



Expected yields after SF application and top tagger studies

Huiling Hua¹ **Fabio Lemmi**¹
Hongbo Liao¹ Hideki Okawa² Yu Zhang²

¹Institute of High Energy Physics (IHEP), Beijing

²Fudan University, Shanghai

July 7, 2021



- Generator weight (not exactly a SF)
- Prefiring weight (not exactly a SF, right?)
- Pileup SF
- Trigger SF (using my 2D histogram `njets_vs_HT`)
- b tagging SF (using latest version of `evalEventSF.C` that I described during last meeting)
- Assume all the weights to be independent: **final weight is the product of weights**

Expected yields: before SFs



	tttt	ttbar	QCD	tt+X	MC total	data	(D-MC)/D	Z
1tau0L	9.85	6371.32	7456.68	191.53	14029.38	13693.00	-0.02	0.08
1tau1L	7.52	1942.57	2.17	81.22	2033.48	1633.00	-0.25	0.17
1tau2L	1.53	32.69	0.00	11.46	45.68	44.00	-0.04	0.23
1tau3L	0.09	0.00	0.00	0.68	0.77	1.00	0.23	0.11
2tau0L	0.50	207.56	1.54	14.73	224.33	225.00	0.00	0.03
2tau1L	0.21	11.53	0.00	4.32	16.05	13.00	-0.23	0.05
2tau2L	0.02	0.10	0.00	0.26	0.37	0.00	nan	0.03

- **Zs do not change dramatically** wrt SF scenario (see next slide)
- The procedure seems OK at first glance
- **Missing minor processes** (we have to agree on what processes and what cross sections to use)

Yields and top
tagger

F. Lemmi

Expected yields: after SFs



	tttt	ttbar	QCD	tt+X	MC total	data	(D-MC)/D	Z
1tau0L	8.77	5392.31	6559.65	162.39	12123.13	13693.00	0.11	0.08
1tau1L	6.68	1628.15	2.15	69.22	1706.20	1633.00	-0.04	0.16
1tau2L	1.33	26.27	0.00	9.33	36.93	44.00	0.16	0.22
1tau3L	0.08	0.00	0.00	0.57	0.65	1.00	0.35	0.10
2tau0L	0.44	168.28	1.70	12.57	183.00	225.00	0.19	0.03
2tau1L	0.18	8.94	0.00	3.62	12.74	13.00	0.02	0.05
2tau2L	0.02	0.09	0.00	0.21	0.31	0.00	nan	0.03

- **Zs do not change dramatically** wrt no-SF scenario (see previous slide)
- The procedure seems OK at first glance
- **Missing minor processes** (we have to agree on what processes and what cross sections to use)

Yields and top
tagger

F. Lemmi



- We are currently **not using the top tagger information**
- Maybe it can be part of BDT input variable, not sure yet though
- Tried to see what happens when **using it in category definition**
 - **Scenario 1**
 - **Scenario 2**
 - **Scenario 3**
- Description of the various scenarios is given in the following slides

Scenario 1



Yields and top
tagger

F. Lemmi

- Scenario 1 is the **loosest scenario** possible
- Ask for **at least 1 resolved top candidate in all categories**
 - Except for 1tau3L and 2tau2L where you do not expect any hadronic tops



- Scenario 2 is the **tightest scenario** possible
- Ask for **exactly the number of resolved top candidates (TCs) you would expect**
 - 1tau0L: ≥ 3 TCs
 - 1tau1L: ≥ 2 TCs
 - 1tau2L: ≥ 1 TCs
 - 1tau3L: no requirements on TCs
 - 2tau0L: ≥ 2 TCs
 - 2tau1L: ≥ 1 TCs
 - 2tau2L: no requirements on TCs



- Scenario 3 is a **compromise between scenarios 1 and 2**
 - 1tau0L: ≥ 2 TCs
 - 1tau1L: ≥ 1 TCs
 - 1tau2L: ≥ 1 TCs
 - 1tau3L: no requirements on TCs
 - 2tau0L: ≥ 1 TCs
 - 2tau1L: ≥ 1 TCs
 - 2tau2L: no requirements on TCs

Scenario 1



Yields and top
tagger

F. Lemmi

	tttt	ttbar	QCD	tt+X	MC total	data	(D-MC)/D	Z
1tau0L	8.73	5350.91	6488.36	159.93	12007.93	13222.00	0.09	0.08
1tau1L	6.48	1493.34	2.15	65.32	1567.28	1502.00	-0.04	0.16
1tau2L	1.18	16.01	0.00	8.00	25.19	26.00	0.03	0.24
1tau3L	0.08	0.00	0.00	0.57	0.65	1.00	0.35	0.10
2tau0L	0.44	165.14	1.70	12.18	179.46	214.00	0.16	0.03
2tau1L	0.17	7.61	0.00	3.30	11.08	12.00	0.08	0.05
2tau2L	0.02	0.09	0.00	0.21	0.31	0.00	nan	0.03

- No dramatic changes wrt no-top-tagger scenario
- Significances unchanged except for 1tau2L (slight increase)
- No significant gain in data/MC agreement

Scenario 2



Yields and top
tagger

F. Lemmi

	tttt	ttbar	QCD	tt+X	MC total	data	(D-MC)/D	Z
1tau0L	3.04	457.08	206.22	17.02	683.36	880.00	0.22	0.12
1tau1L	4.28	507.82	0.00	28.78	540.87	556.00	0.03	0.18
1tau2L	1.18	16.01	0.00	8.00	25.19	26.00	0.03	0.24
1tau3L	0.08	0.00	0.00	0.57	0.65	1.00	0.35	0.10
2tau0L	0.35	103.61	1.27	9.28	114.50	118.00	0.03	0.03
2tau1L	0.17	7.61	0.00	3.30	11.08	12.00	0.08	0.05
2tau2L	0.02	0.09	0.00	0.21	0.31	0.00	nan	0.03

- Nice **increase in significance for 1tau0L**
 - **QCD is hugely reduced!**
- Data/MC agreement looks OK

Scenario 3



	tttt	ttbar	QCD	tt+X	MC total	data	(D-MC)/D	Z
1tau0L	7.61	4023.86	3865.60	117.90	8014.98	8309.00	0.04	0.09
1tau1L	6.48	1493.34	2.15	65.32	1567.28	1502.00	-0.04	0.16
1tau2L	1.18	16.01	0.00	8.00	25.19	26.00	0.03	0.24
1tau3L	0.08	0.00	0.00	0.57	0.65	1.00	0.35	0.10
2tau0L	0.44	165.14	1.70	12.18	179.46	214.00	0.16	0.03
2tau1L	0.17	7.61	0.00	3.30	11.08	12.00	0.08	0.05
2tau2L	0.02	0.09	0.00	0.21	0.31	0.00	nan	0.03

- Decrease in significance in 1tau0L and 1tau1L wrt scenario 2
 - **QCD is back**
- Data/MC agreement looks OK
- Only slight increase in significances (1tau0L, 1tau2L) wrt no-top-tagger scenario

Yields and top
tagger

F. Lemmi



- Scenario 2 seems nice in terms of significances, but are we left with enough statistics (e.g., for BDT training or final fit)?
- Let's have a look at raw entries
- Produce same tables, replace `h->Integral(-1,-1)` with `h->GetEntries()`

Raw entries: scenario 2



Yields and top
tagger

F. Lemmi

	tttt	ttbar	QCD	tt+X	MC total	data	(D-MC)/D	Z
1tau0L	89273	3716	12	9224	102225	880	-115	494
1tau1L	116756	8383	0	15625	140764	556	-252	514
1tau2L	32526	408	0	4576	37510	26	-1442	294
1tau3L	2215	0	0	321	2536	1	-2535	78
2tau0L	9093	1287	1	5015	15396	118	-129	97
2tau1L	4763	152	0	1984	6899	12	-574	82
2tau2L	500	2	0	135	637	0	nan	31

- Just 12 QCD events in 1tau0L...
 - **Not enough for training**
- QCD would still be 30% of the background, probably not safe to neglect it in training

Raw entries: scenario 3



Yields and top
tagger

F. Lemmi

	tttt	ttbar	QCD	tt+X	MC total	data	(D-MC)/D	Z
1tau0L	207363	33472	110	47547	288492	8309	-34	563
1tau1L	169380	26201	2	30790	226373	1502	-150	535
1tau2L	32526	408	0	4576	37510	26	-1442	294
1tau3L	2215	0	0	321	2536	1	-2535	78
2tau0L	10969	2239	2	6521	19731	214	-91	100
2tau1L	4763	152	0	1984	6899	12	-574	82
2tau2L	500	2	0	135	637	0	nan	31

- Just 110 QCD events in 1tau0L...
 - **Not enough for training**
- QCD would still be 50% of the background, not safe to neglect it in training