Fake rates in 2016

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

[1]Institute of High Energy Physics (IHEP), Beijing

[2]Fudan University, Shanghai

April 20, 2022

QCD-enriched control region (CR)

- QCD is major background yield in $1\tau\,0\text{L}$
- All the remaning **major backgrounds** ($t\bar{t}$ and $t\bar{t}$+X) **and signal involve** top quarks, i.e., **bottom quarks** in the final state
- **Revert the request on the number of b tagged jets in the event**

	N_{τ_h}	N_ℓ	N_{jets}	N_{bjets}
$1\tau\,0\text{L}$	1	0	≥ 8	≥ 2
$1\tau\,0\text{L}$ ctl	1	0	≥ 8	0

- $N_{\text{bjets}} = 0$ is meant to reject all the top-related processes

Fake rate method

- Estimate the background completely from data by doing

$$N_{\text{fake-}\tau} = \sum_{p_T,\eta} N_{\text{fake-}\tau}(p_T,\eta) = \sum_{p_T,\eta} \left[N_{F,\overline{T}}(p_T,\eta) \times \frac{\text{FR}(p_T,\eta)}{1-\text{FR}(p_T,\eta)} \right]$$

- $\text{FR}(p_T,\eta)$ is computed in the CR: enriched in fake taus
- $N_{F,\overline{T}}(p_T,\eta)$ is computed in the application region (AR): just the SR but with fakeable taus
- Parametrize as a function of p_T, η of fakeable tau
- Binning in (p_T,η): $p_T \in [20, 30, 75, 150, 300, \text{Inf}]$; $\eta \in [0, 1.5, 2.3]$

Yields full 2016

	$t\bar{t}t\bar{t}$	$t\bar{t}$	QCD	$t\bar{t}$+X
1τ 0L SR	8	4186	1902.53	120
1τ 0L CR	0	222	3567	5

- **The CR is QCD-dominated**
- QCD events form 94% of the events in CR
- The **large QCD simulated yield** that we get **in CR comes from fake taus** (i.e., QCD jets misidentified as hadronic taus)

Closure test in MC QCD

- Compute FR in CR, **apply the method in** the same **CR**
- **Compare with number** of events in CR **you count from MC**
- These numbers should close

	Value	Raw entries
Counting	3264 ± 450	471
Fake rate method	3384 ± 804	–

- Values are in **agreement** within the uncertainties, **closure is not perfect** (3% discrepancy)
  - **Due to approximations** in weighting and summing `TEfficiency` objects
  - See my discussion with ROOT developer Lorenzo Moneta here
  - Also **not applying SFs**!!!

Fake rate method in MC QCD

- Compute FR in CR, **apply the method in** the **application region** (same as signal region, but use fakeable-not-tight taus)
- **Compare with number** of events in SR **you count from MC**

	Value	Raw entries
Counting	2223 ± 454	331
Fake rate method	1834 ± 534	–

- Values are in **agreement within the uncertainties**
- Uncertainties are big due to poor statistics in MC samples

Closure test in data

- Compute FR in CR, **apply the method in** the same **CR**
- **Compare with number** of events in CR **you count from data**
- This should close

	Value	Raw entries
Counting	8084 ± 90	8084
Fake rate method	8084 ± 329	–

- Values are in **agreement** within the uncertainties, **perfect closure**
  - No weighting of any kind of objects is needed for data

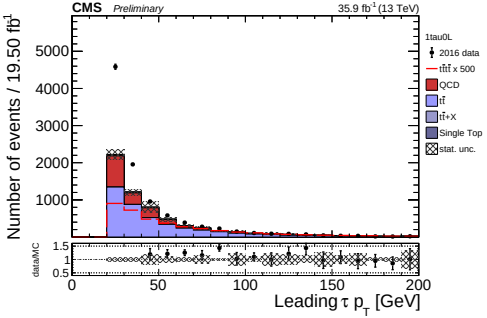
Fake rate method in data

- Compute FR in CR, **apply the method in** the **application region** (same as signal region, but use fakeable-not-tight taus)
- **Do not compare with number** of events in SR **you count from data**: **we are blinded**
- **IMPORTANT!**: take care of subtracting $t\bar{t}$ and $t\bar{t}$+X from $N_{\mathrm{F},\overline{\mathrm{T}}}(p_{\mathrm{T}}, \eta)$

	Value	Raw entries
Counting	–	–
Fake rate method	5617 ± 246	–

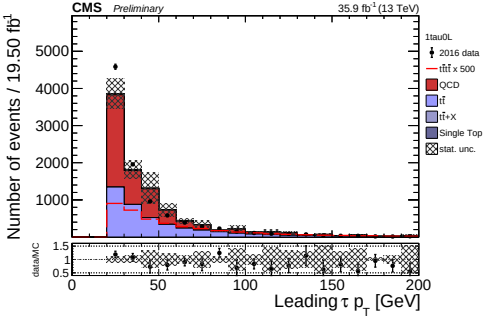
- **4% uncertainty on yield**

1tau0L (UL2016, MC QCD)




```
======== 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%
```

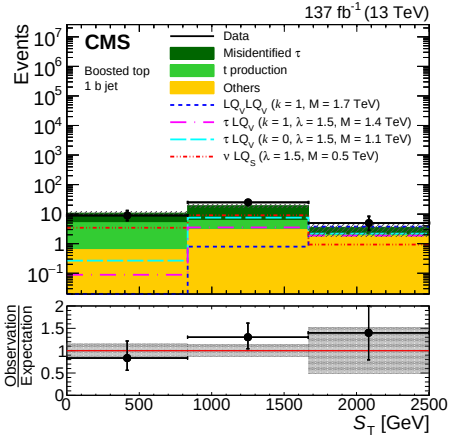
1tau0L (UL2016, DATA QCD)




```
======== 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%
```

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



Definition of the validation region

- As a **reminder**: we **compute fake rates in** the so-called **control region** (CR): same requirements as SR, but no b tagged jets
- I defined the **validation region (VR)** to be both close to CR and SR: same definition of SR but **exactly 1 b tagged jet**
- Orthogonal to both CR and SR
- Being orthogonal to SR, we can look at data here (not blinded)

	N_{τ_h}	N_ℓ	N_{jets}	N_{bjets}
CR	1	0	≥ 8	0
VR	1	0	≥ 8	1
SR	1	0	≥ 8	≥ 2

Definition of the validation region

- The **VR background composition is similar to the one in the SR**: lots of QCD, non-negligible $t\bar{t}$, some $t\bar{t}$+X

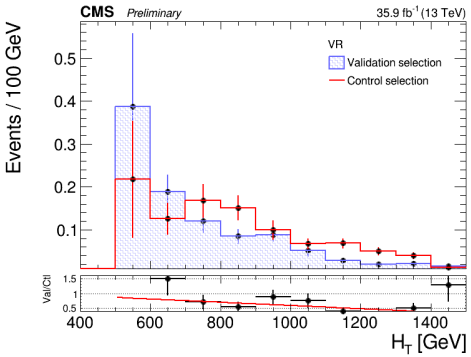
```
======== jets_HT 1tau0L_VR ====================
Data events:  10582
signal events: 0.644113
ttbar events: 1733.67
QCD events: 8739
tt+X events:   34.1828
single top events:   -0.0105877
total MC events: 10506.8
data/MC agreement: 0.715339%
```

QCD shape estimation: general idea

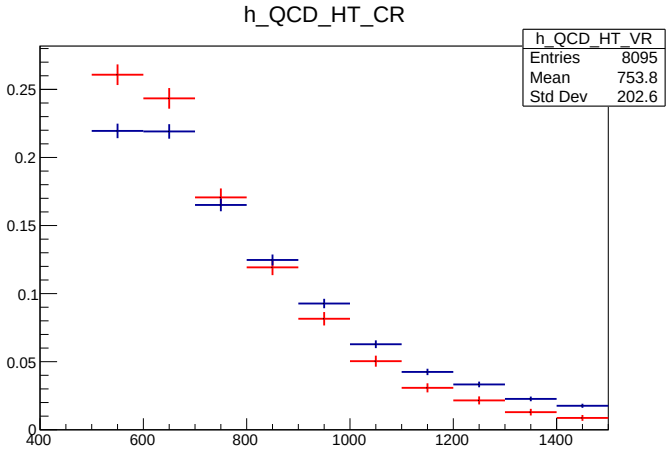
- First, we **need a QCD-dominated region** which is sufficiently close to the VR
  - **We have it already, it's the CR** used in the FR method
  - 96% QCD purity in the CR
- Take the **QCD shape from** the **CR in data**
- **Correct for kinematic differences between CR and VR using the simulation**
- Take the ratio of H_T shapes in VR and CR, fit it and **get a transition function from CR to VR**
- **Apply the transition function** to the data distribution in CR **to get the final shape in the VR**

Transition function

- **Just compare shapes**: normalize areas to 1
- Of course, QCD spikes are present here, so we cannot hope for a precise ratio
- Smoothen the ratio by **fitting with a straight line**
- This straight **transition factor is applied to the H_T distribution of data in the CR** to obtain the final shape

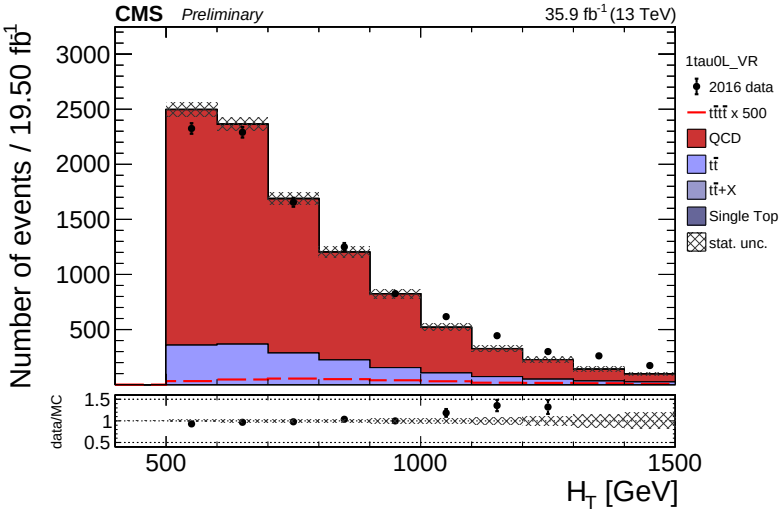


Corrected data shape

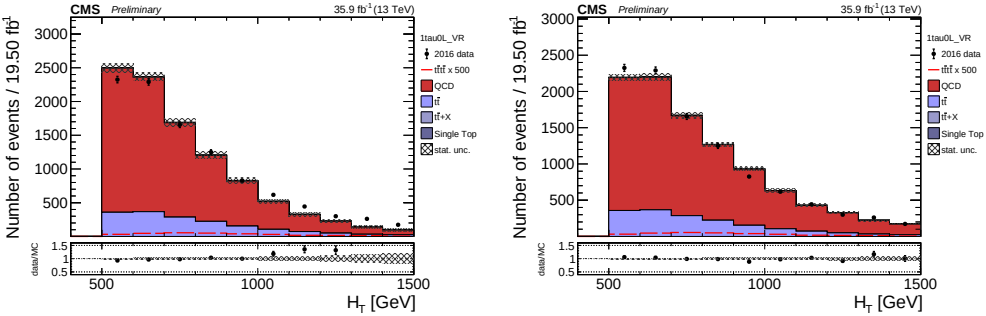


- Blue: H_T shape from data in CR; red: H_T shape from data in CR corrected

Validation of the FR method: QCD shape from data



Using the shape from CR (no correction)



- Left: corrected; right: uncorrected
- Agreement is better when using uncorrected shape!
- I wonder how reliable is the QCD simulation even for the TF computation