Tagging Boosted Top Quarks and Higgs Bosons in ATLAS

BOOST 2015, University of Chicago

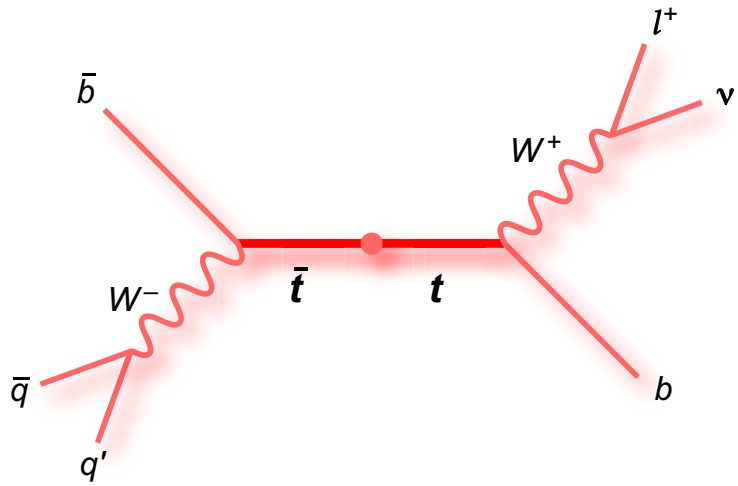
Matt LeBlanc, on behalf of the ATLAS Collaboration
10^{th}-14^{th} August 2015

Overview

- Run-I Top Tagging: ATLAS-CONF-2015-36
  - Tagging Techniques
  - Data/MC Comparisons
  - Efficiency and Fake Rate Measurements in Data
  - Performance Comparison in MC
- Run-II $H \to b\bar{b}$ Tagging: ATL-PHYS-PUB-2015-035
  - Developing the tagger
  - Systematics
- W tagging and W/Z discrimination covered by Julien!

Top Tagging Methods

- Comparison of different ATLAS Run I top taggers:
  - Substructure Cut-Based[1]
  - HEPTopTagger[2]
  - Shower Deconstruction[3]
  - HTT04

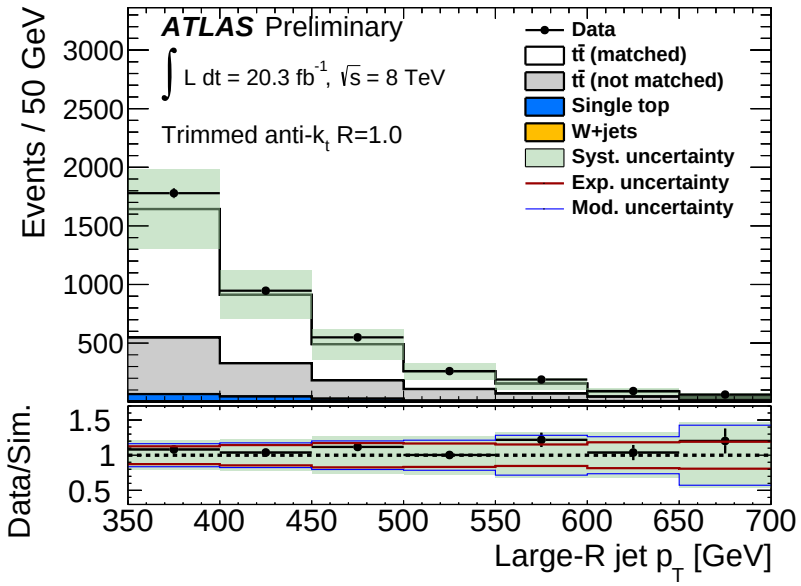


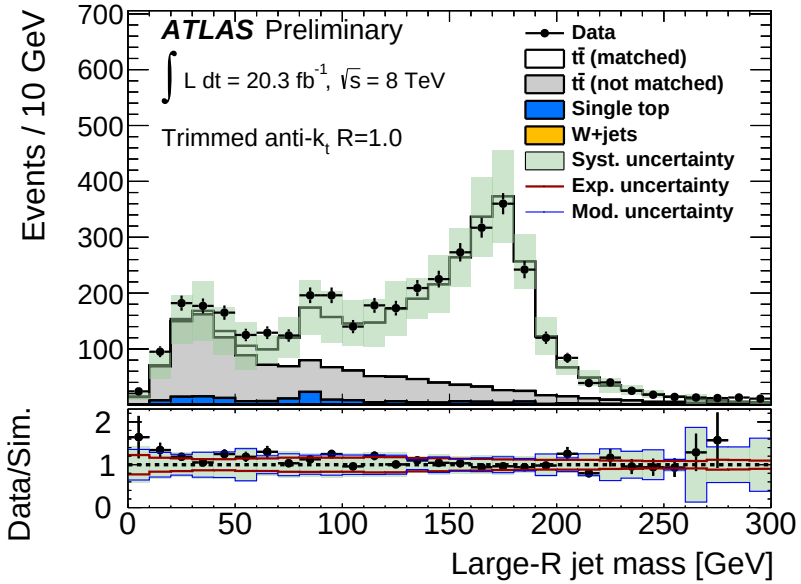
tagger	jet algorithm	grooming	radius parameter	p_T range	$\lvert\eta\rvert$ range
Tagger I–V W' top tagger Shower Deconstruction	anti-k_t	trimming	$R = 1.0$	> 350 GeV	< 2
Shower Deconstruction	C/A	none	$R = 1.2$	> 350 GeV	< 2
HEPTopTagger	C/A	none	$R = 1.5$	> 200 GeV	< 2

1. N-Subjettiness; Thaler & Van Tilberg
2. Plehn, Salam, Spannowsky, Takeuchi et al.
3. Soper & Spannowsky

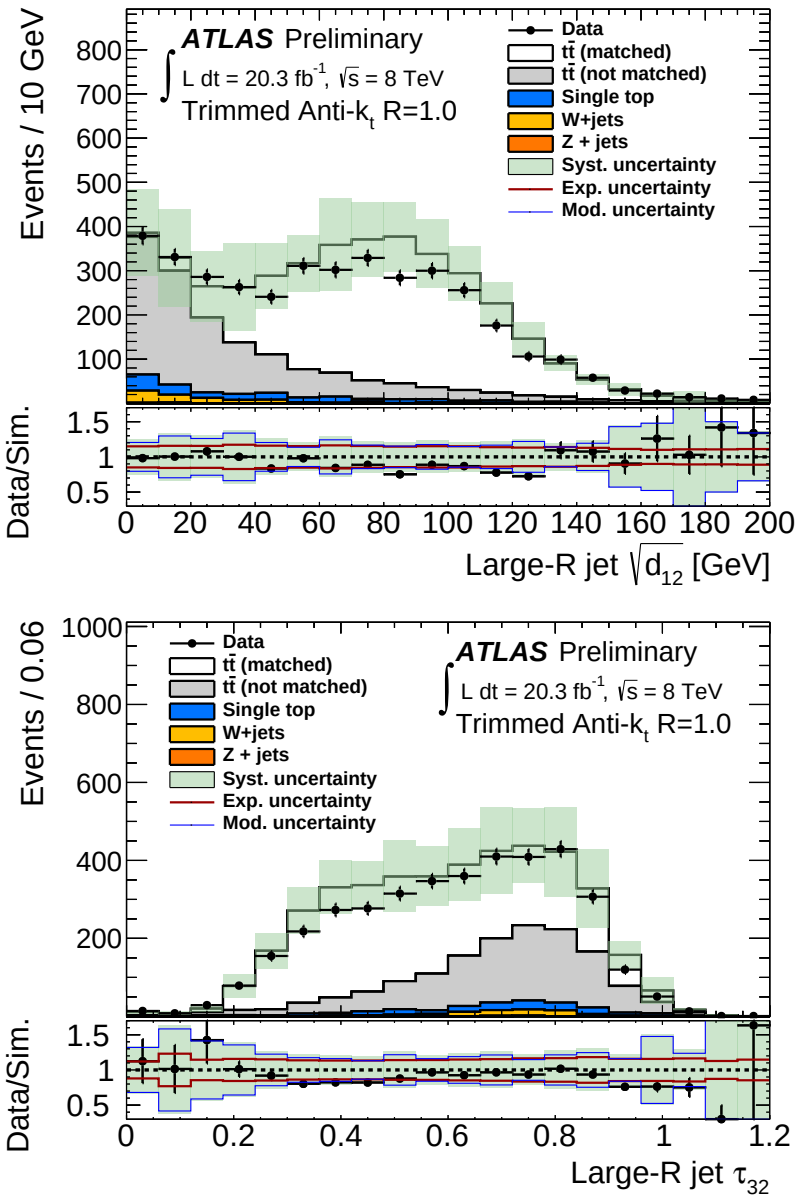
Selection

- Lepton+jets $t\bar{t}$ selection.
  - 2 b-tagged jets.
- Try tagging large-R jet away from lepton.
- Decompose $t\bar{t}$ into truth-matched (signal) and non-matched (background) components.

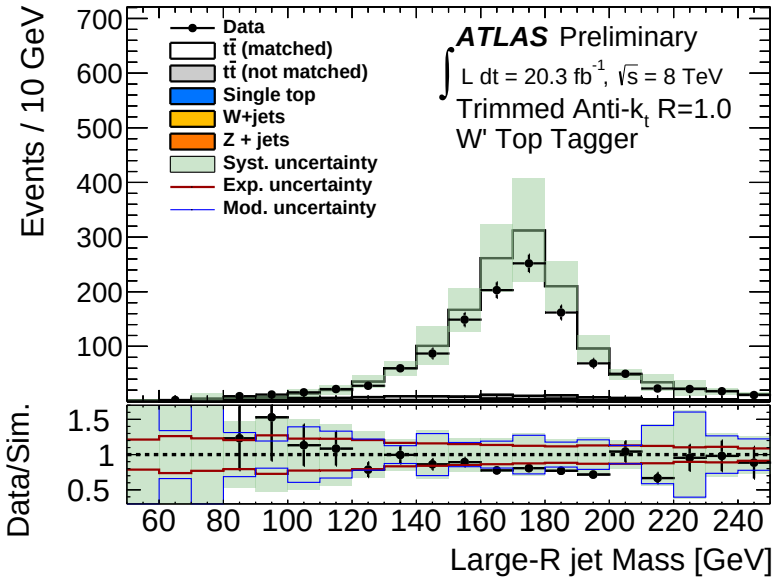




Substructure-Based Top Tagging

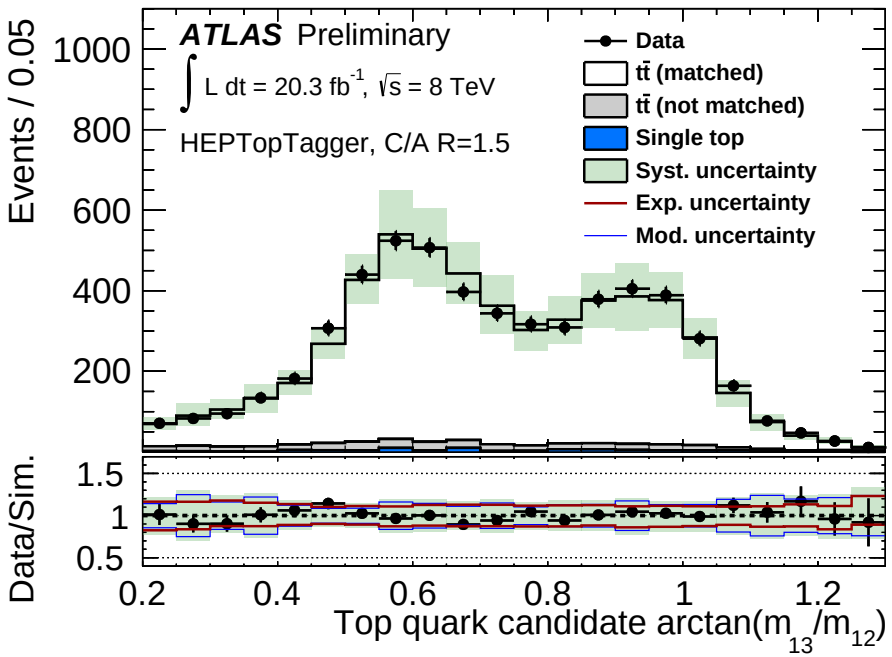


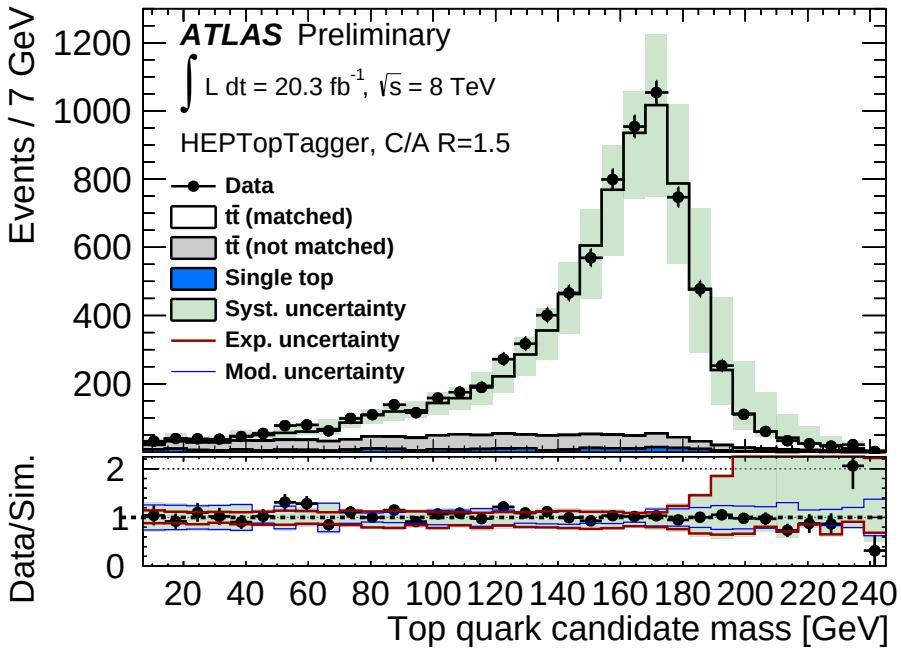
tagger	top tagging criterion
Substructure tagger I	$\sqrt{d_{12}} > 40$ GeV
Substructure tagger II	$m > 100$ GeV
Substructure tagger III	$m > 100$ GeV and $\sqrt{d_{12}} > 40$ GeV
Substructure tagger IV	$m > 100$ GeV and $\sqrt{d_{12}} > 40$ GeV and $\sqrt{d_{23}} > 10$ GeV
Substructure tagger V	$m > 100$ GeV and $\sqrt{d_{12}} > 40$ GeV and $\sqrt{d_{23}} > 20$ GeV
W' top tagger	$\sqrt{d_{12}} > 40$ GeV and $0.4 < \tau_{21} < 0.9$ and $\tau_{32} < 0.65$



HEPTopTagger

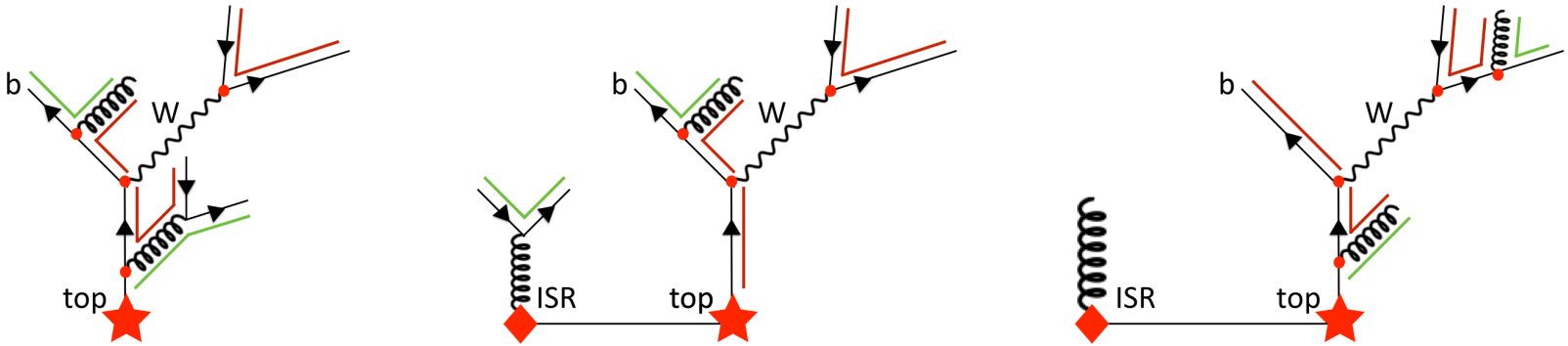
- HTT explained elsewhere (cf. Torben's talk later today, and the CMS talk by Justin from yesterday.).
- Parameters used here: $m_{\text{cut}} = 50$ GeV, $N_{\text{filt}} = 5$, $R_{\text{max}^{\text{filt}}} = 0.25$, $f_W = m_W \times (1 \pm 0.15)$.
- Operates at lower p_T than other taggers (C/A $p_T > 200$ GeV).



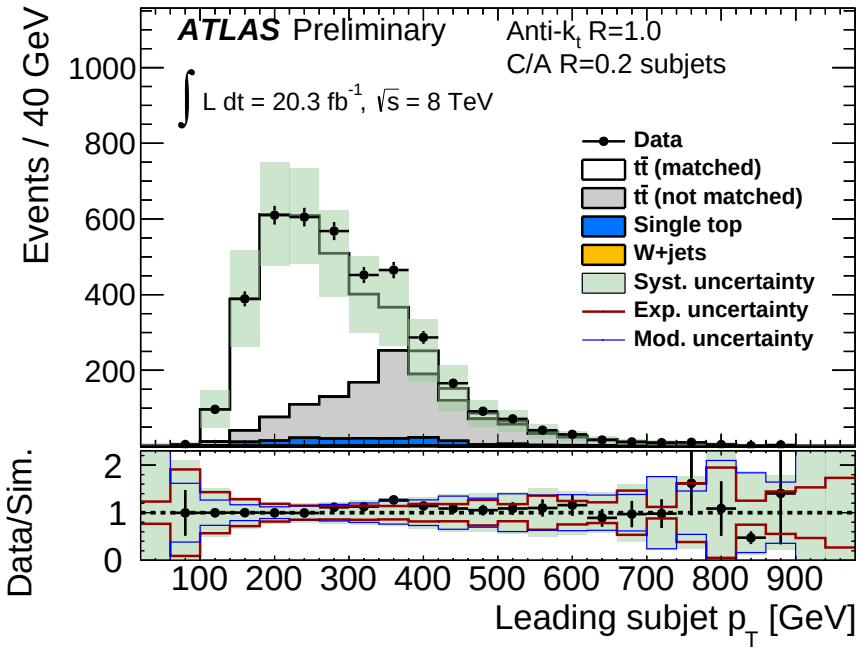


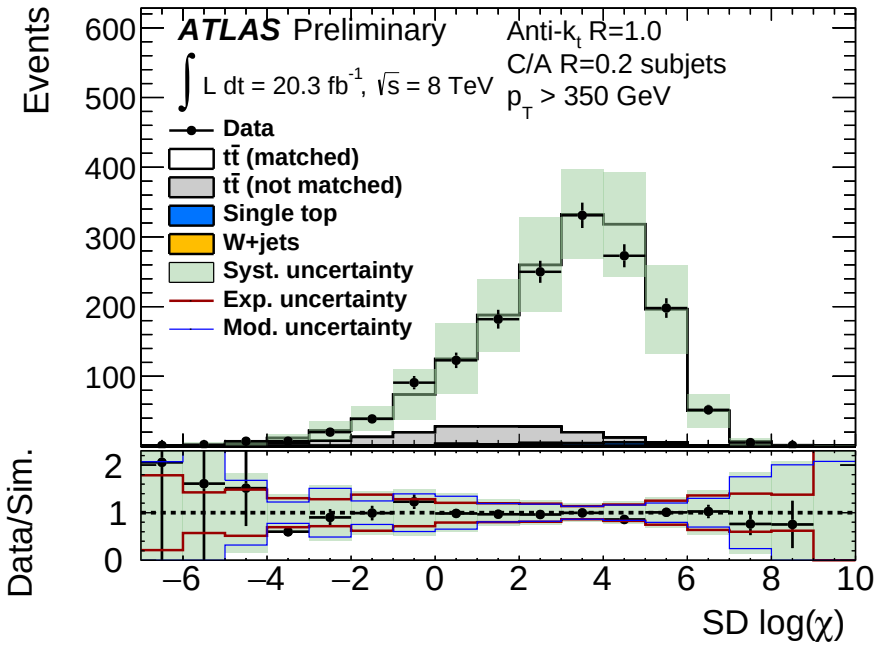
Shower Deconstruction

- Calculate, for each subjet of the input, the probability that the subjet is associated with a certain source of radiation (ISR, light quark, etc.).
  - ... for signal (top) and background (QCD).
  - ... for all possible combinations of radiation sources and subjets.
- Discriminant χ is ratio of sum of signal probabilities to sum of background probabilities.



Shower Deconstruction

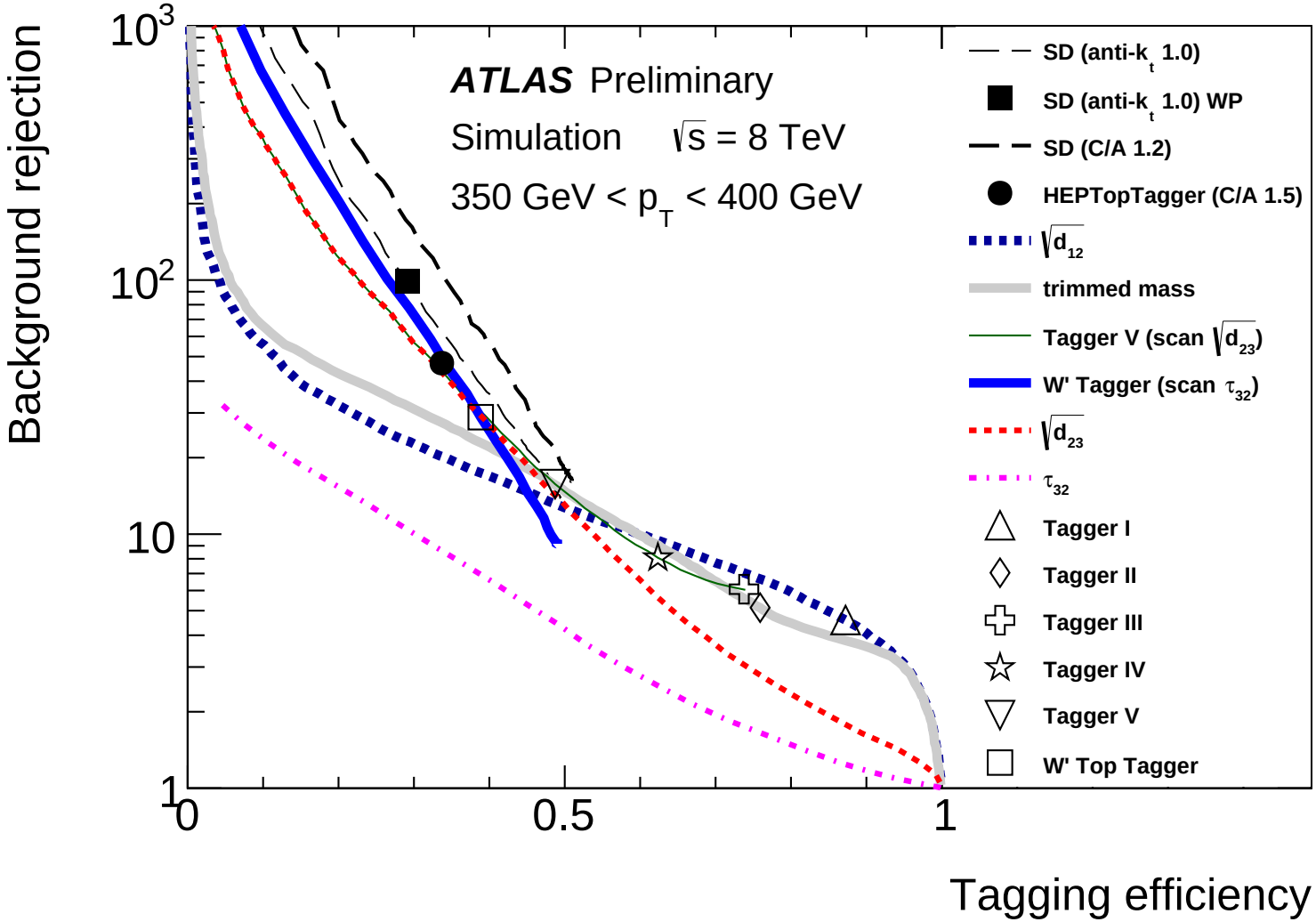




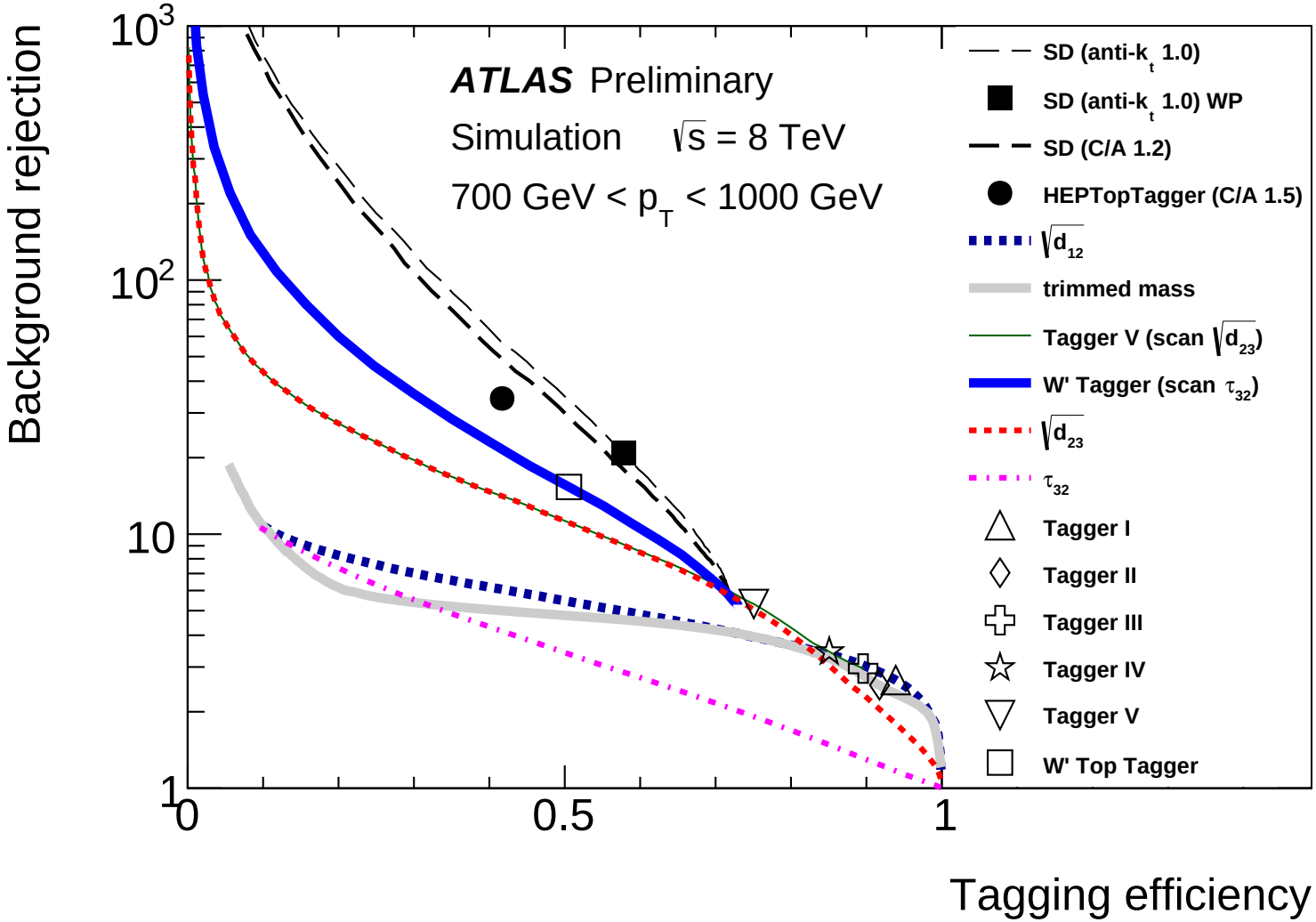
Performance Comparison in Simulation

- Characterising different taggers in terms of their tagging efficiency & background rejection.
- ROC curves made in many p_T bins.
- Which one is best?
  - ... this depends on what you want to do, of course!

Performance Comparison in Simulation



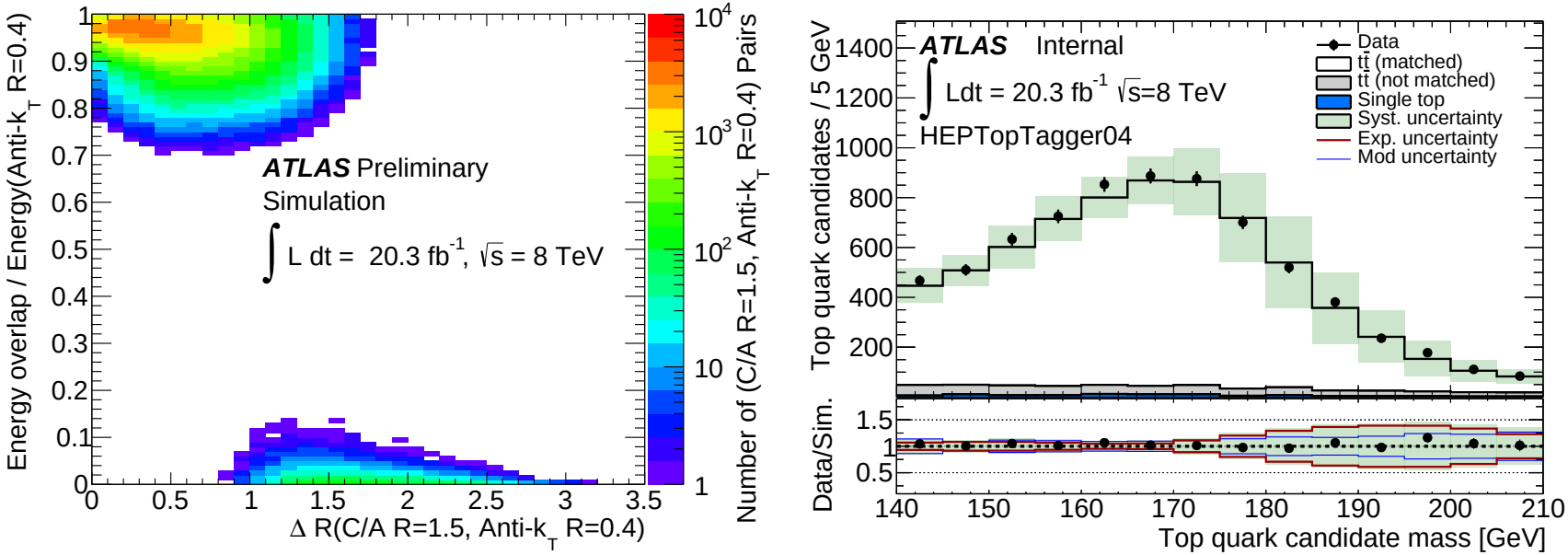
Performance Comparison in Simulation



HEPTopTagger04

- Also studied here: HTT04, a modified approach to the HTT method which uses anti-k_T $R = 0.4$ jets as inputs, rather than C/A $R = 1.5$.
- Designed for busy environments, e.g.

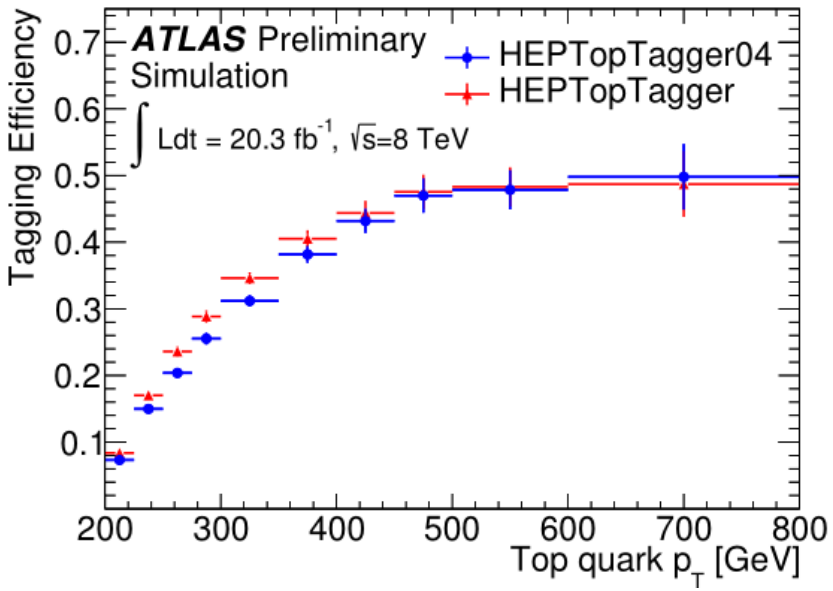
$$pp \to H^+\bar{t}b \to tb\bar{t}\bar{b}$$



HEPTopTagger04

- Also studied here: HTT04, a modified approach to the HTT method which uses anti-k_T $R = 0.4$ jets as inputs, rather than C/A $R = 1.5$.
- Designed for busy environments, e.g.

$$pp \to H^+ \bar{t} b \to t b \bar{t} \bar{b}$$



Top Tagging Efficiency Measurement

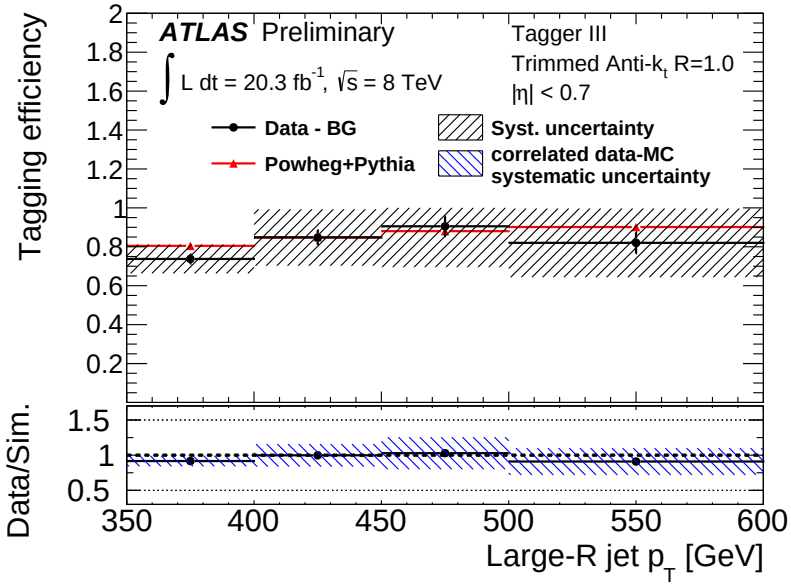
- Efficiency calculated as the fraction of tagged to preselected large-R jets.
- For measurement in data, backgrounds are statistically subtracted.
- For measurement in simulation, only truth-matched $t\bar{t}$ is considered.

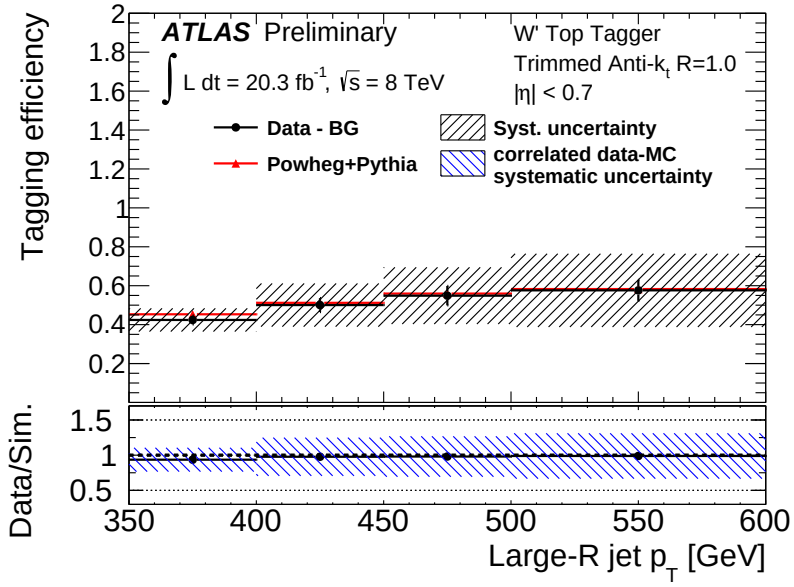
$$f_{\text{data}} = \left(\frac{N^{\text{tag}}_{\text{data}} - N^{\text{tag}}_{t\bar{t}\ \text{unmatched}} - N^{\text{tag}}_{\text{non-}t\bar{t}}}{N_{\text{data}} - N_{t\bar{t}\ \text{unmatched}} - N_{\text{non-}t\bar{t}}} \right) \tag{1}$$

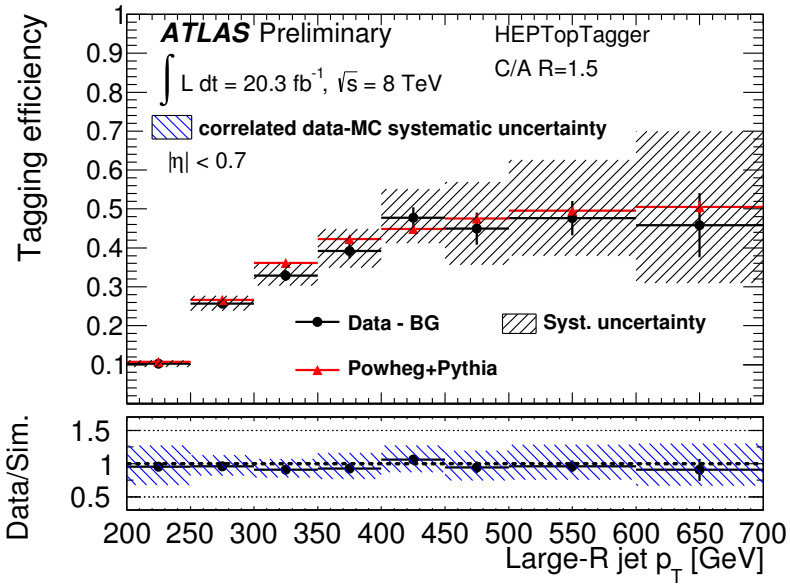
$$f_{\text{MC}} = \left(\frac{N^{\text{tag}}_{\text{MC}}}{N_{\text{MC}}} \right) \tag{2}$$

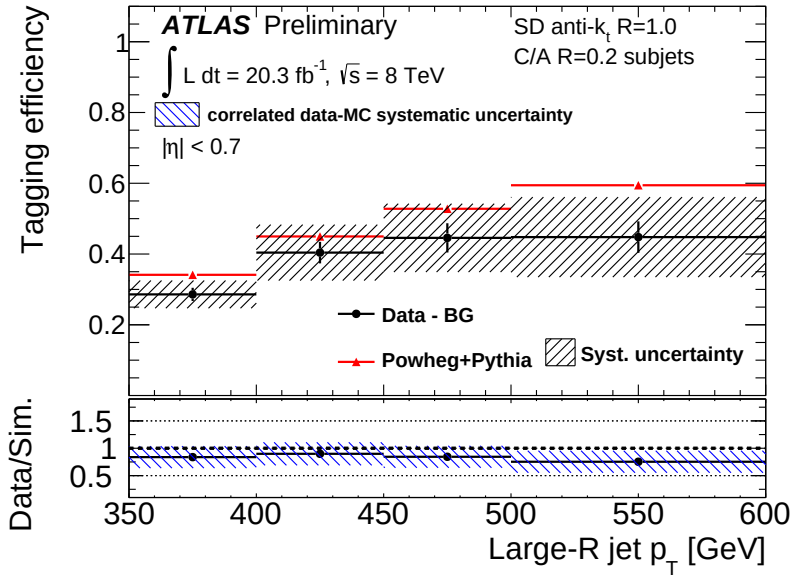
- Unfortunately, systematic uncertainties **do not** fully cancel in this ratio!

Top Tagging Efficiencies in Data



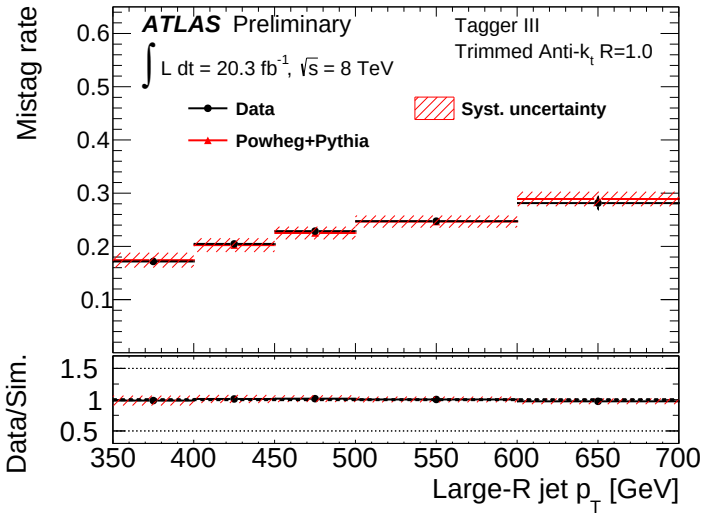


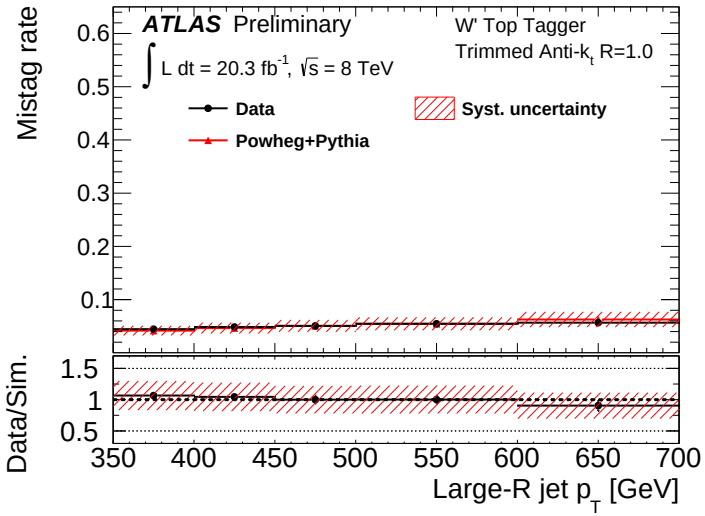


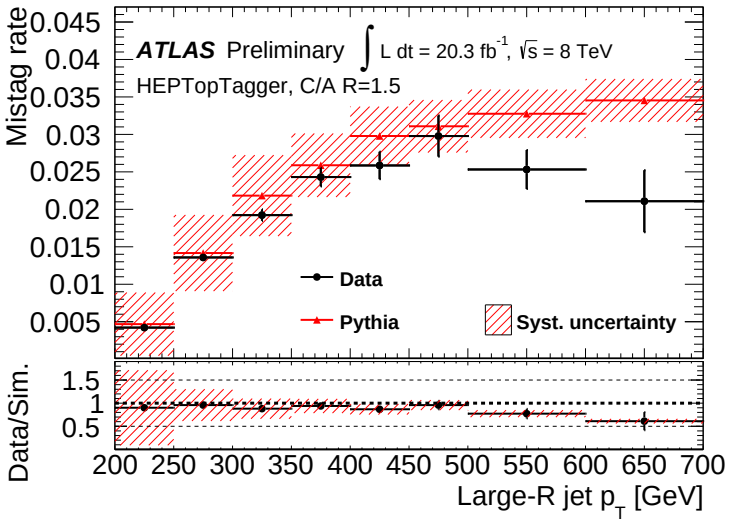


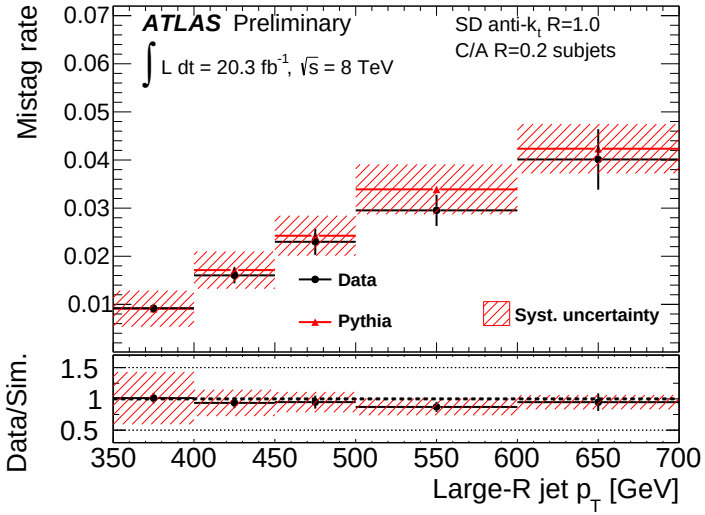
Top Tagging Fake Rates in Data.

- Fake rates measured in QCD-enriched sample 'with small top contamination.'



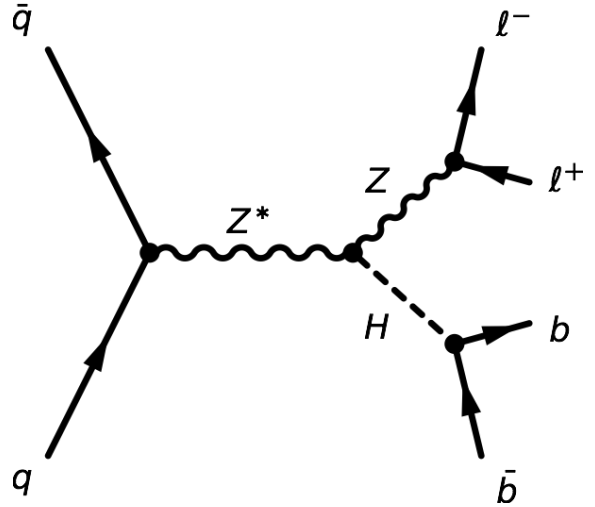


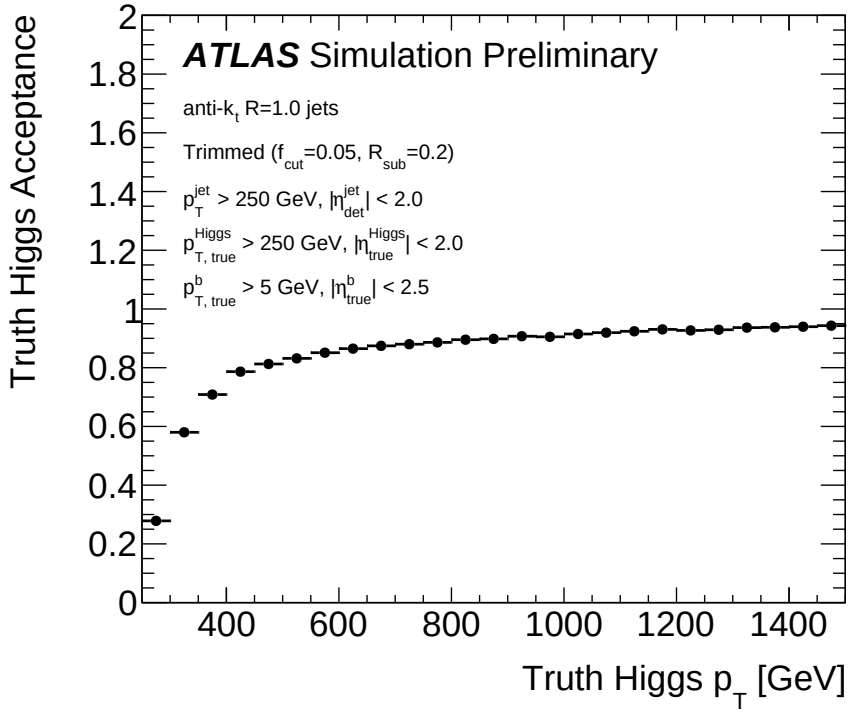




Tagging Boosted $H \to b\bar{b}$

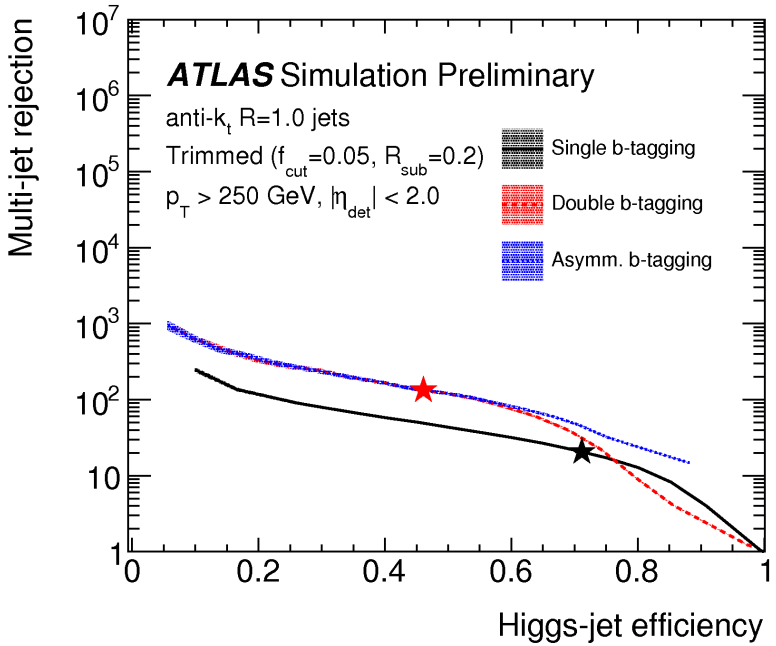
- Want to identify boosted $H \to b\bar{b}$ reconstructed as trimmed anti-k_t $R = 1.0$ jet.
- Backgrounds considered:
  - Multi-jet events.
  - Boosted, hadronically decaying top quarks.
- Three handles for S/B discrimination:
  - Heavy flavour content of large-R jet.
  - Large-R jet mass.
  - Large-R jet substructure.

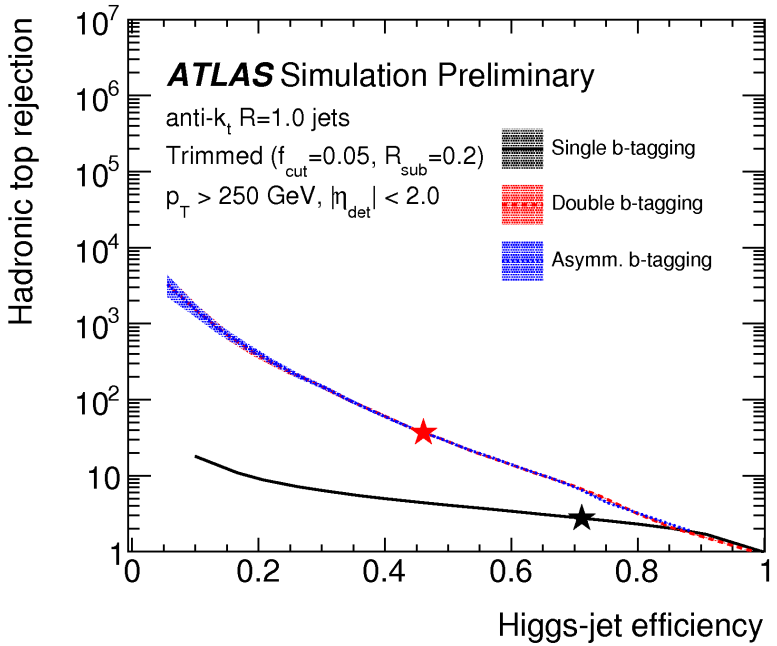




b-Tagging in Large-R Jets

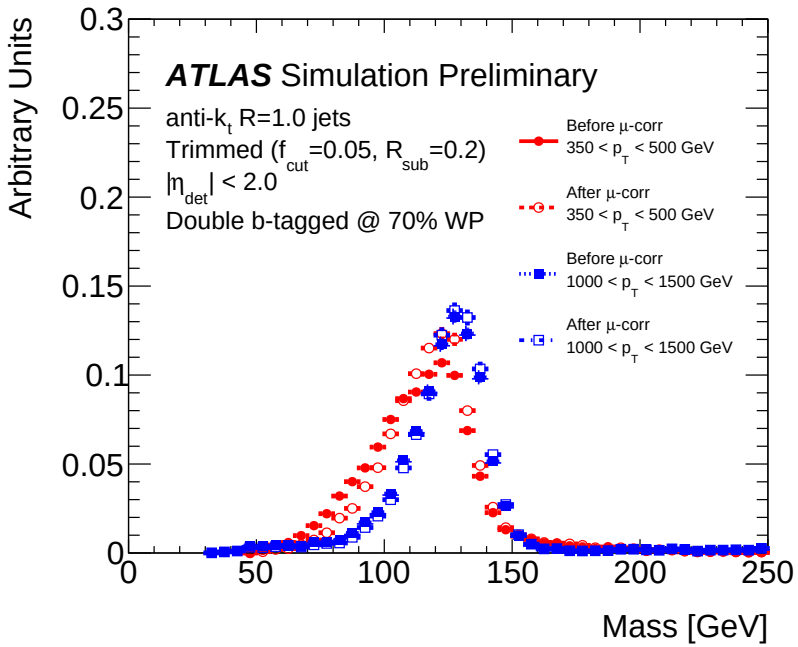
- Find leading pair of anti-k_t $R = 0.2$ track jets ghost-associated with a large-R jet (recall Michael's talk from yesterday).
- Attempt to tag track jets with MV2c20 algorithm. Study single, double and one-tight-one-loose scenarios.

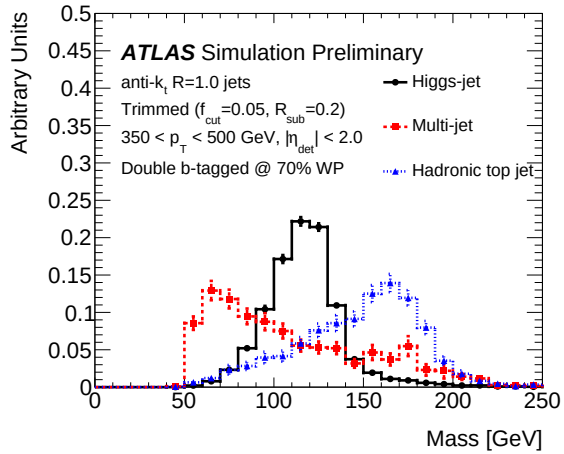


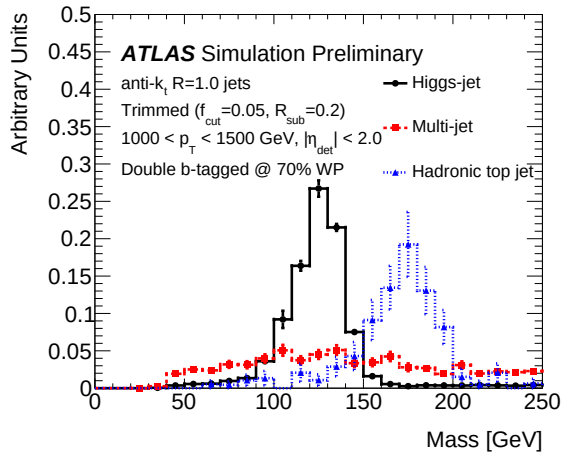


Mass Window and Muon-in-b-Jet Correction

- Before setting mass window, can increase mass resolution by correcting for semi-leptonic b hadron decays to muons.

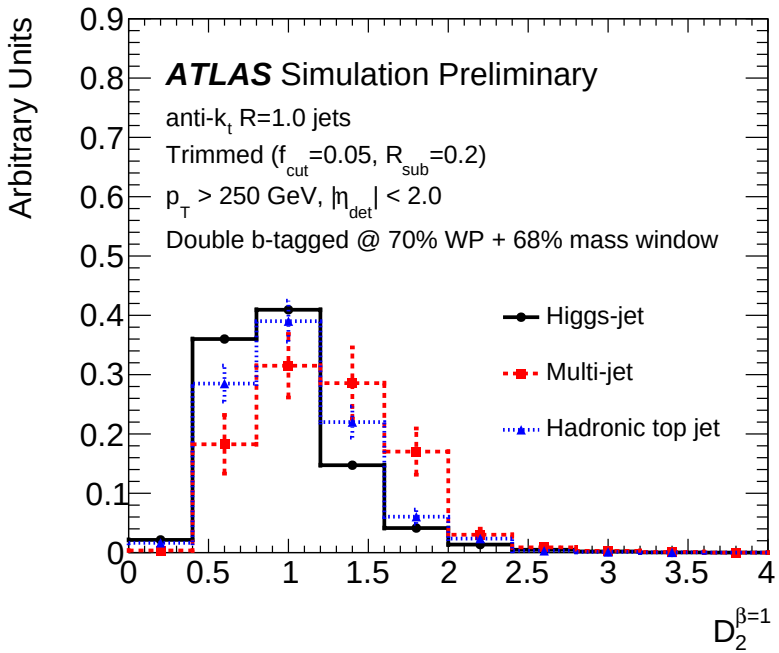


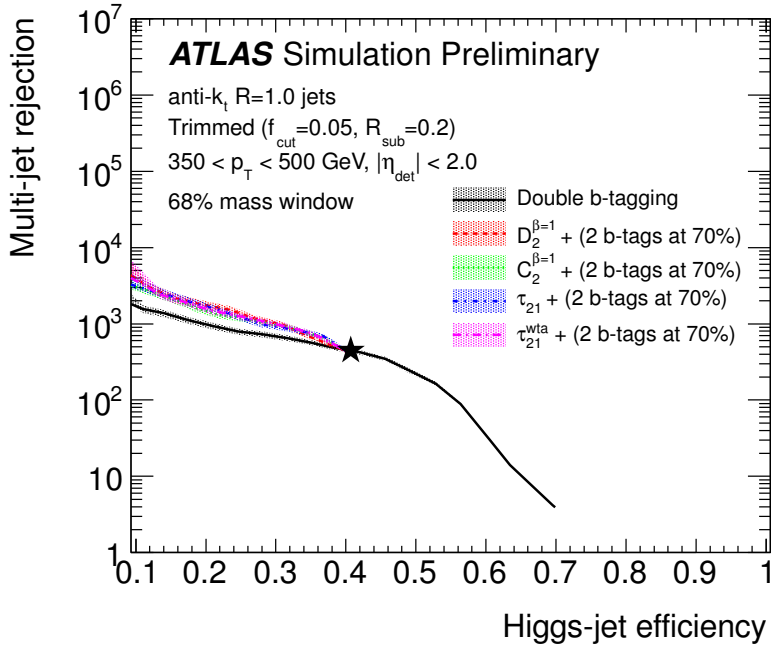




Including Substructure Information

- Examined performance of tagger with 68% mass window and b-tagging WP fixed at 70% efficiency, when scanning additionally on one of three jet substructure variables: $D_2^{\beta=1}$, $C_2^{\beta=1}$ and τ_{21}^{wta}.
  - Performance similar for all variables considered, $D_2^{\beta=1}$ provides slightly better improvement overall.

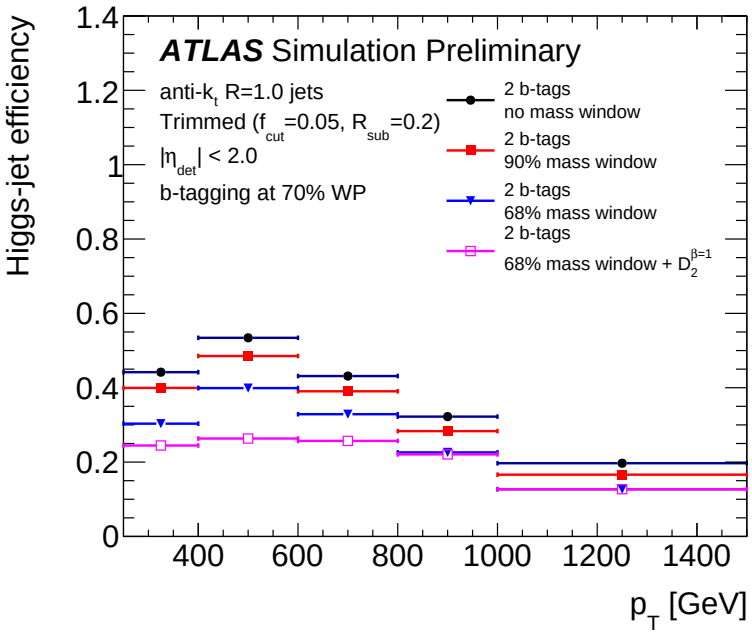


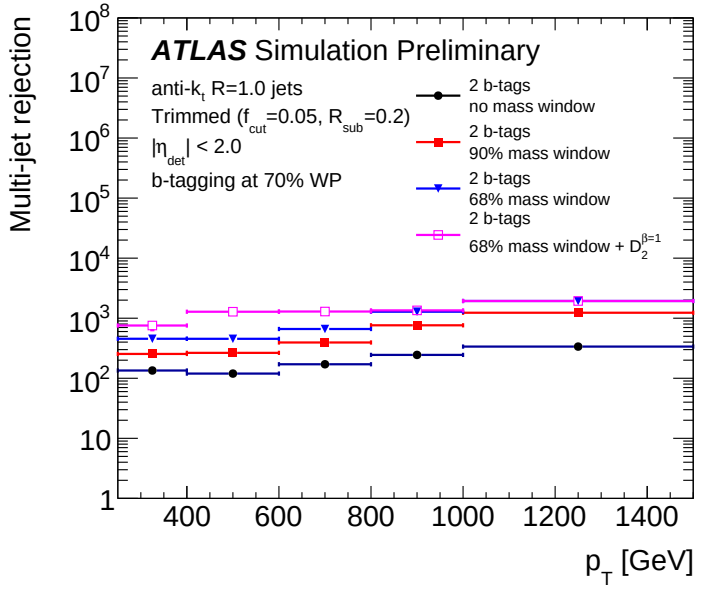


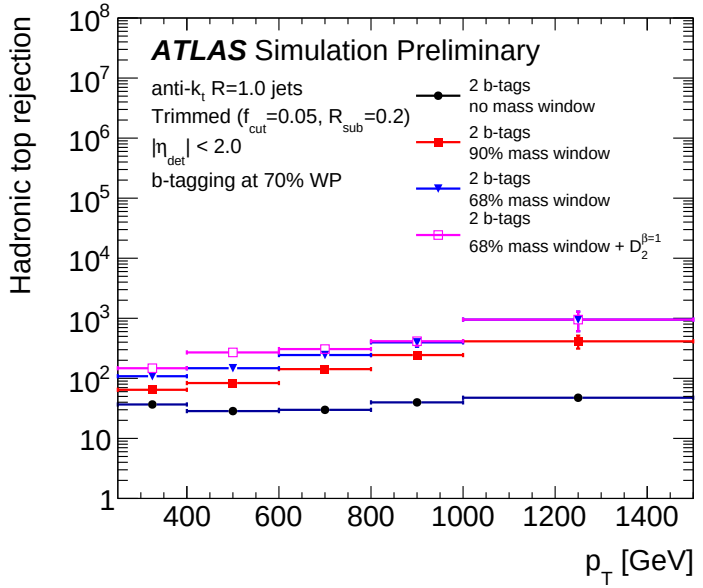
Benchmark $H \to b\bar{b}$ Selections

Selection	double b-tagging	large-R jet Mass	$D_2^{(\beta=1)}$
`Loose`	70% WP	90% window, $m \in [76, 148]$ GeV	-
`Medium`	70% WP	68% window, $m \in [95, 137]$ GeV	-
`Tight`	70% WP	68% window, $m \in [95, 137]$ GeV	p_T-dependent cut

Table 1: Criteria used for the different Higgs-jet tagging selections.

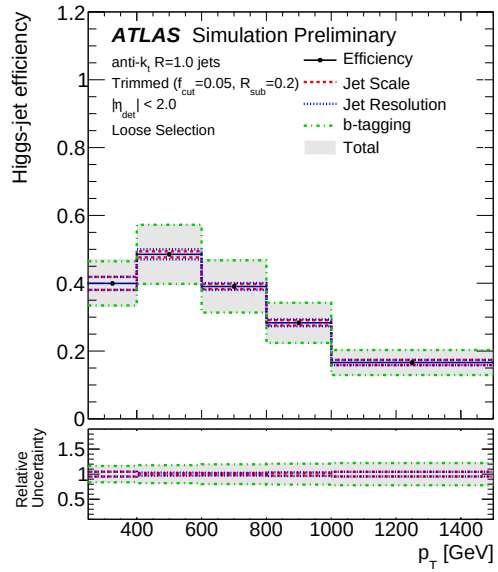


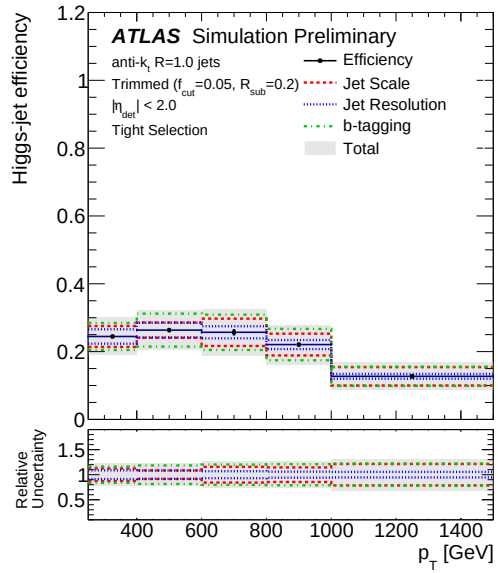




Systematic Uncertainty Estimates on $H \to b\bar{b}$ Tagging

	Loose	Medium	Tight
efficiency	0.41 ±0.07	0.32 ±0.06	0.25 ±0.05
Multi-jet rejection			
Inclusive	260 ±50	460 ±90	800 ±210
Light-flavor	$\mathcal{O}(10^5)$	$\mathcal{O}(10^5)$	$\mathcal{O}(10^6)$
cl	$\mathcal{O}(10^3)$	$\mathcal{O}(10^3)$	$\mathcal{O}(10^4)$
bl	$\mathcal{O}(10^2)$	$\mathcal{O}(10^2)$	$\mathcal{O}(10^3)$
bc	$\mathcal{O}(10)$	$\mathcal{O}(10)$	$\mathcal{O}(10^2)$
cc	250 ±150	480 ±310	1200 ±900
bb	11 ±2	19 ±4	31 ±9
Hadronic top rejection			
Inclusive	67 ±17	110 ±30	160 ±50
bl	360 ±230	660 ±460	810 ±600
bc	24 ±6	39 ±11	53 ±16





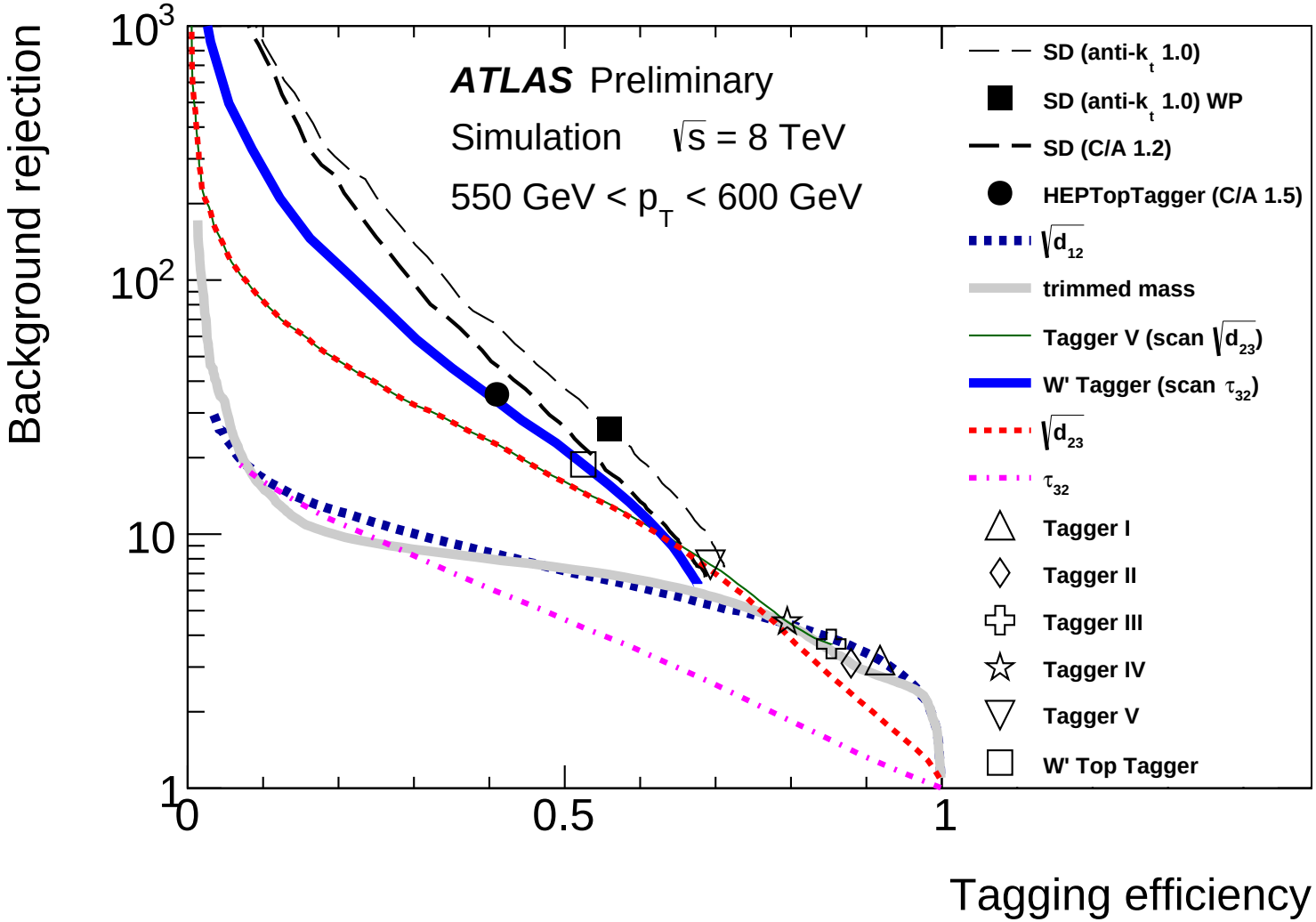
- For loose working point, b-tagging uncertainties largest.
- With tigher working points, jet scale & resolution uncertainties increase.

Concluding Remarks

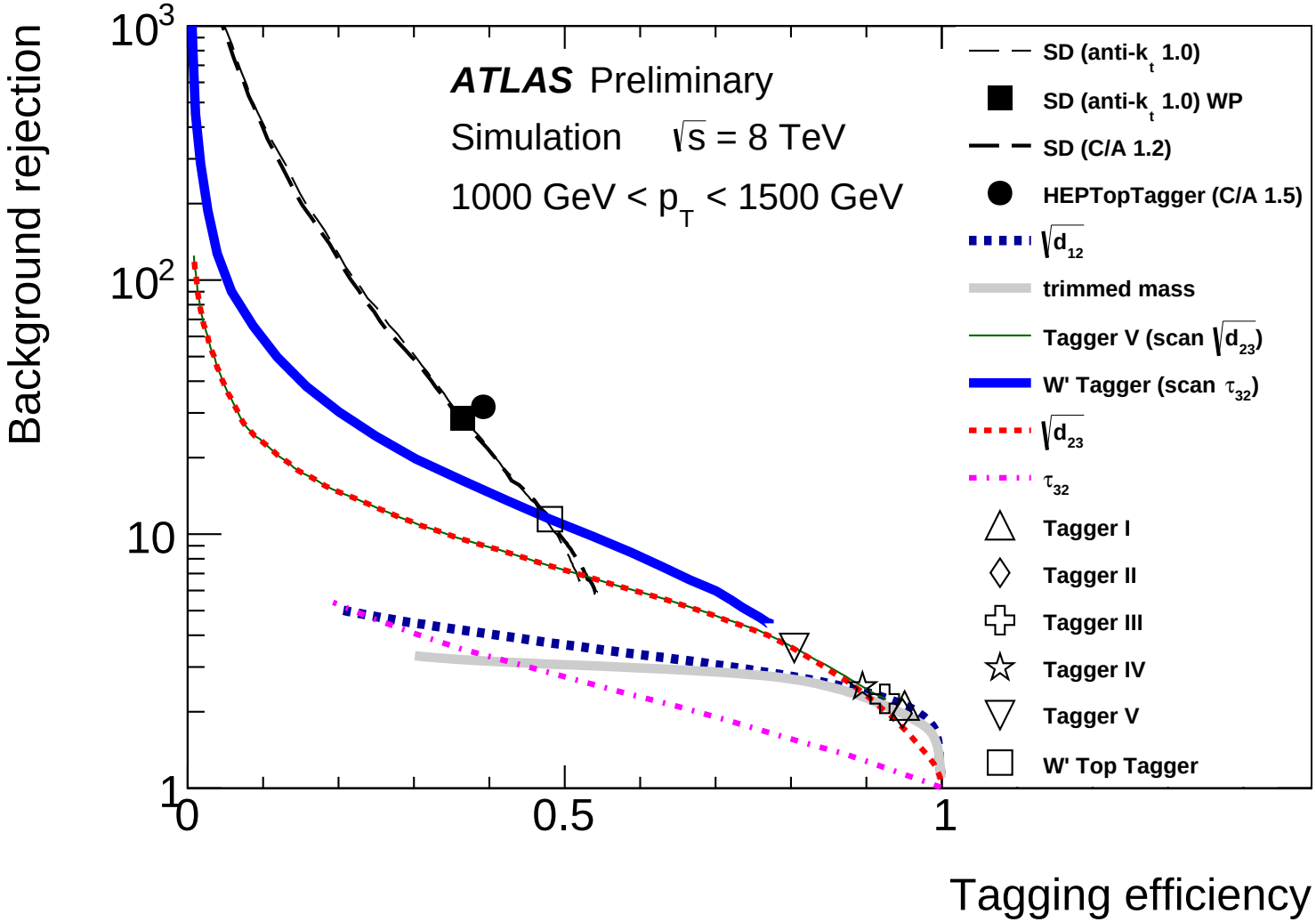
- Thorough performance comparisons for several different methods of top-tagging utilized by ATLAS in run I.
  - New in-situ efficiency and fake rate measurements in data for different taggers: substructure-based, HTT and SD.
    - Performance of HTT04 studied in the context of a high jet multiplicity final state (charged Higgs search).
  - ROC curve sets comparing tagger performance in simulation produced in many p_T bins.
- $H \to b\bar{b}$ tagger developed for use in run-II analyses.
  - Tagging efficiency and background rejection for QCD multijet and $t\bar{t}$ backgrounds evaluated in simulation.
  - Expected performance studied in terms of three benchmark selections.
  - Estimates of systematics based on extrapolations from Run I.

Supplemental Material

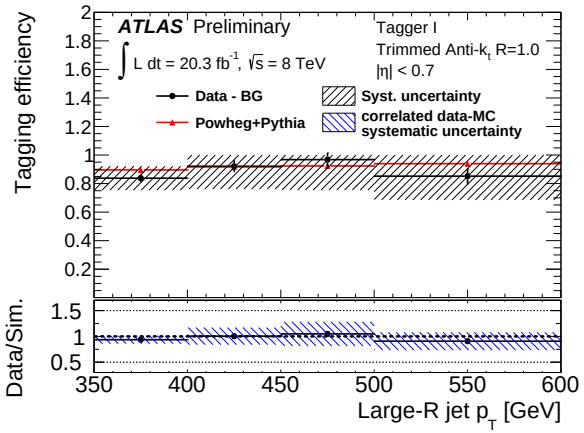
Performance Comparison

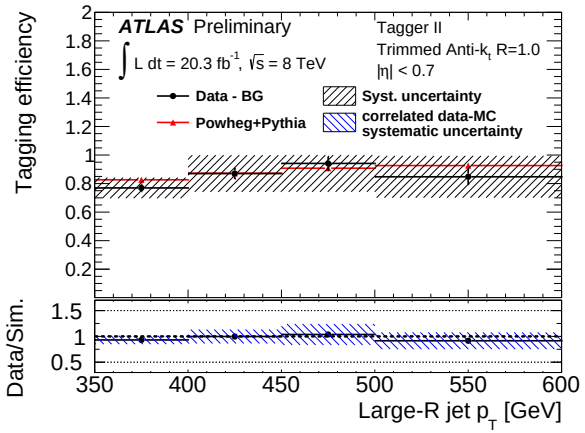


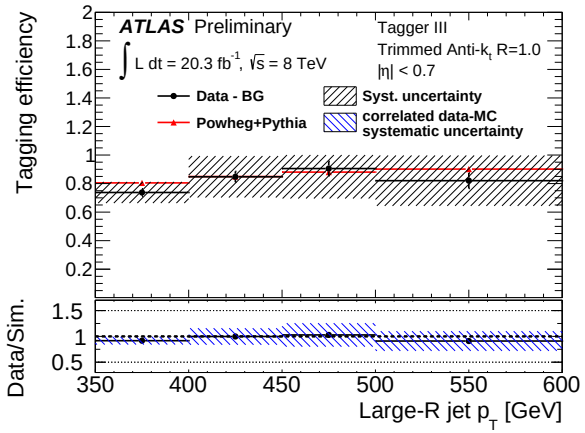
Performance Comparison

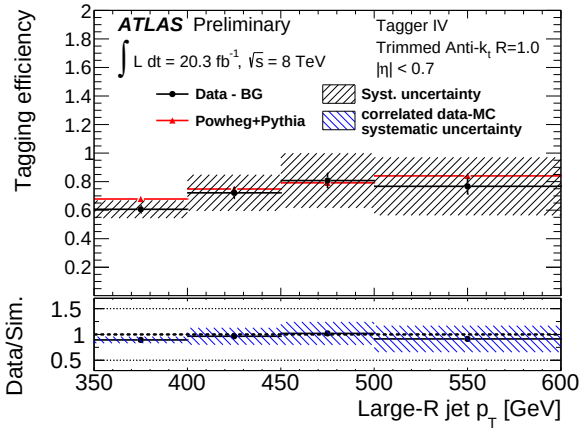


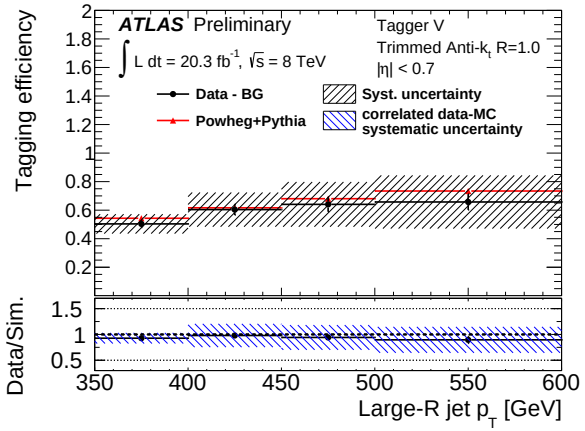
Substructure-Based Top Tagging: Forward Tagging Efficiency in Data.

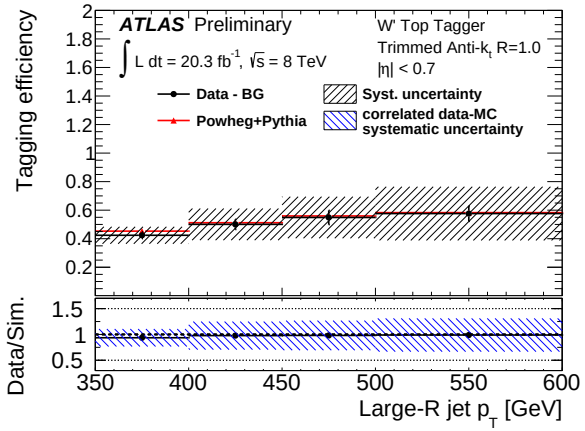




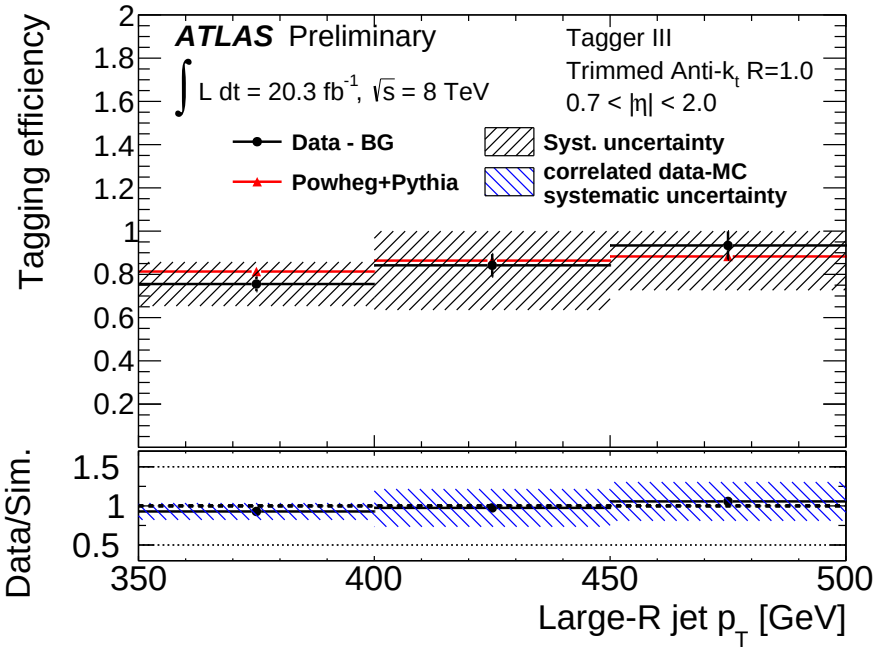


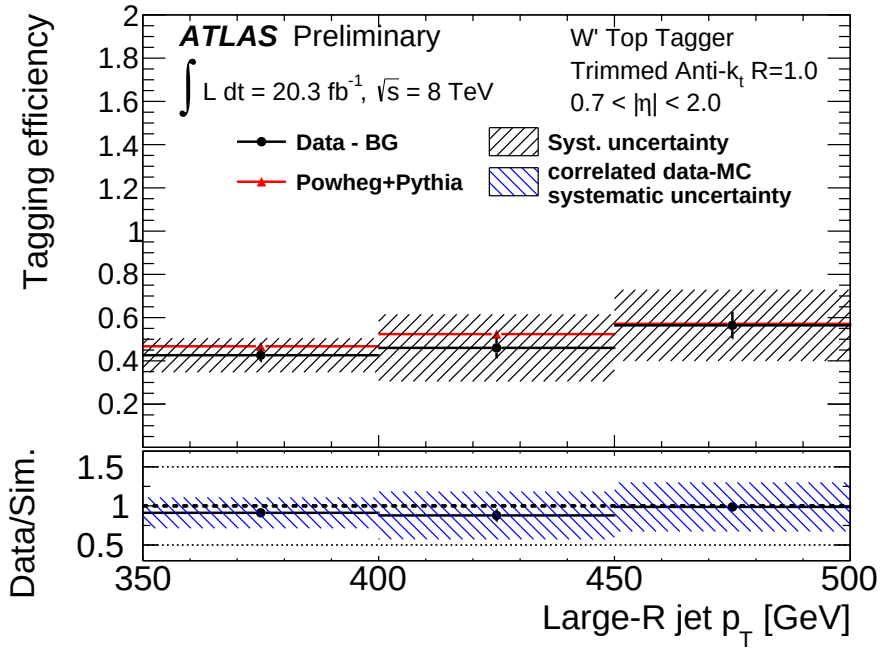






Substructure-Based Top Tagging: Forward Tagging Efficiency in Data.





HEPTopTagger & SD: Forward Tagging Efficiencies in Data