Uncertainties on the QCD estimation in 1tau0L category

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 8, 2021

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%

Remarks on validation procedure

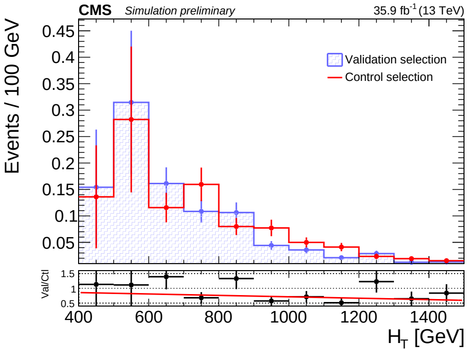
- MC QCD spikes make it hard to decide the level of agreement
- **Try to get the shape of QCD from data as well**
  - Statistics would be increased a lot

QCD shape estimation: general idea

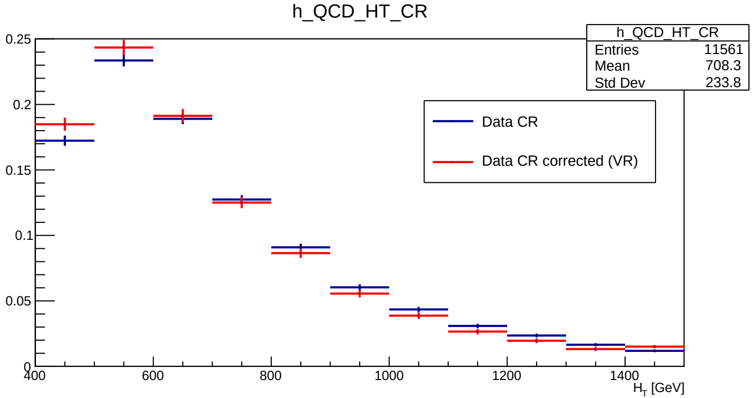
- First, we **need a QCD-dominated region** which is sufficiently close to the SR
  - **We have it already, it's the CR** used in the FR method
  - 96% QCD purity in the CR
- Take the **QCD shape from** the **CR in data**
- **Correct for kinematic differences between CR and VR using the simulation**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

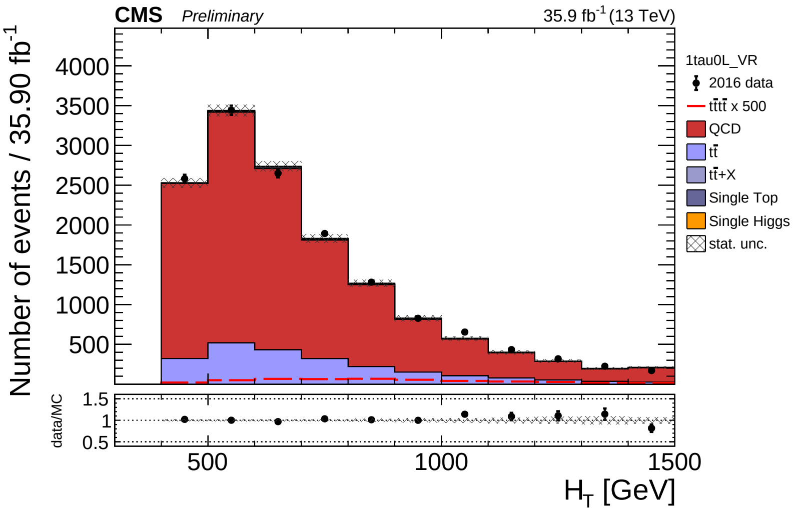


Corrected data shape



- 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



QCD uncertainties in 1tau0L

F. Iemmi

Uncertainties on FR method

HT in SR

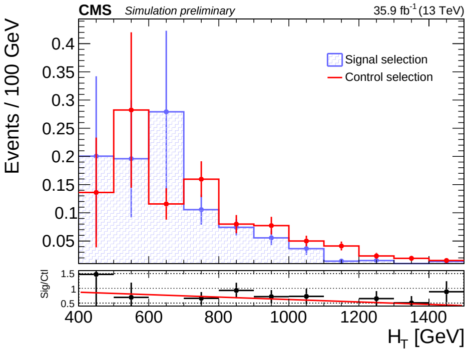
Uncertainties on QCD shape

H_T distribution in SR

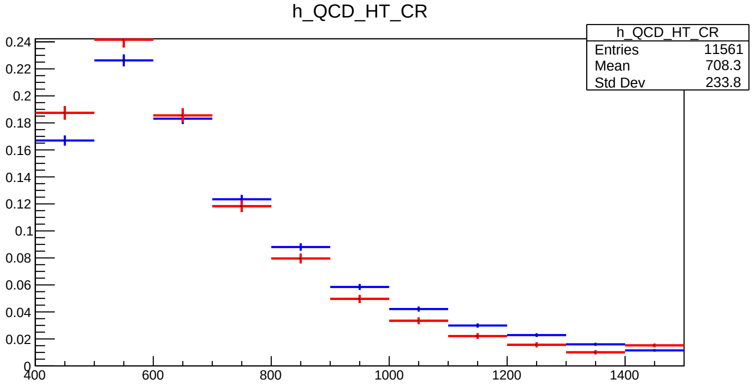
- **Estimate QCD shape in the SR with identical method as for the VR** (see following slide)
- Of course **do not plot data here**: **we are blinded!**

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

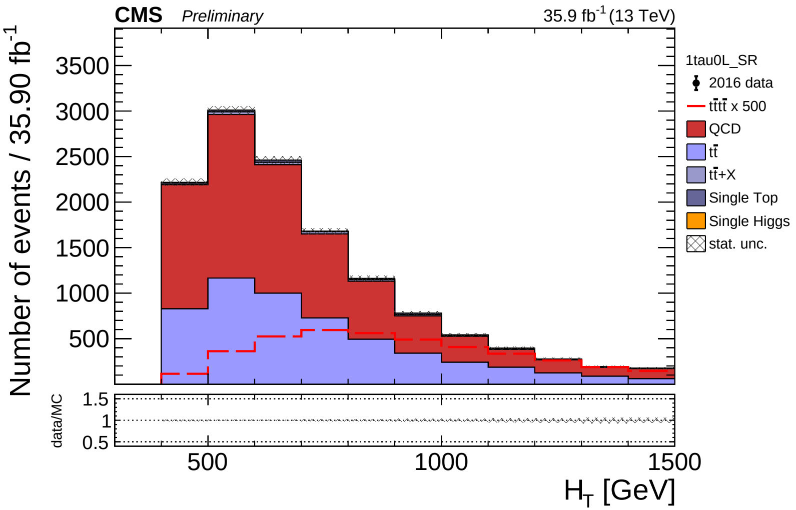


Corrected data shape



- 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



Uncertainties on the QCD shape

- QCD H_T shapes are taken from the CR in data and translated to VR or SR **using** the corresponding **transition functions (TFs)**
- TFs are the result of a fit: ROOT gives you the **fitted parameters and the correlation matrix** $\mathcal{V}$ of the fit
- In our case, we fitted with straight lines of the form

$$y = mx + q,$$

so the correlation matrix will look like

$$\mathcal{V} = \begin{bmatrix} \sigma_q^2 & \rho_{qm} \\ \rho_{mq} & \sigma_m^2 \end{bmatrix},$$

where $\sigma^2_{q/m}$ are the variances of the parameters and $\rho_{qm} = \rho_{mq}$ are the correlation coefficients between m and q

Uncertainties on the QCD shape

- In general, $\rho_{qm} = \rho_{mq} \neq 0$, i.e., some degree of correlation exists between the two parameters
- This means **one cannot shift** m **and** q **up and down independently**
- $\mathcal{V}$ is a real, symmetric matrix $\implies$ it can always been diagonalized by means of an orthogonal transformation
- This means **it exists some auxiliary parameter space in which** m **and** q **are fully decorrelated**
  - One can shift them up/down independently in this space
- Linear algebra theorem: the orthogonal diagonalizing matrix $\mathcal{O}$ has the eigenvectors of $\mathcal{V}$ as columns

$$\mathcal{D} = \mathcal{O}^{-1}\mathcal{V}\mathcal{O} = \mathcal{O}^{T}\mathcal{V}\mathcal{O}$$

Uncertainties on the QCD shape

- **Idea**: Starting from the "real" parameters, described by the vector $\mathbf{p}^T = (q, m)$, we first transform them to some auxiliary parameters $\tilde{\mathbf{p}}^T = (\tilde{q}, \tilde{m})$:

$$\tilde{\mathbf{p}} = \mathcal{O}\mathbf{p}$$

- In the auxiliary space, the correlation matrix is diagonal and its non-zero elements are the variances of $\tilde{m}, \tilde{q}$

$$\mathcal{D} = \begin{bmatrix} \tilde{\sigma}^2_{\tilde{q}} & 0 \\ 0 & \tilde{\sigma}^2_{\tilde{m}} \end{bmatrix}.$$

- Now the parameters can be shifted independently, so we define the shifted TFs in the auxiliary space to be described by

$$\tilde{\mathbf{p}}^T_{\text{up}} = (\tilde{q} + \tilde{\sigma}_{\tilde{q}}, \tilde{m} + \tilde{\sigma}_{\tilde{m}})$$
$$\tilde{\mathbf{p}}^T_{\text{down}} = (\tilde{q} - \tilde{\sigma}_{\tilde{q}}, \tilde{m} - \tilde{\sigma}_{\tilde{m}})$$

Uncertainties on the QCD shape

- Finally, we perform the **inverse transformation** to go back and get the parameters describing the **TFs in the original space**

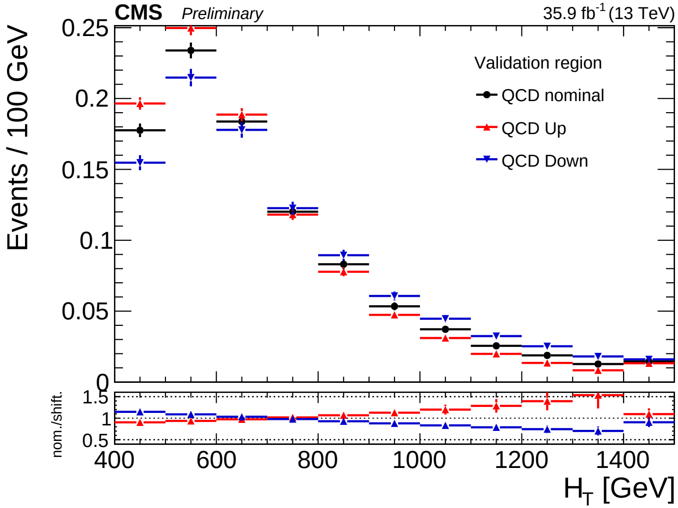
$$\mathbf{p}_{\text{up}} = \mathcal{O}^{-1}\tilde{\mathbf{p}}_{\text{up}}$$
$$\mathbf{p}_{\text{down}} = \mathcal{O}^{-1}\tilde{\mathbf{p}}_{\text{down}}$$

- Now compare nominal shapes with the upwards/downwards shifted shapes
- Scale all areas to one: we are **interested in the shape differences**
  - The yield will be coming from FR for all of them
- **The upwards/downwards shapes are what Combine needs to implement shape uncertainties**

Validation region

	q	m
Nominal	0.95	-0.00024
Up	1.31	-0.00064
Down	0.59	0.00016

QCD shape uncertainty: VR



Signal region

	q	m
Nominal	1.034	-0.00041
Up	1.42	-0.00083
Down	0.65	-1.3572e-06

QCD shape uncertainty: SR