



Variable Validation with Data MC Comparison

Search for Four top in Tau Final States

Anshul Kapoor¹ Fabio Lemmi¹ Huiling Hua¹
Hongbo Liao¹ Hideki Okawa² Yu Zhang²

¹IHEP

²Fudan University

July 6, 2022



Outline

- ① $1\tau_0$
- ② $1\tau_1$
- ③ Miscellaneous

Section 1

1tau0l



Control region definition (1tau0l)

- Signal depleted, background dominated
- Close to signal region
- To examine if background are well modeled or corrected or estimated.

	N_τ	N_l	N_{jets}	N_b
SR	1	0	≥ 8	≥ 2
CR	1	0	≥ 8	1
VR	1	0	≥ 8	0
CR2	1	0	< 8	≥ 2
CR3	1	0	< 7	≥ 2
CR4	1	0	$= 7$	≥ 2

Table 1: 1tau0l, CR and VR from Fabio

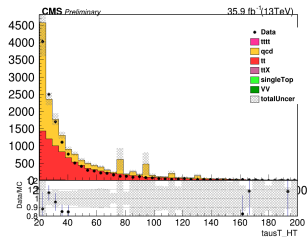
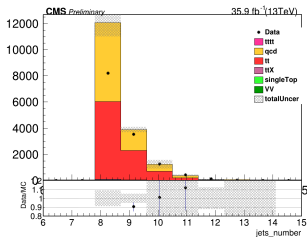
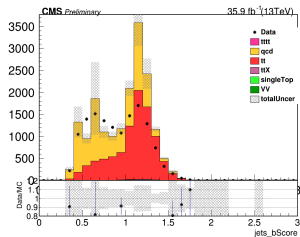
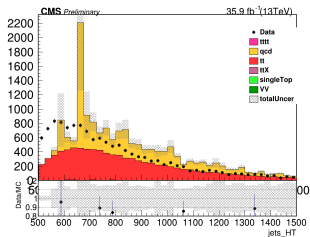
CR/VR background composition(1tau0l)

process	SR	CR	VR	CR2	CR3	CR4
TTTT	9.29	0.09	0.92	2.79	0.73	2.07
TT	6190.80	479.11	3046.66	12579.78	5885.25	6694.5
TTX	190.38	14.45	76.94	319.85	145.24	174.61
VV	0.09	0.27	0.18	0.38	0.15	0.23
SingleTop	131.85	15.23	83.37	348.22	173.53	174.70
QCD	3945.73	12504.31	8315.00	8196.65	3283.88	4912.7
allBg	10458.86	13013.37	11522.15	21444.90	9488.05	11956.8

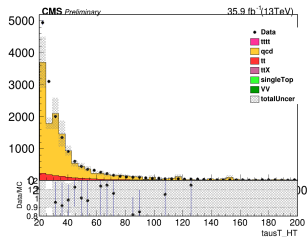
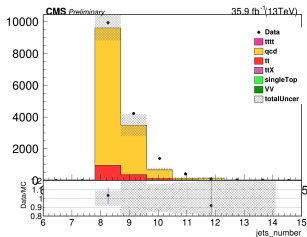
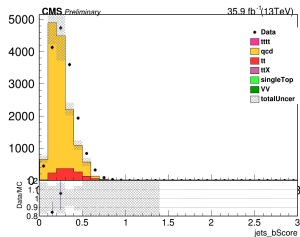
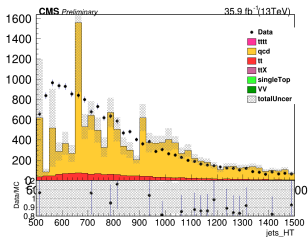
Table 2: 1tau0l, correction applied, all MC prediction

- basic weight applied(genWeight, prefiringWeight, pileupWeight); JES and JER applied
- CR: QCD control region
- CR2 : maybe TT control region

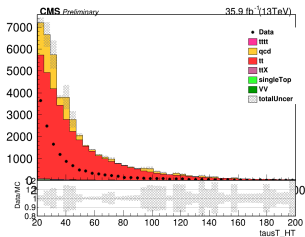
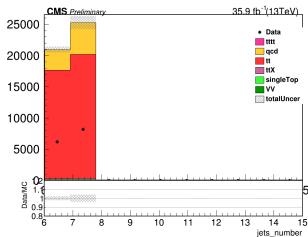
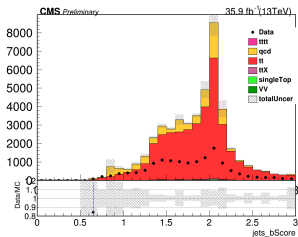
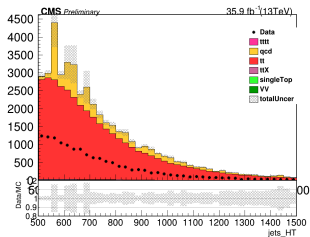
Variable distribution in CR



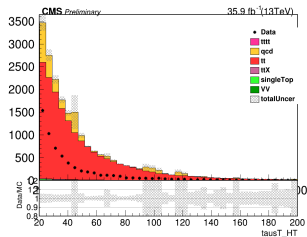
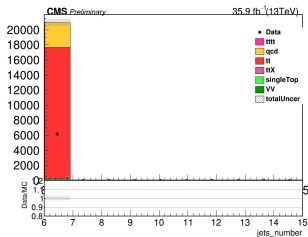
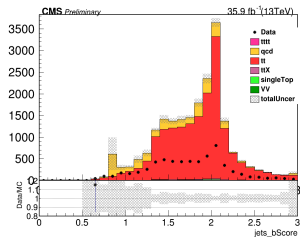
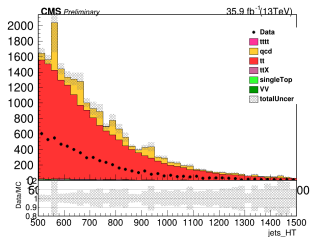
Variable distribution in VR



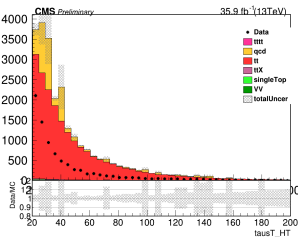
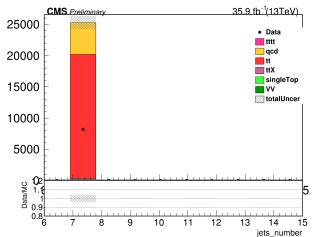
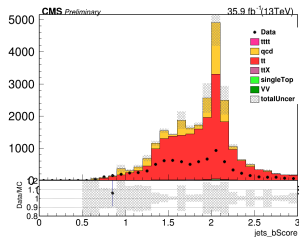
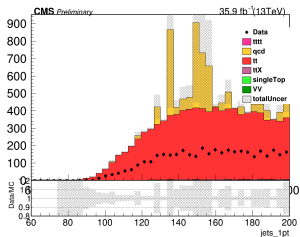
Variable distribution in CR2



Variable distribution in CR3



Variable distribution in CR4



Section 2

1tau1l



Control region definition(1tau1l)

- Signal depleted, background dominated
- Close to signal region
- To examine if background are well modeled or corrected or estimated.

	N_τ	N_l	N_{jets}	N_b
SR	1	1	≥ 7	≥ 2
CR0	1	1	≥ 7	1
CR1	1	1	≥ 7	0
CR2	1	1	6	≥ 2
CR3	1	1	6	< 2

Table 3: 1tau1l

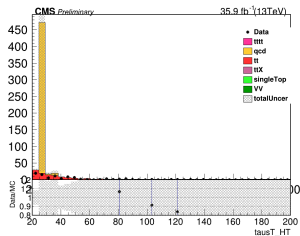
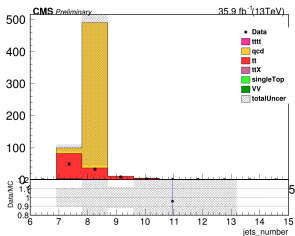
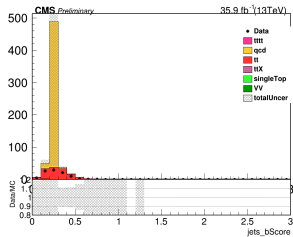
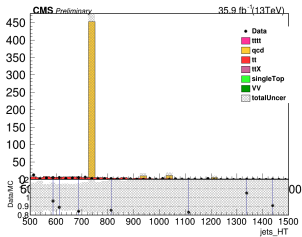
CR background composition(1tau1l)

process	SR	CR0	CR1	CR2	CR3
tttt	5.09	0.51	0.04	0.76	0.13
tt	1233.46	788.81	152.66	751.29	609.64
qcd	0.00	5.71	471.60	5.24	97.87
ttX	3790	22.06	2.86	21.37	13.08
VV	0.02	0.04	0.02	0.00	0.07
singleTop	13.76	10.56	2.17	8.89	8.94

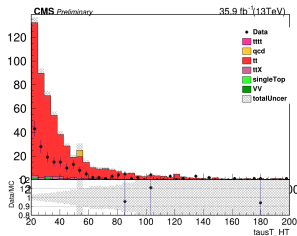
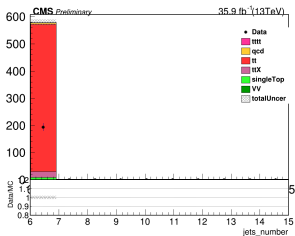
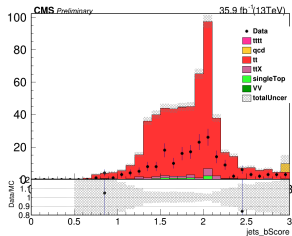
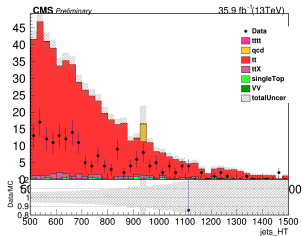
Table 4: 1tau1l, correction applied, all MC prediction

- 1tau1l major background in SR: tt
- CR0 and CR2: tt control region
- CR1: qcd control region
- basic weight applied(genWeight, prefiringWeight, pileupWeight); JES and JER applied

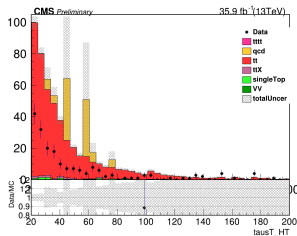
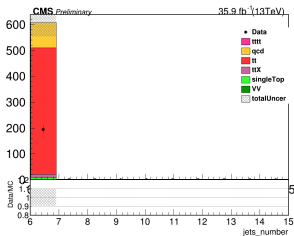
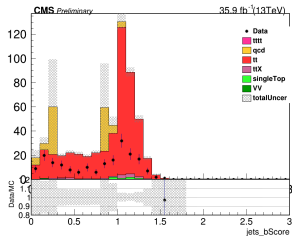
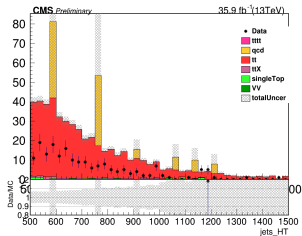
Data/mc in CR1



Data/mc in CR2



Data/mc in CR3



Section 3

Miscellaneous



Analysis in progress

channel	strategy	object	HLT	bg estimation	variable validation	BDT	sytematic	overlap study
1tau0l	HT	✓	✓	✓	✓	no need	ongoing	
1tau1l	BDT	✓	✓	no need	✓	✓		
1tau2l	BDT	✓	✓	no need		✓		

- AN writing is starting
- Table is just for 2016
- Big goal is getting CADI line before this September!

To do next week

- Figure out why the t_t in CR are over estimated(1 day)
- Continue systematic study
- Write AN

Section 4

Back up



Data/MC comparison plot: how to deal with uncertainty

- Uncertainty in stack for all MC processes
 - $\sigma_{total} = \sum_{n=1}^N \sqrt{\sigma_i^2}$
- Uncertainty in the *data/MC* ratio
 - shadow uncertainties around 1: $\frac{\sigma_{total}}{n_i}$
 - uncertainty on data/mc : simply Divide data and mc histogram, error bar calculated by root