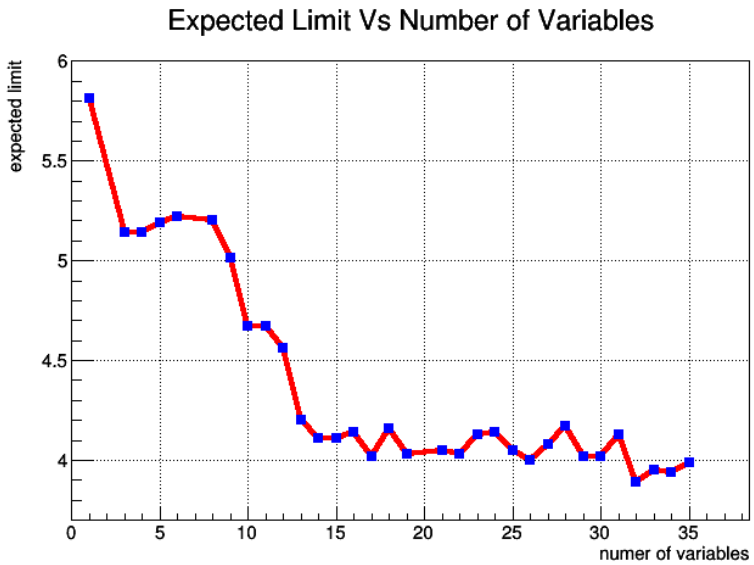


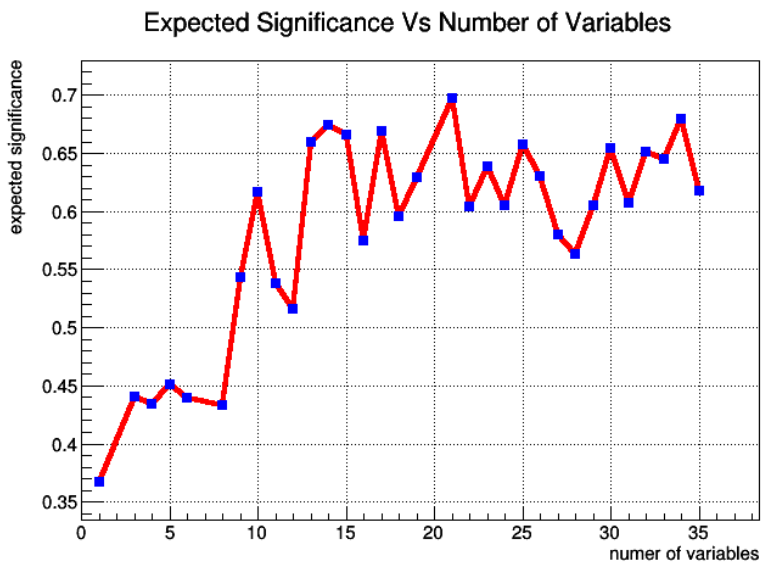
Outline

MVA optimazation stradegy

- Correlation removal method
  - Start with 36(50) most powerful variables
  - Generate 36 variable list by remove one variable at a time according to Correlation
  - Run BDT training with the 36 lists for 36 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

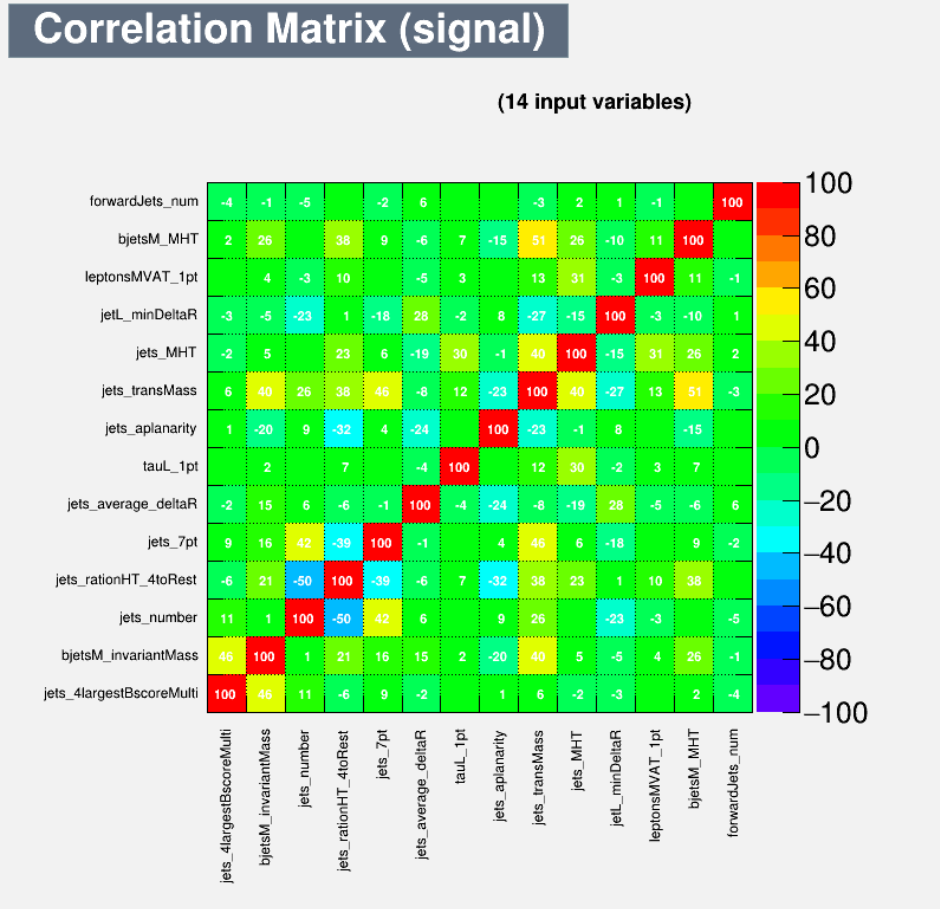
Training performance Vs number of input variables(1tau1l)

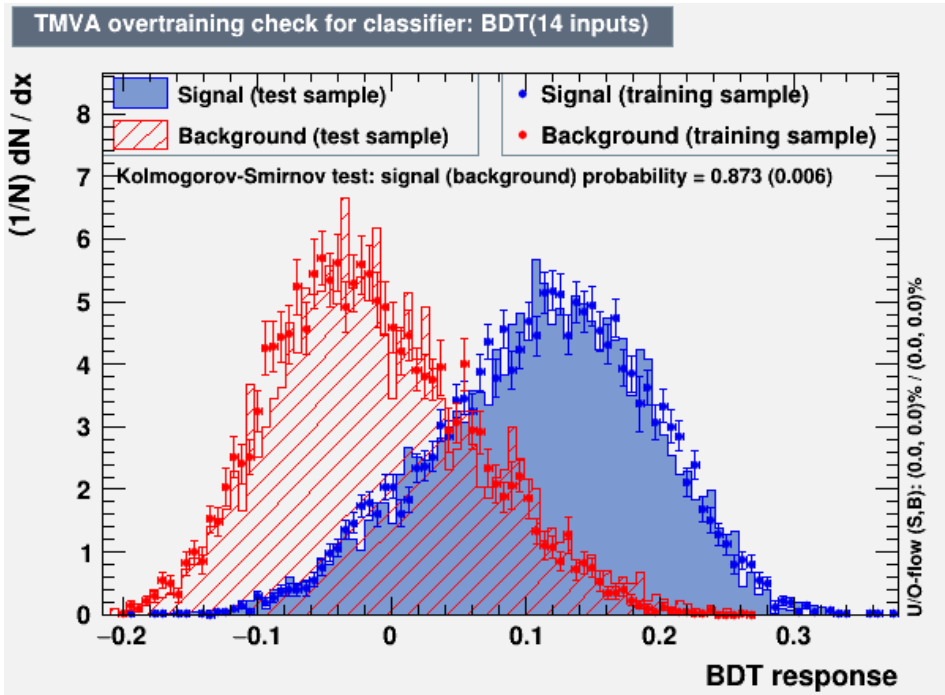




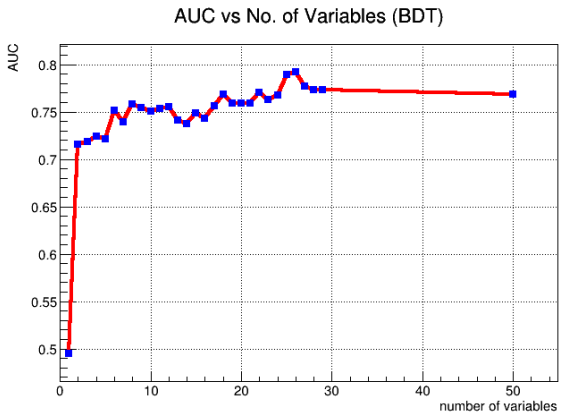
- Reached plateau at around 14 variables
- Expected significance 0.66, expected limit 4.1

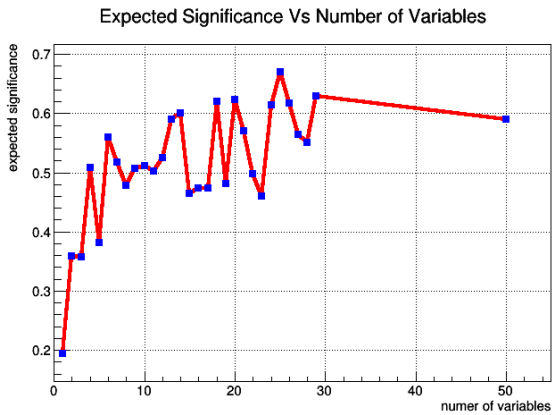
14 input variables training performance(1tau1l)

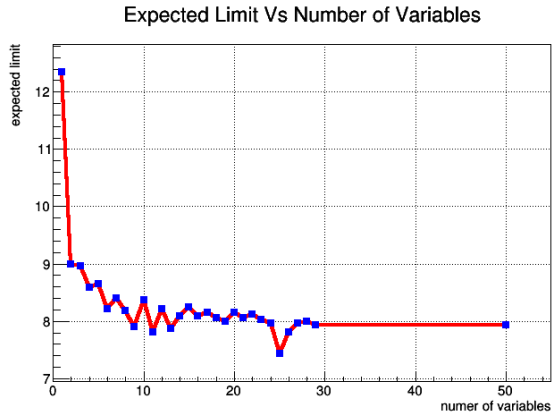




Training performance Vs number of input variables(1tau2l)







- Reached plateau of 0.5 significance at around 8 input variables

Outline

Combination of subchannels in 2016

- Strategy
  - BDT score of BDT channels(1tau1l, 1tau2l) and HT of non BDT channels(1tau0l, 2tau1l, 2tau0l)
  - Writen a datacard for combination
  - Blinded with combine
  - Statistic uncertainty only
- Results
  - Expected significance 0.832571
  - Expected limit 3.2969

Summary and plan

- Summary of progress
  - Switched to UL NanoAOD dataset
  - Have done HLTstudy, BDT optimazation and QCD Estimation
  - Got prelimilary expected significance and limit results(2016)
- Future plan
  - Incorperate systematics
  - Analyze 2017 and 2018 data
  - Optimazation with BDT hyper parameter tuning
  - Looking for combination with SS group

Backup

back up

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: 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: 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

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

Trigger efficiency in 2016postVFP

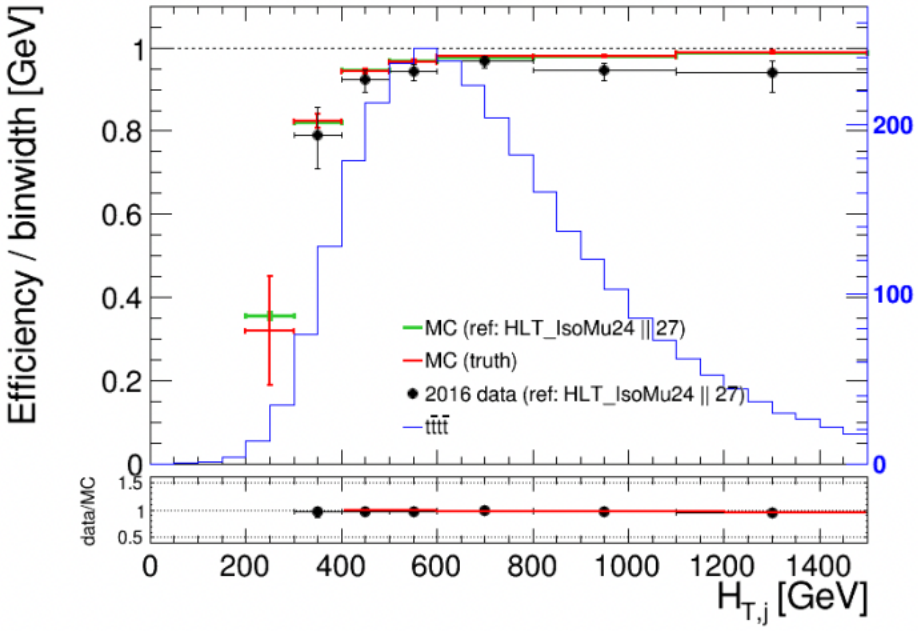


Figure: data/MC

- Idea is to perform **validation of the FR method in a region with similar background composition to the signal region**
- Validation is a **data/MC agreement** check on the variable they are going to use in final fit
- Something similar is done in EXO-19-015

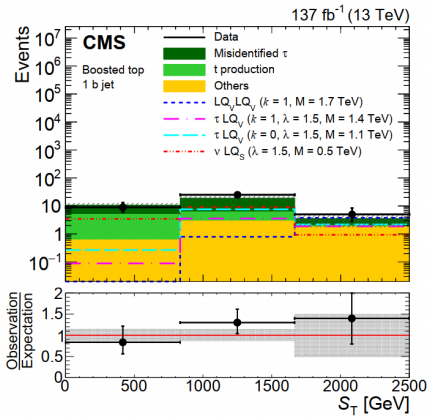
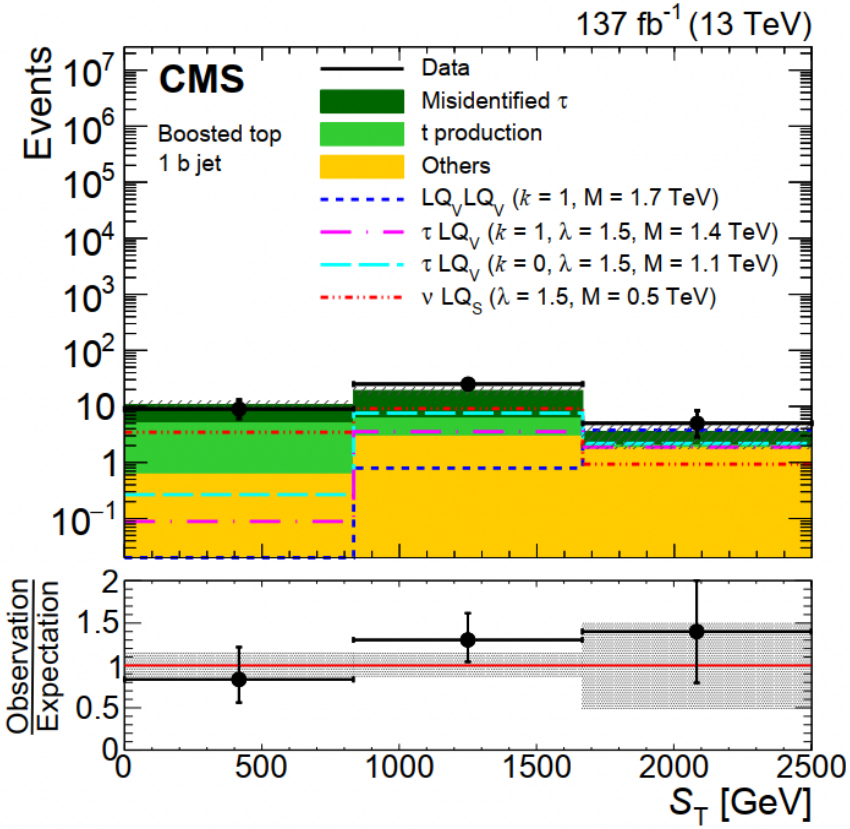


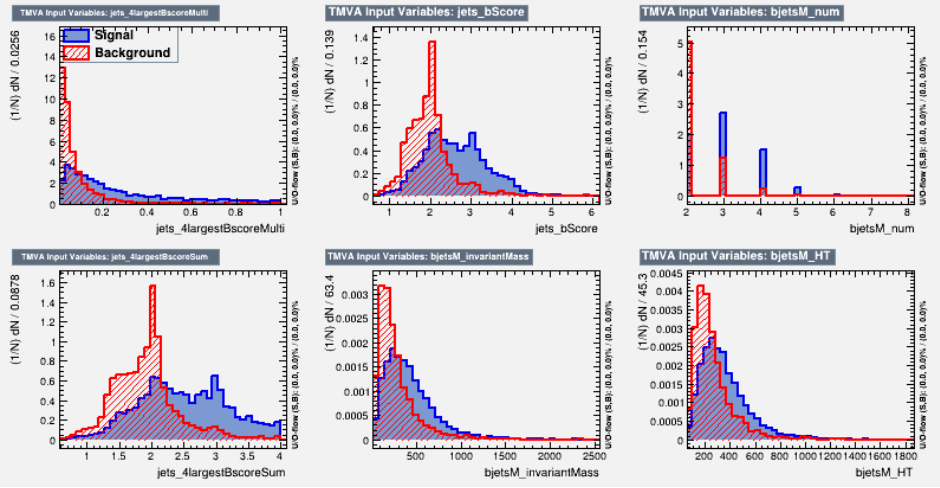
Figure: data/MC rario error

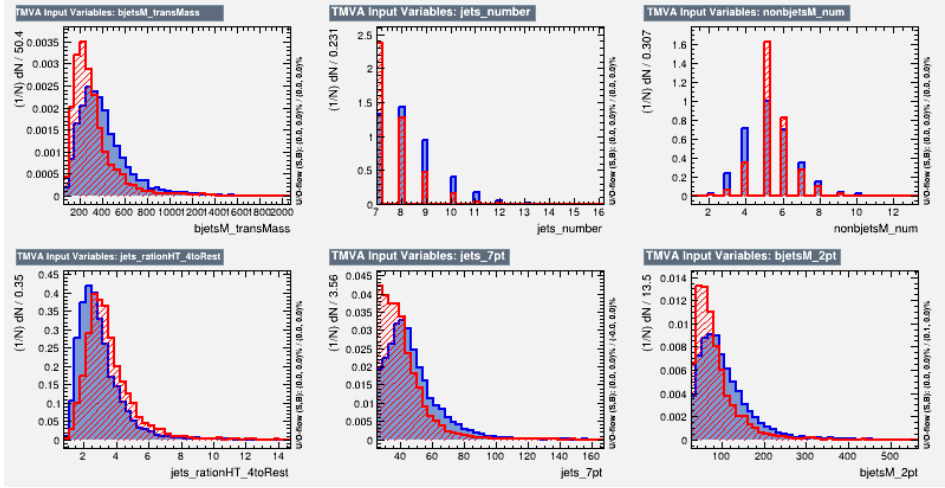
Uncertainties on FR method

- Idea is to perform **validation of the FR method in a region with similar background composition to the signal region**
- Validation is a **data/MC agreement** check on the variable they are going to use in final fit
- Something similar is done in EXO-19-015

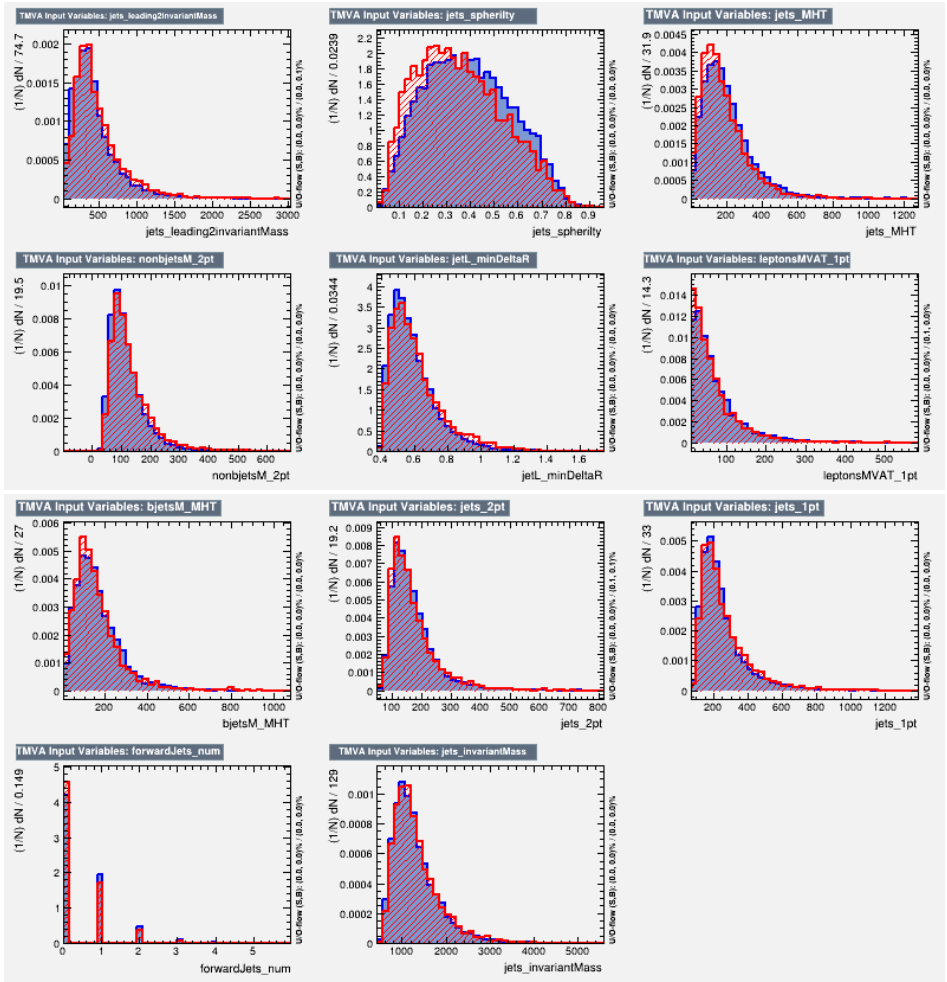


BDT variables(1tau1l)

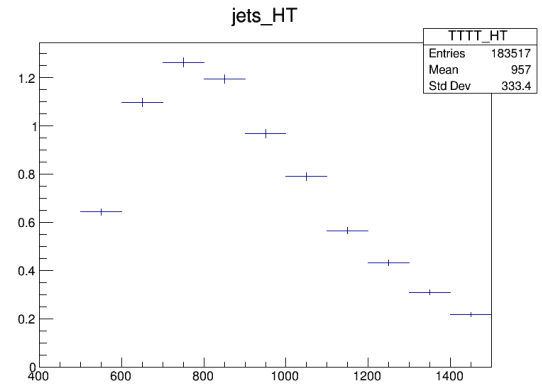


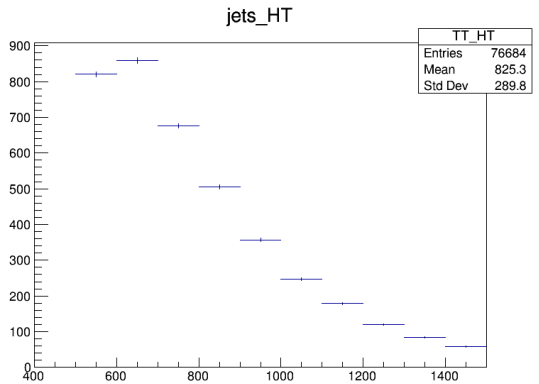


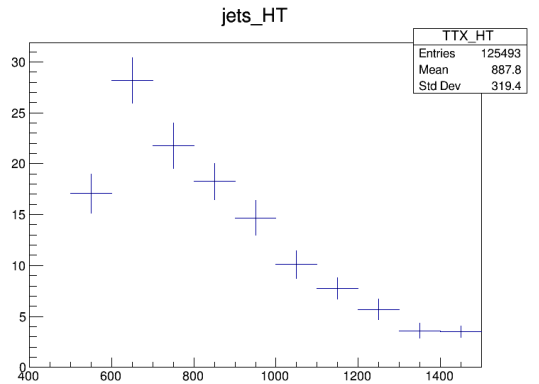
BDT variables(1tau1l)

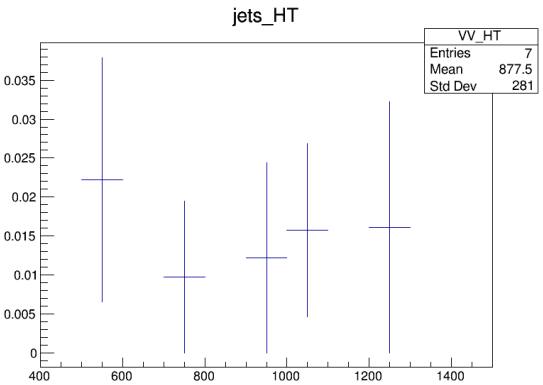


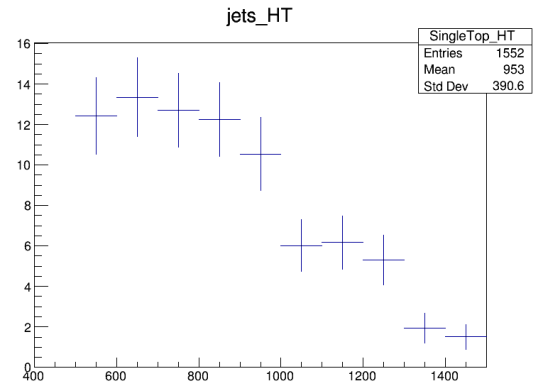
1tau0l HT distribution

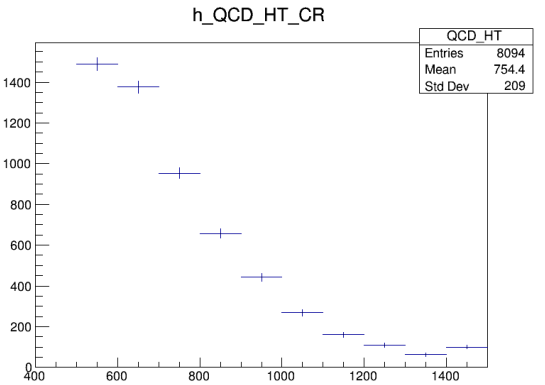












- 1tau0l use HT as templates for combine
- HT of QCD from data driven method
- HT of other processes from MC

Fitting results of 1tau0l

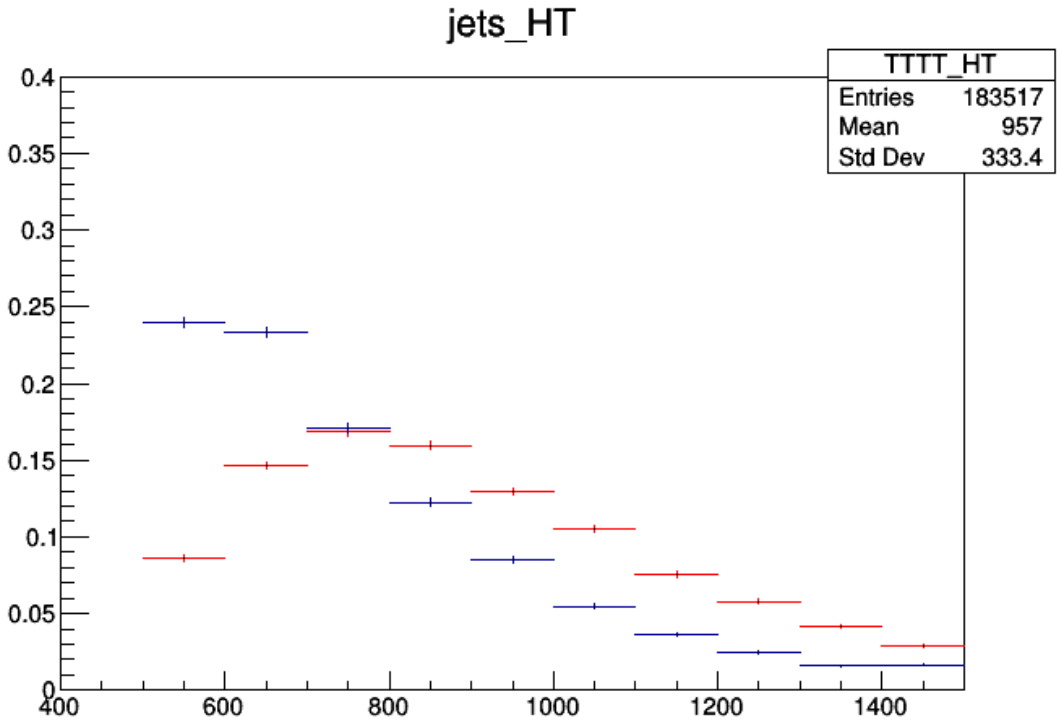


Figure: HT shape of TTTT and bg

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