Universität Hamburg
DER FORSCHUNG | DER LEHRE | DER BILDUNG

Measurement of top tagging efficiencies

Dennis Schwarz

February 19, 2020

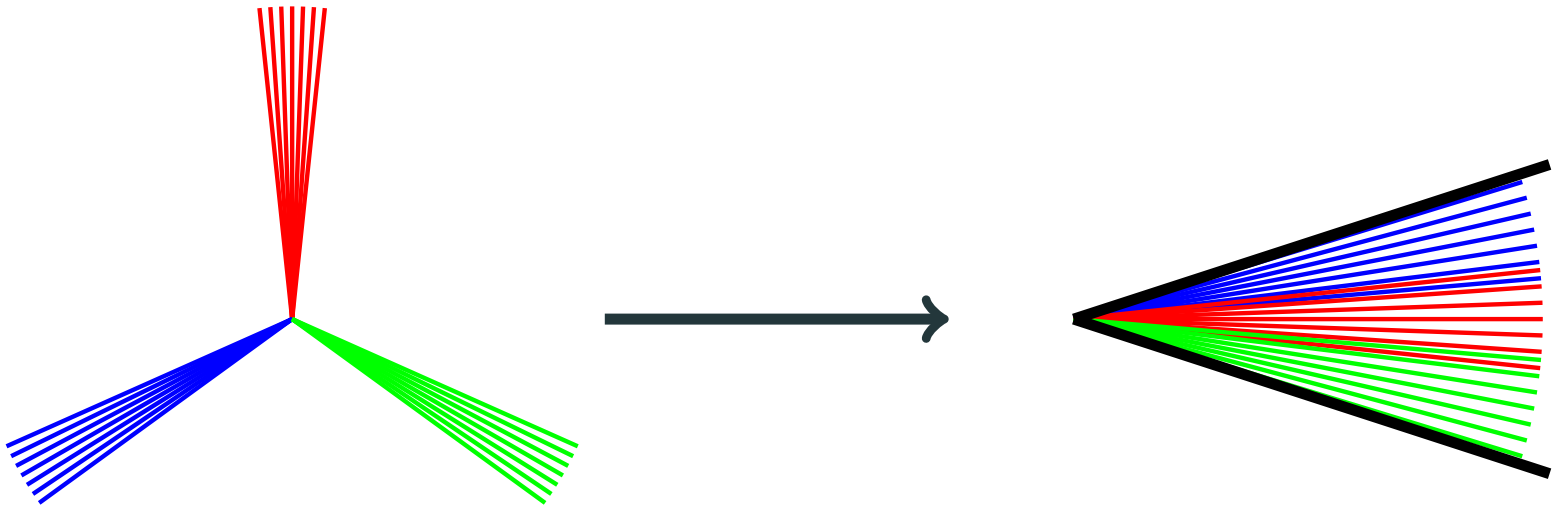
Group meeting

Why top tagging?

- Abundant production of high-p_T top quarks at LHC
- High-p_T tops are often part of searches or measurements
- Hadronic top decays are highly collimated

$\rightarrow$ reconstruction with one jet

$\rightarrow$ Important to identify those jets



resolved boosted

Which jets to use?

[Lapsien et al., Eur. Phys. J. C 76, 600 (2016)]

Large radius jets to include full top decay

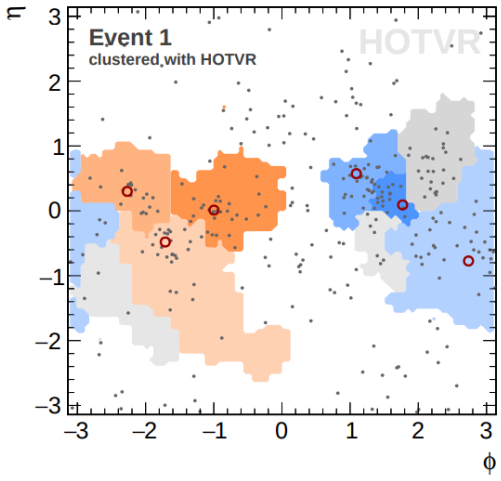
AK8

- Constant radius of $R = 0.8$
- Standard in CMS

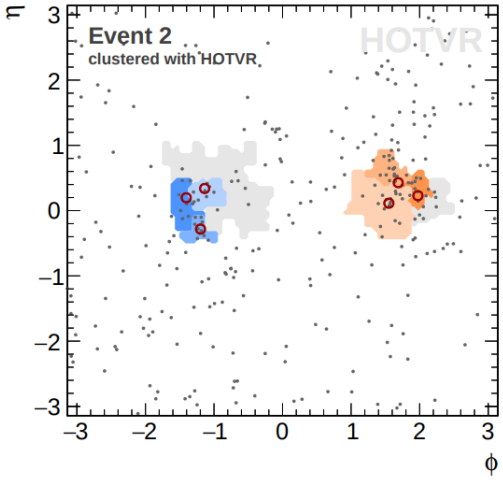
HOTVR

- Adapt radius with $R = \frac{\rho}{p_T}$
- Mass jump criterion
  - $\rightarrow$ suppressing soft radiation
  - $\rightarrow$ defining subjets

low p_T



high p_T

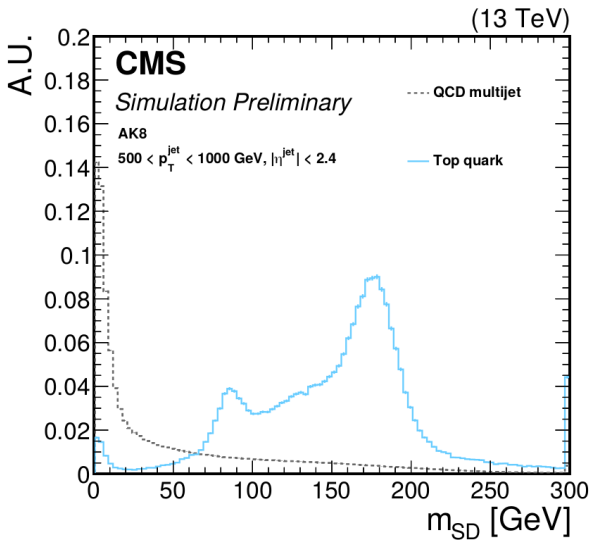


How are jets top-tagged?

[CMS-PAS-JME-18-002, Thaler et al., JHEP 1103:015,2011]

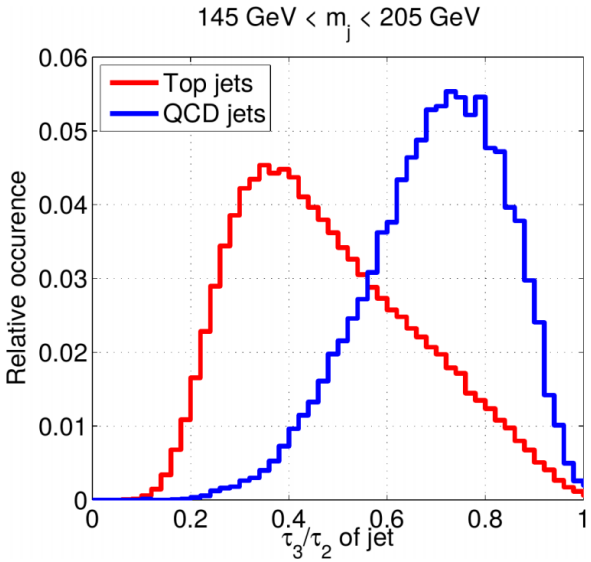
Identification of top jets with substructure

- Jet mass (soft drop grooming)
- N-subjettiness ratio $\tau_{32} = \tau_3/\tau_2$



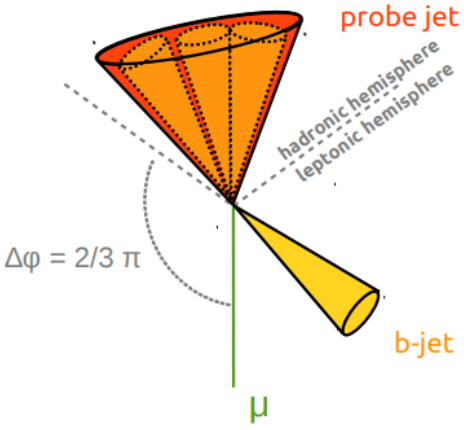
Additional requirements for HOTVR

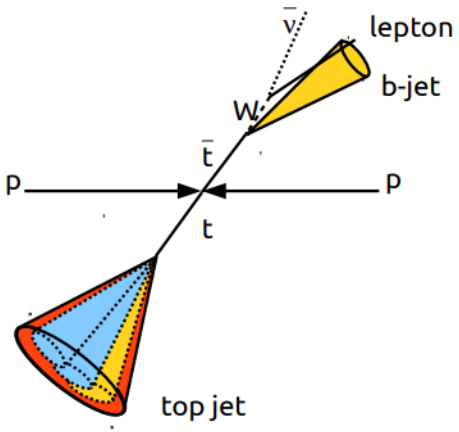
- Number of subjets
- p_T fraction of leading subjet
- Minimal pairwise mass of two subjets



Measurement strategy

- tag and probe method
- μ+jets $t\bar{t}$ events
- leptonic decay leg as tag





- high p_T muon
- b-tag in leptonic hemisphere
- probe jet: AK8/HOTVR jet in hadronic hemisphere

Goal: Provide p_T dependent data-to-simulation scale factors

Overview

Jet collection	Working points
AK8 PUPPI	10 (5 with + 5 without subjet b tagging)
AK8 CHS	8 (4 with + 4 without subjet b tagging)
HOTVR	1

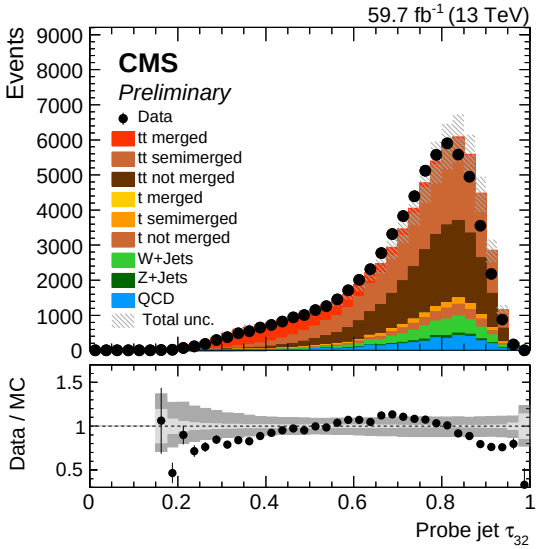
→ 19 working points per year

→ Provide scale factors for 57 working points in total!

Disclaimer: In this talk I will focus on AK8 PUPPI in 2018

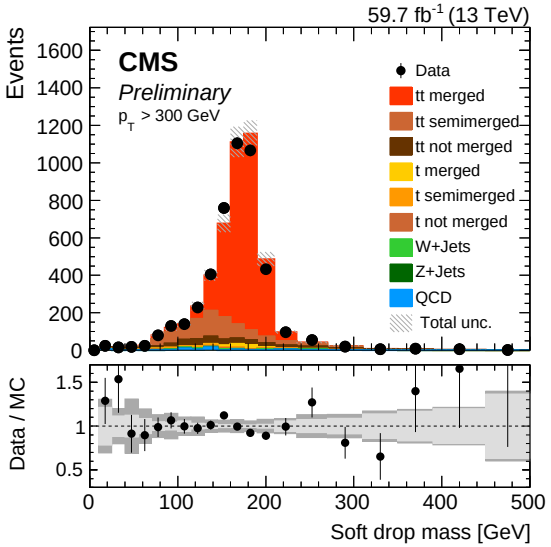
Measurement of efficiencies

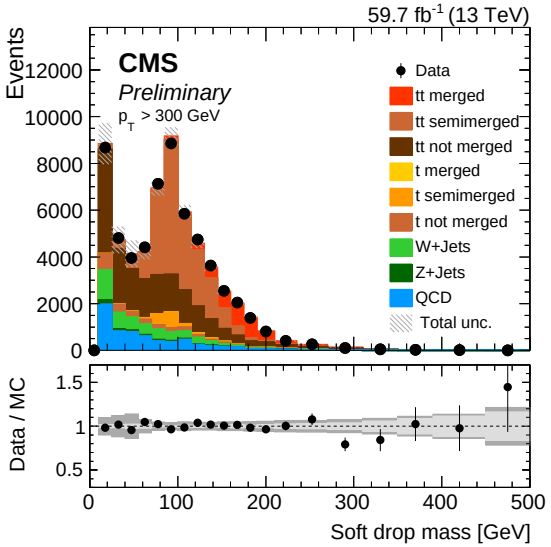
1. Split $t\bar{t}$ in merged, semimerged, not merged
2. Define pass and fail regions (per WP)
   - AK8: τ_{32} cut
   - HOTVR: $N_{\text{subjets}} \geq 3$, $f < 0.8$, $m_{ij} > 50$ GeV, τ_{32} cut



Measurement of efficiencies

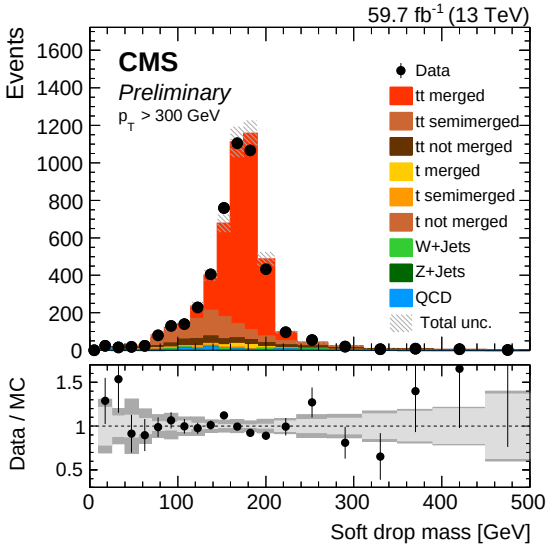
1. Split $t\bar{t}$ in merged, semimerged, not merged
2. Define pass and fail regions (per WP)
   - AK8: τ_{32} cut
   - HOTVR: $N_{\text{subjets}} \geq 3$, $f < 0.8$, $m_{ij} > 50$ GeV, τ_{32} cut
3. Fit jet mass in pass and fail

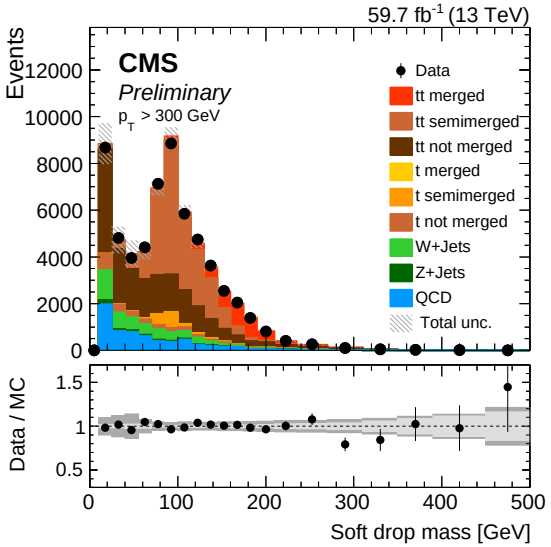




Measurement of efficiencies

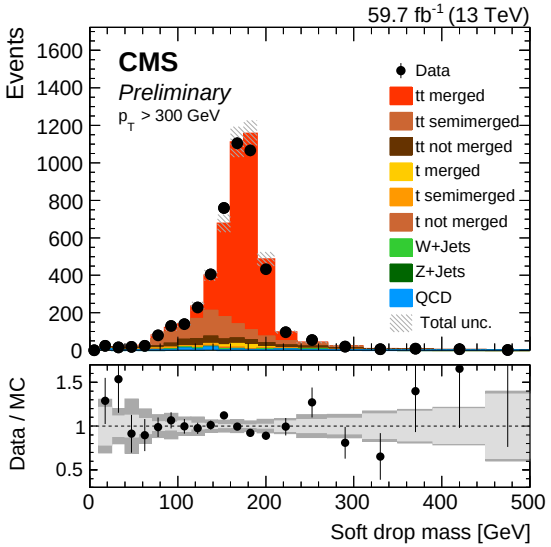
1. Split $t\bar{t}$ in merged, semimerged, not merged
2. Define pass and fail regions (per WP)
   - AK8: τ_{32} cut
   - HOTVR: $N_{\text{subjets}} \geq 3$, $f < 0.8$, $m_{ij} > 50$ GeV, τ_{32} cut
3. Fit jet mass in pass and fail
4. Get efficiencies $\epsilon_{\text{tag}} = \frac{N_{\text{pass}}}{N_{\text{pass}} + N_{\text{fail}}}$
   - $\epsilon_{\text{tag}}^{\text{MC}}$ from MC **pre** fit
   - $\epsilon_{\text{tag}}^{\text{data}}$ from MC **post** fit

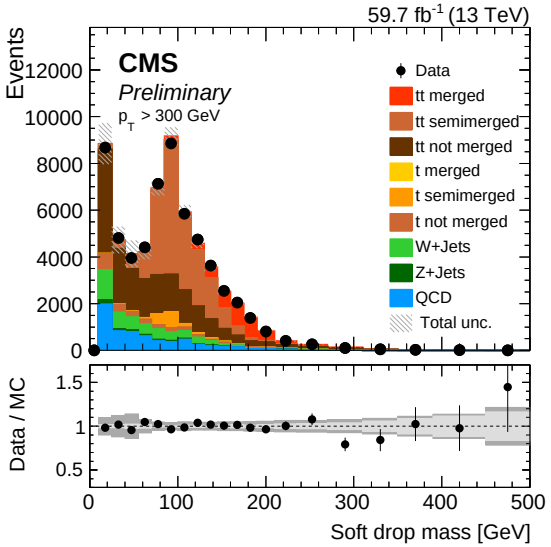




Measurement of efficiencies

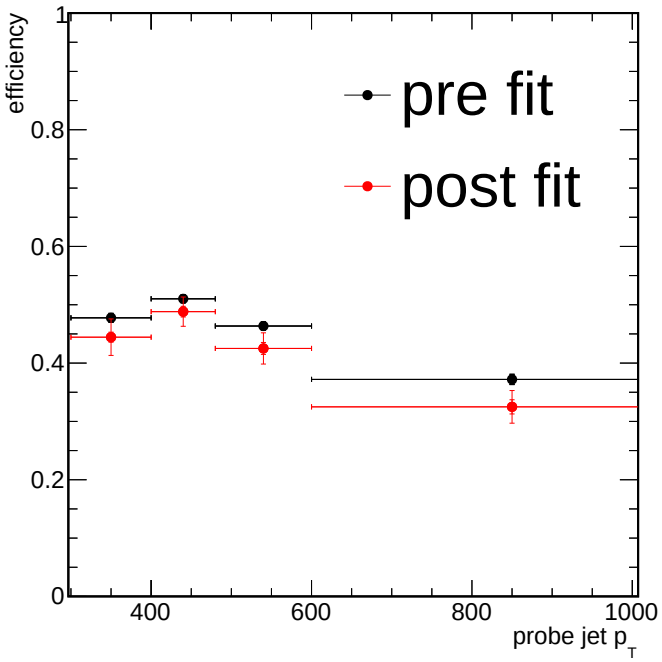
1. Split $t\bar{t}$ in merged, semimerged, not merged
2. Define pass and fail regions (per WP)
   - AK8: τ_{32} cut
   - HOTVR: $N_{\text{subjets}} \geq 3$, $f < 0.8$, $m_{ij} > 50$ GeV, τ_{32} cut
3. Fit jet mass in pass and fail
4. Get efficiencies $\epsilon_{\text{tag}} = \frac{N_{\text{pass}}}{N_{\text{pass}} + N_{\text{fail}}}$
   - $\epsilon_{\text{tag}}^{\text{MC}}$ from MC **pre** fit
   - $\epsilon_{\text{tag}}^{\text{data}}$ from MC **post** fit
5. Get efficiency of mass window from cut and count $\epsilon_{\text{mass}} = \frac{N_{140 < m_{\text{jet}} < 220}}{N_{\text{all}}}$



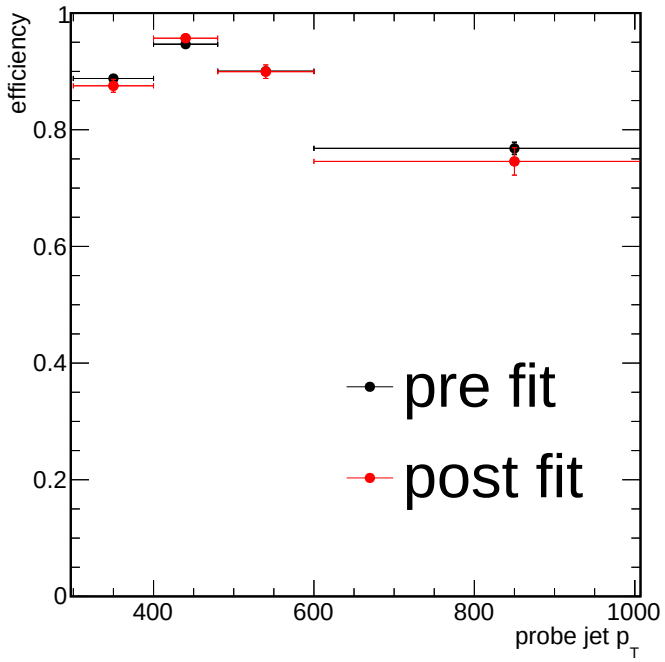


Calculate scale factors

Efficiency τ_{32}



Efficiency mass cut

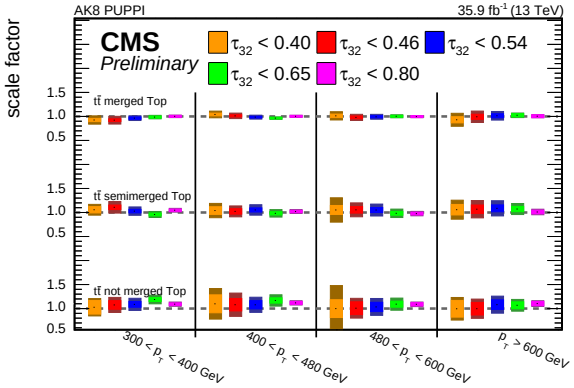


6. Extract $\epsilon_{\text{tot}} = \epsilon_{\text{tag}} \cdot \epsilon_{\text{mass}}$
7. Calculate data-to-simulation scale factor:

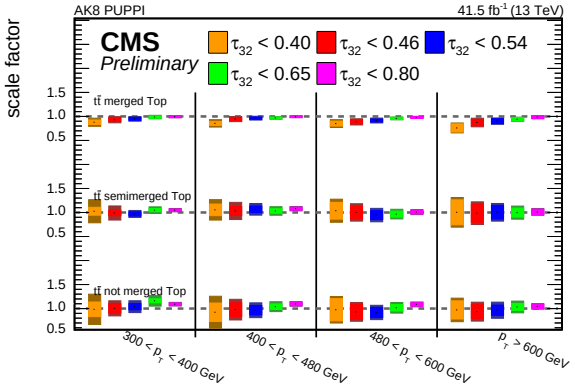
$$\text{SF} = \frac{\epsilon_{\text{tot}}^{\text{postfit}}}{\epsilon_{\text{tot}}^{\text{prefit}}}$$

AK8 scale factors

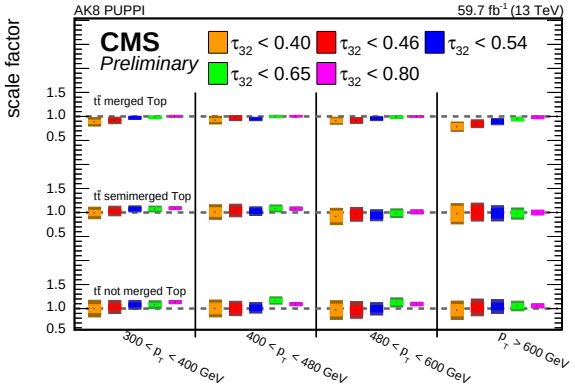
2016



2017

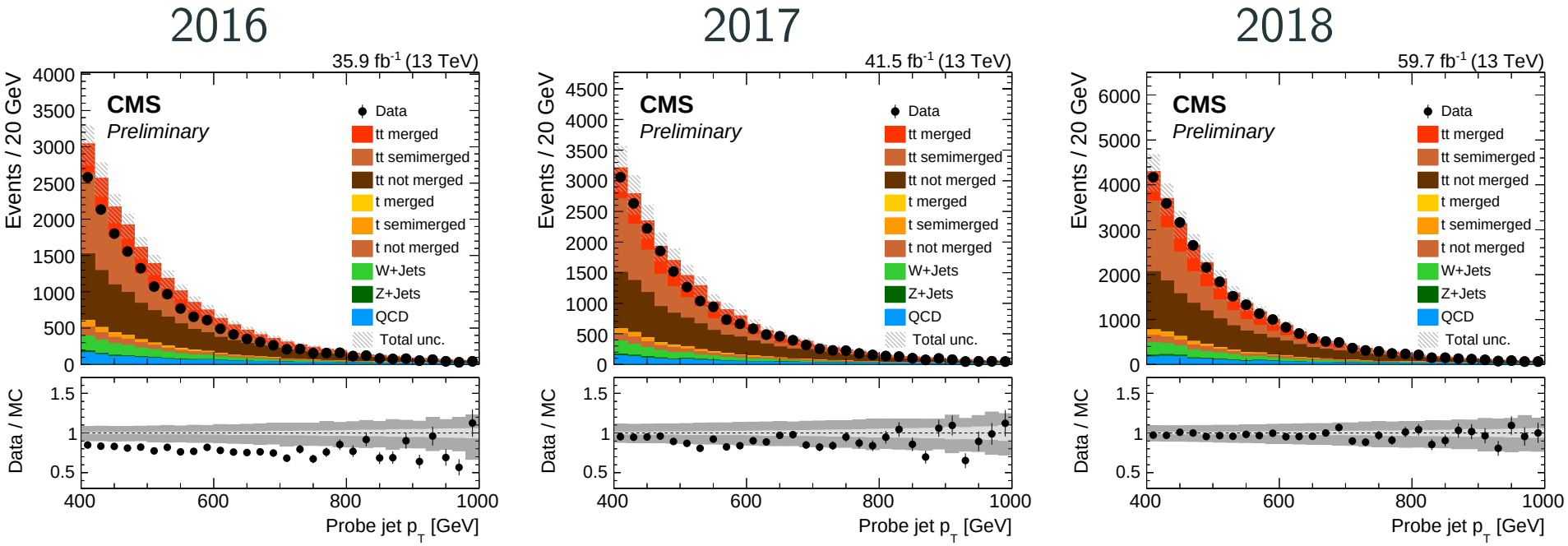


2018



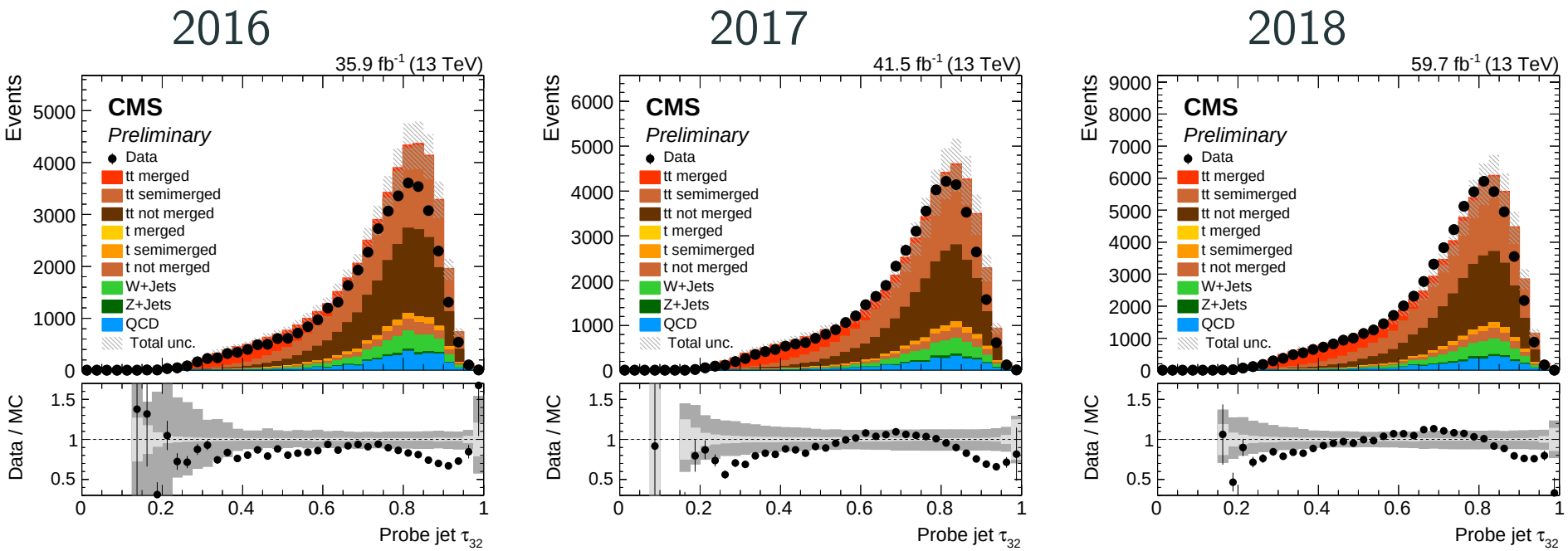
- Scale factors measured for full Run 2
- Split into merged, semimerged, not merged
- Working points defined by τ_{32} cut
- We also provide a set of scale factors with additional subjet b tagging applied (not shown here)
- Differences between the years can further be investigated

Differences between the years



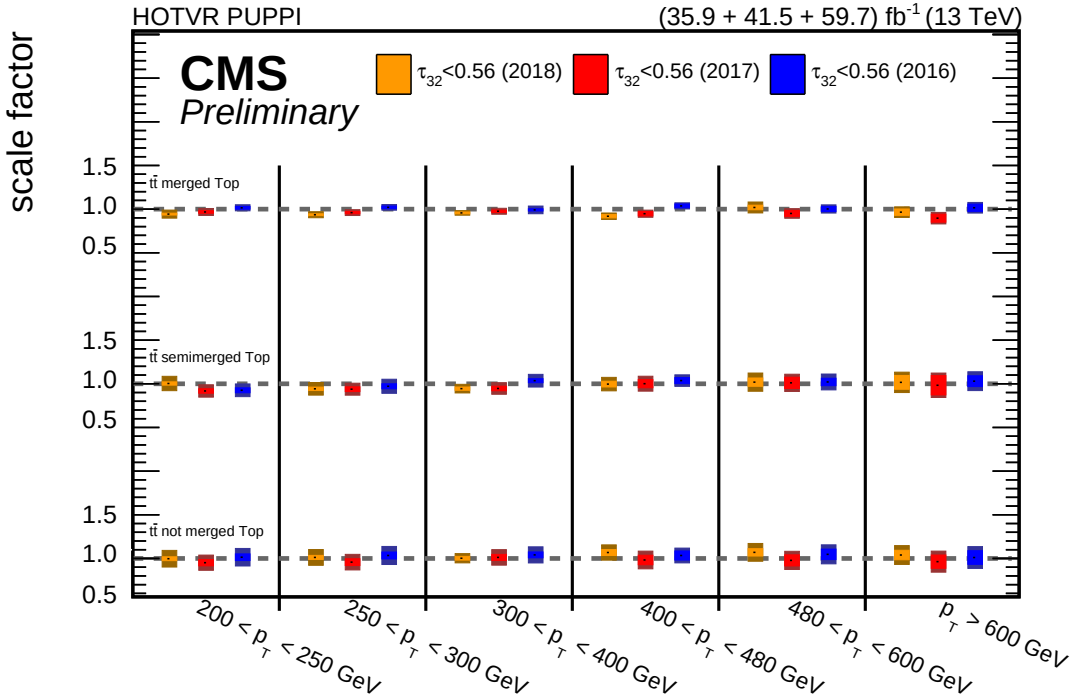
- Data/MC mostly flat against p_T
- Normalization off in 2016
  (changed with new tune in 2017 and 2018)
- Very good prediction in 2018

Differences between the years



- Worse τ_{32} description in 2017 and 2018
- New $t\bar{t}$ tune responsible
- Leading to smaller SF in 2017 and 2018 (2016 close to 1)

HOTVR scale factors



- Only one working point for HOTVR (here: all 3 years)
- HOTVR possible for smaller p_T
- SF very close to 1 and flat

Summary

- Top tagging is an important tool for a variety of searches and measurements
- Measurement of p_T dependent top tagging scale factors
- Wide range of working points covered
- Targeting publication of plots in a DP note (in progress)