Uncertainties on the fake rate method
H_T distributions

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

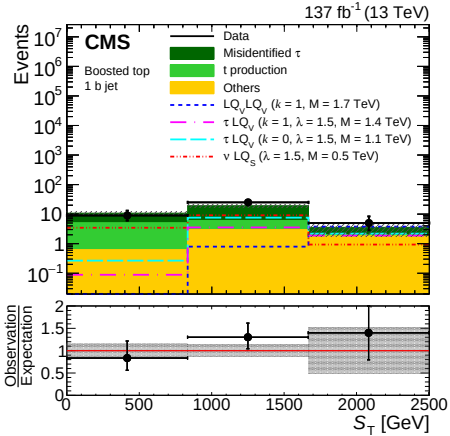
[1]Institute of High Energy Physics (IHEP), Beijing

[2]Fudan University, Shanghai

September 1, 2021

Uncertainties on FR method

- We are inspired by EXO-19-015
- Their idea is to perform **validation of the FR method in a region with similar background composition as the signal region**
- Validation is a **data/MC agreement** check on the variable they are going to use in final fit
- We follow a similar approach



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

	$t\bar{t}t\bar{t}$	$t\bar{t}$	QCD	$t\bar{t}$+X
CR	0.09	287.46	6051.20	8.17
VR	0.98	2321.43	7792.01	78.91
SR	8.79	5389.60	6539.06	162.25

- It looks fine to perform validation in this region
- **Compute the QCD yield expected by the FR method in the VR**

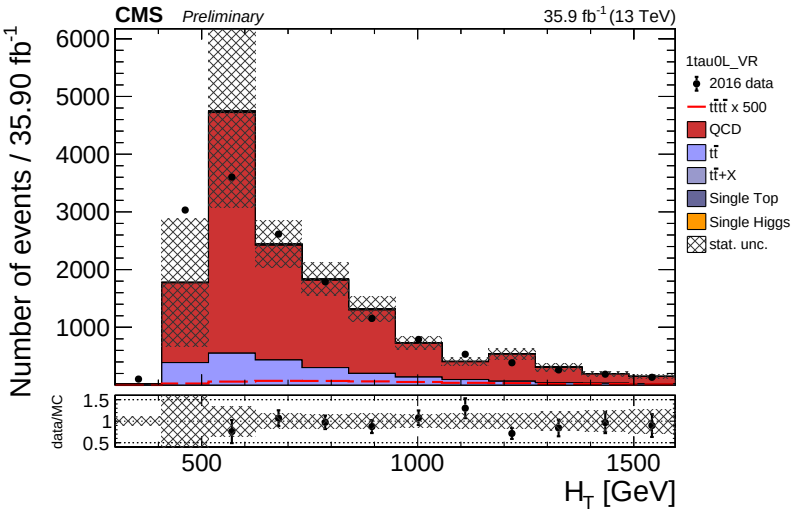
	MC QCD yield	FR QCD yield
exp. yield	7792	12392

Validation of the FR method

- Assumed we are going to fit H_T distribution in this category
  - We don't have a BDT here
- Perform **data/MC agreement for H_T** distribution in the VR
- **Scale the MC QCD shape to yield coming from FR method**
- Interestingly, using the FR yield **enhances the data/MC agreement**:

	MC QCD yield	FR QCD yield
data/MC	28%	0.2%

Validation of the FR method



Remarks on validation procedure

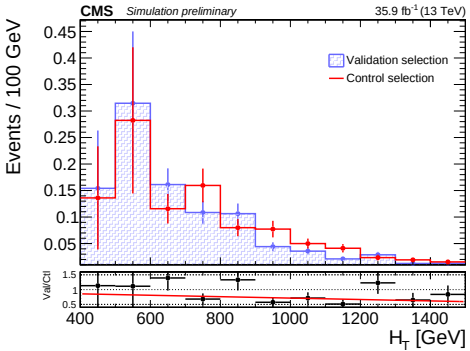
- Based on previous slide agreement, **we should assess the uncertainty on this method**
- I propose to assign **two uncertainties** in the datacard
  - **One** log-normal unc. of $\approx 4\%$ **for the statistical uncertainty** on the yield
  - **One** log-normal unc. of some value **for the above level of agreement**
- **MC QCD spikes make it hard to decide the level of agreement**
- Binning in EXO-19-015 is pretty coarse, rebinning could work but I don't like the idea so much
- **Try to get the shape of QCD from data as well**
  - Statistics would be increased a lot

QCD shape estimation: general idea

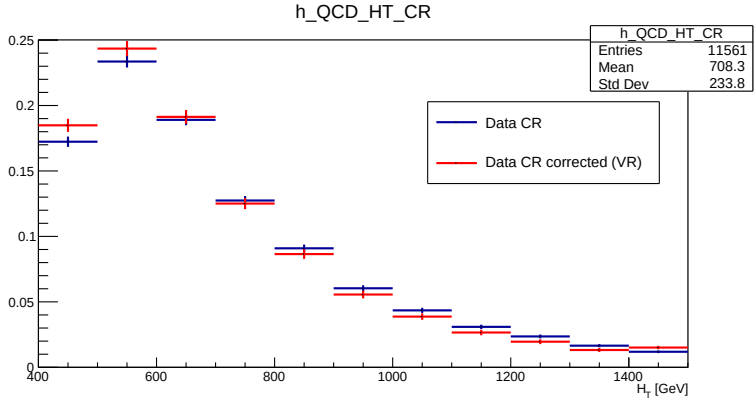
- First, we **need a QCD-dominated region** which is sufficiently close to the SR
  - **We have it already, it's the CR** used in the FR method
  - 96% QCD purity in the CR
- Take the **QCD shape from** the **CR in data**
- **Correct for kinematic differences between CR and VR using the simulation**
- 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

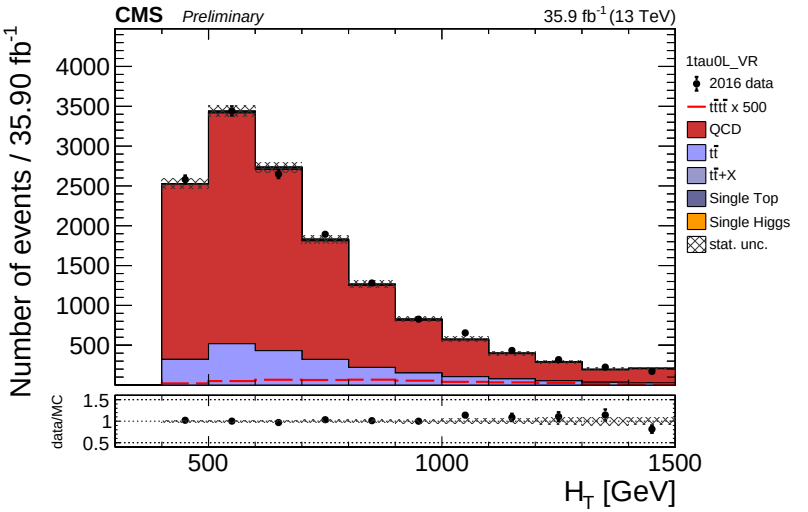


Corrected data shape



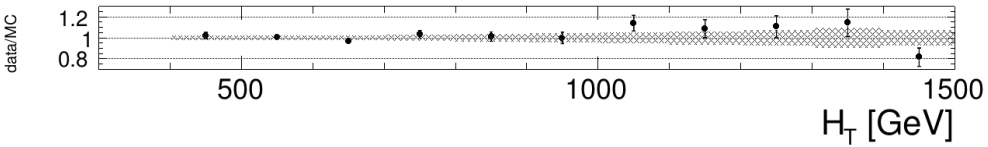
- Blue: H_T shape from data in CR; red: H_T shape from data in CR corrected with CRtoVR transition function

Validation of the FR method: QCD shape from data



Validation of the FR method: QCD shape from data

- **What level of uncertainty should we assess** for this procedure?
- By a closer look at the ratio plot, we see that **none of the points disagrees by more than 20%**
- Actually, all of them are compatible with one except for $1000 < H_T < 1100$ GeV and $1400 < H_T < 1500$ GeV bins
- Given that some degree of uncertainty also comes from the shape estimation, I would say that assigning a **15% uncertainty on the QCD estimation looks fair** (and maybe conservative)
- **Room for discussion here**

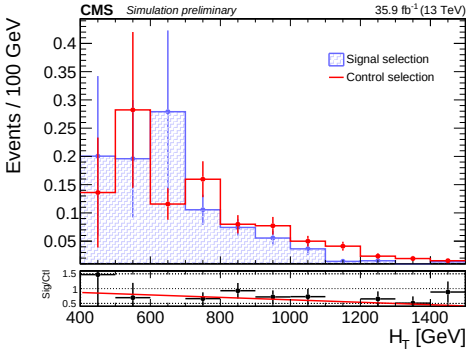


H_T distributions

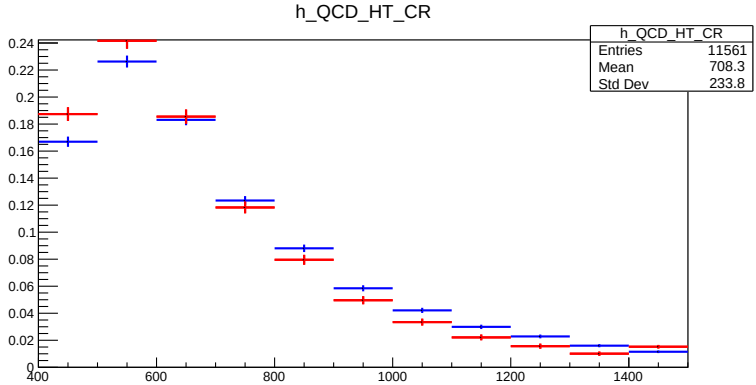
- In categories where we didn't train a BDT, we **plan to fit H_T distributions**
- **Check** the distributions to see **if** this variable **really separates signal from backgrounds**
- Of course **do not plot data here: we are blinded!**
- **1tau0L** has a special treatment. **Estimate QCD shape in the SR with identical method as for the VR** (see following slide)

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

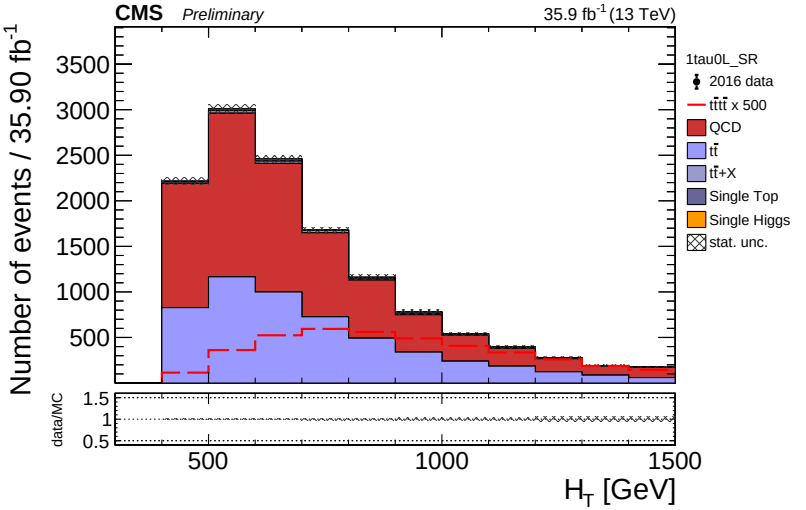


Corrected data shape

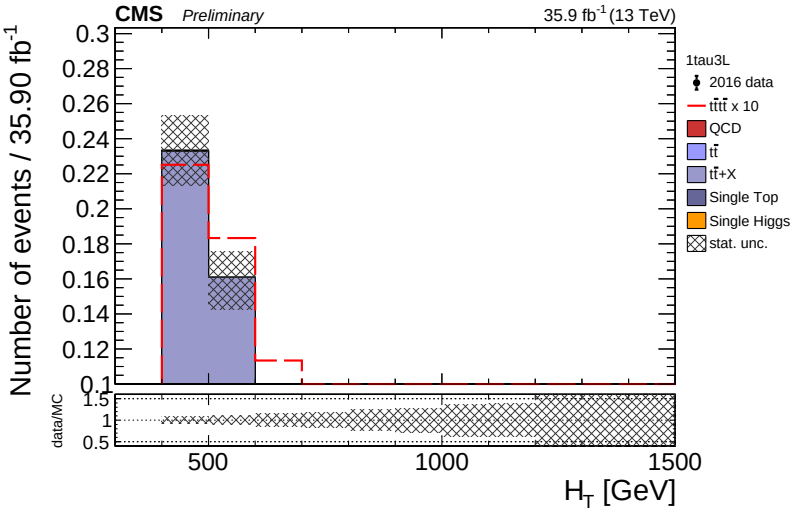


- Blue: H_T shape from data in CR; red: H_T shape from data in CR corrected with CRtoVR transition function

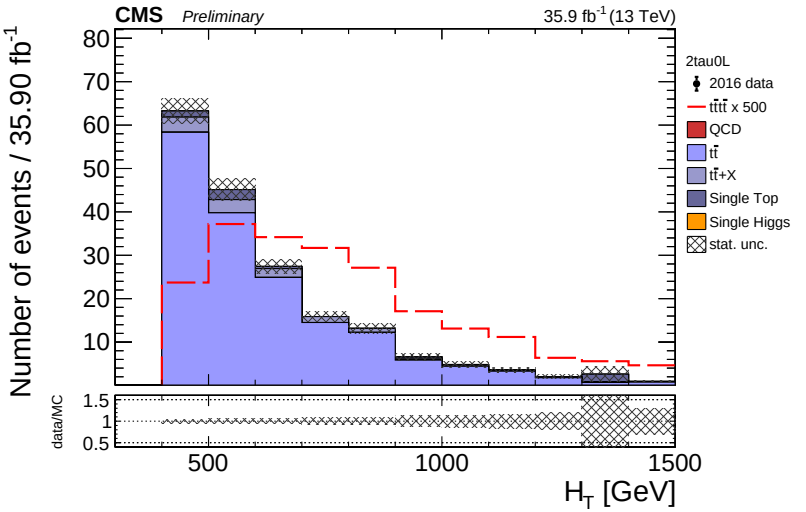
H_T distributions: 1tau0L



H_T distributions: 1tau3L



H_T distributions: 2tau0L



H_T distributions: 2tau2L