

Identification of b -jets in boosted and dense jet event topologies

Nikola Whallon

University of Washington

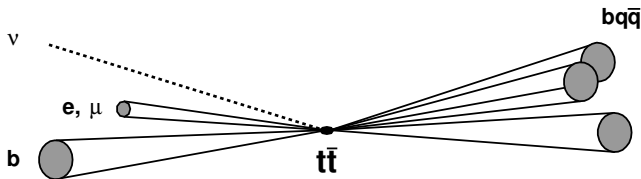
July 19, BOOST 2016 in Zurich



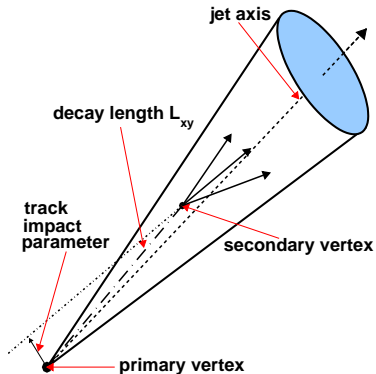
Introduction

- ▶ b -tagging very important for physics measurements and searches at LHC
- ▶ higher \sqrt{s} energy in Run 2 \rightarrow denser/boosted environments
- ▶ denser environments \rightarrow worse standard b -tagger performance
- ▶ during Run 1, ATLAS developed and calibrated MVb to increase performance in dense environments
- ▶ MVb description: ATL-PHYS-PUB-2014-014
- ▶ MVb calibration: ATLAS-CONF-2016-001 (focus of this talk)

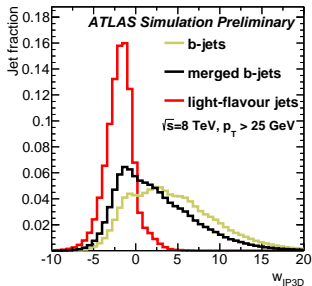
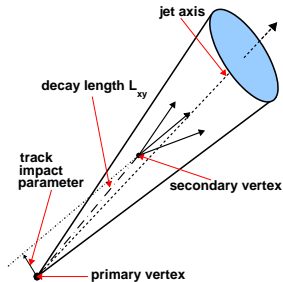
example of boosted $t\bar{t}$ event with a b -jet merged with other jets:



- ▶ b -hadrons have a large mass (~ 5 GeV) \rightarrow large lifetime ($O(\text{ps})$) \rightarrow large distance traveled before decaying ($O(\text{mm})$).
- ▶ This long-lived nature results in tracks originating from b -hadrons having large impact parameters (IPs) w.r.t. the primary vertex, and displaced secondary vertices (SVs).
- ▶ These observables allow us to develop methods of tagging jets originating from b -hadrons.



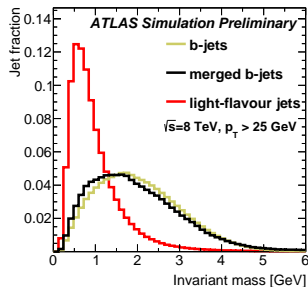
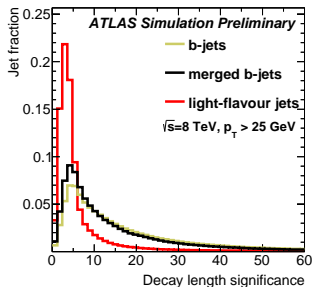
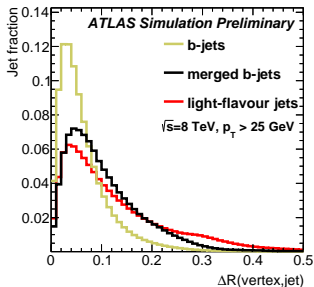
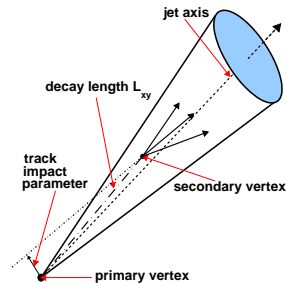
- ▶ signed IP offers good discriminating power between b -jets and other jets
- ▶ the sign of the IP is positive if the angle between the IP and the jet axis is < 90 degrees and negative otherwise



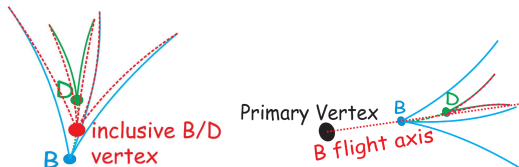
- ▶ the 3D (2D) IP information for all tracks selected by the IP algorithms are combined into the IP3D (IP2D) discriminants

Secondary Vertex Discriminants ATL-PHYS-PUB-2014-014

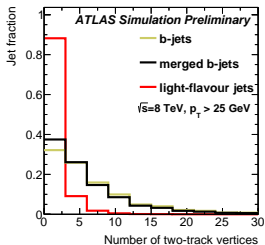
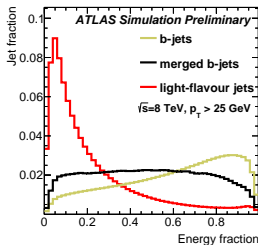
- ▶ reconstructing an inclusive SV gives many discriminating variables for b -jets
- ▶ we use an iterative vertex finding algorithm - SV1
- ▶ examples of discriminating variables from the SV found by SV1 are shown below



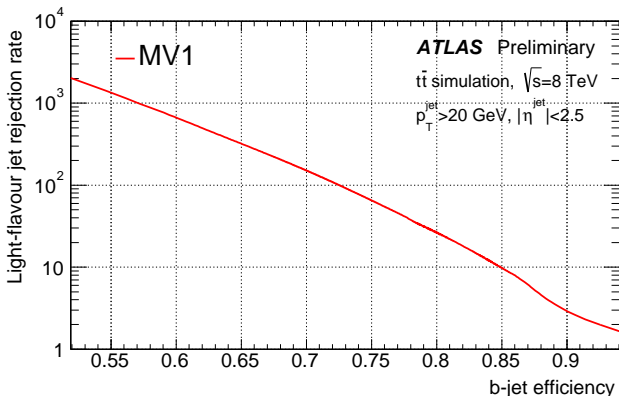
- ▶ JetFitter: algorithm which uses a Kalman Filter to reconstruct the b - to c -hadron decay chain



- ▶ advantage over SV1: if b - and c -hadron decay vertices are well separated, JetFitter offers much more precise information
- ▶ discriminating variables are created by JetFitter, e.g.:



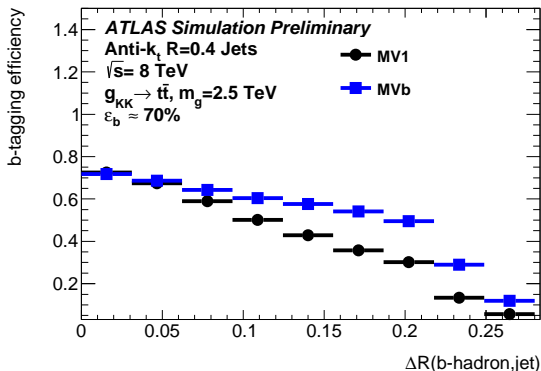
- ▶ ATLAS combines variables from the IP, SV, and JetFitter algorithms in an artificial neural network called MV1.
- ▶ MV1 was the standard b -tagging algorithm used during Run 1 and is powerful when b -jets are isolated.



Degradation of MV1 in Dense Environments ATLAS-CONF-2016-001

in dense environments (e.g. collimated/boosted jets), MV1 performance degrades due to:

- ▶ jet axis shifts away from b -hadron flight direction
- ▶ contamination of tracks from nearby light jets

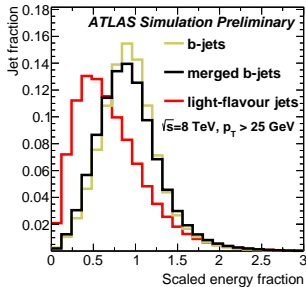
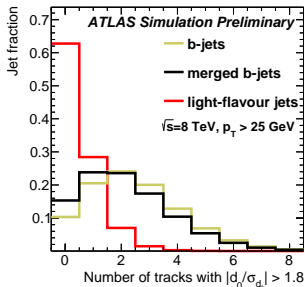
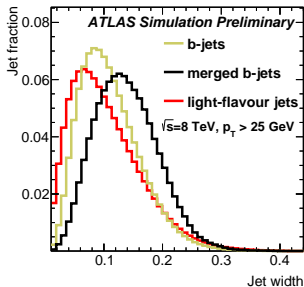


to combat this, ATLAS developed a new algorithm: MVb

MVb Multi-Variate B-Tagger ATL-PHYS-PUB-2014-014

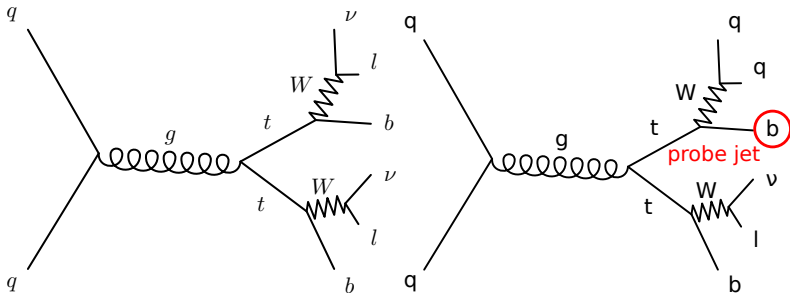
MVb = dedicated dense environment b -tagger, a BDT with MV1 variables + dense-environment-robust variables, e.g.:

- ▶ the jet width, $w_{\text{jet}} = \sum_{i=1}^N p_T^{\text{trk}_i} \Delta R(\text{trk}_i, \text{jet}) / \sum_{i=1}^N p_T^{\text{trk}_i}$
- ▶ the number of tracks with $|d_0/\sigma_{d_0}| > 1.8$
- ▶ energy fractions of SV1 and JetFitter vertices, scaled by the ratio of tracks associated to both the vertex and the jet to the tracks associated to the vertex



B-Tagging Efficiency Determination in Data with $t\bar{t}$

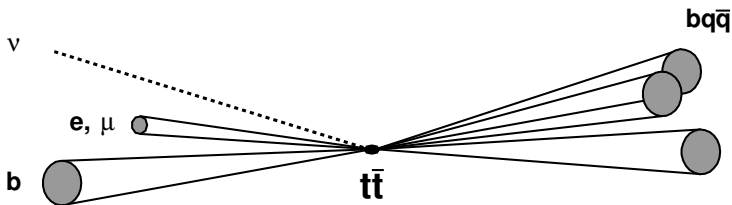
- ▶ interested in deriving b -tagging efficiency scale factors to match MC to data (for both MV1 and MVb)
- ▶ b -tagging calibration most precise using the leptonic $t\bar{t}$ channel (left)
- ▶ tag-and-probe method with the semi-leptonic $t\bar{t}$ channel (SL T&P) better for this study (right)



- ▶ SL T&P better b/c high jet multiplicity, boost from hadronic side \rightarrow collimated jets/dense environment we are interested in

Event/Object Reconstruction ATL-PHYS-PUB-2014-014

- ▶ The final state consists of 1 lepton, E_T^{Miss} from the neutrino, and at least 4 jets.
- ▶ Jets are reconstructed using the anti- k_T algorithm with $R = 0.4$.
- ▶ The 70% MV1 and MVb working point is applied to jets to reconstruct the b-jets.



χ^2 Minimization

χ^2 minimization procedure used to reduce combinatorics

$$\chi_{\text{total}}^2 = \chi_{W_h}^2 + \chi_{t_h - W_h}^2 + \chi_{t_l}^2 + \chi_{\Delta p_T}^2$$

$$\text{recon. (hadronic) } W \text{ mass} \quad \chi_{W_h}^2 = \left(\frac{m_{jj} - M_{W_h}}{\sigma_{M_{W_h}}} \right)^2$$

$$\text{recon. (hadronic) } t \text{ mass} \quad \chi_{t_h - W_h}^2 = \left(\frac{m_{jjj} - m_{jj} - M_{t_h - W_h}}{\sigma_{M_{t_h - W_h}}} \right)^2$$

$$\text{recon. (leptonic) } t \text{ mass} \quad \chi_{t_l}^2 = \left(\frac{m_{jl\nu} - M_{t_l}}{\sigma_{M_{t_l}}} \right)^2$$

$$\text{recon. hadronic-leptonic } p_T \text{ balance} \quad \chi_{\Delta p_T}^2 = \left(\frac{(p_{T,jjj} - P_{T,t_h}) - (p_{T,jl\nu} - P_{T,t_l})}{\sigma_{\Delta p_T}} \right)^2$$

Measuring the B-Tagging Efficiency

to measure the b -tagging efficiency with our probe jet (ϵ_b), we need to know:

- ▶ fraction of probe jets truth matched to a $b/c/l$ -jet or fake jet from multijet background ($f_{b/c/l\text{-jets}}$, f_{fake})
- ▶ mistag efficiencies of our b -tagger ($\epsilon_{c/l}$, ϵ_{fake})

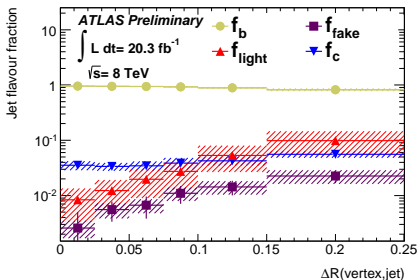
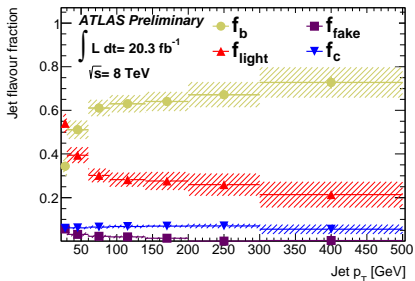
then, the b -tagging efficiency is given by the following:

$$\epsilon_b = \frac{f_{\text{tag}}}{f_{b\text{-jets}}} \cdot (1 - \epsilon_c f_{c\text{-jets}} - \epsilon_l f_{l\text{-jets}} - \epsilon_{\text{fake}} f_{\text{fake}})$$

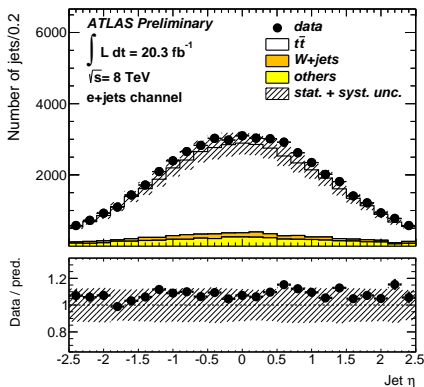
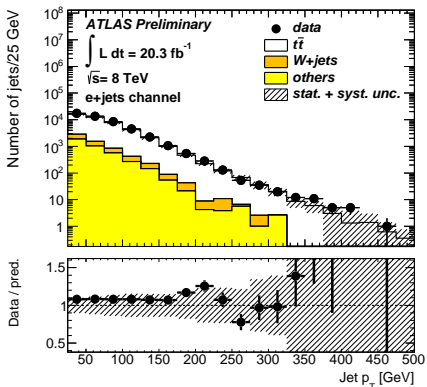
where f_{tag} is the fraction of probe jets passing a cut of our tagger (in this study, we used the 70% working point cut for MV1 and MVb)

Flavour Fraction and Truth Labeling ATLAS-CONF-2016-001

- ▶ $f_{b\text{-jets}}$, $f_{c\text{-jets}}$, etc, critical for evaluating b -tagging efficiency (any large uncertainty in flavour fractions will result in a large uncertainty in b -tagging efficiency).
- ▶ flavour fraction determined by truth labeling in MC: jet labeled as b -/ c -jet if $\Delta R(\text{jet}, b\text{-}/c\text{-hadron}) < 0.3$, labeled as light jet otherwise
- ▶ c - and light jet contamination of the probe jet sample occurred primarily when the χ^2 minimization procedure failed (assigned wrong permutation).



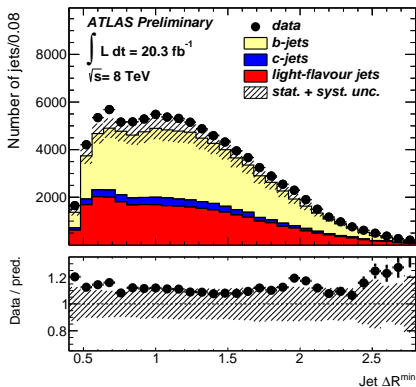
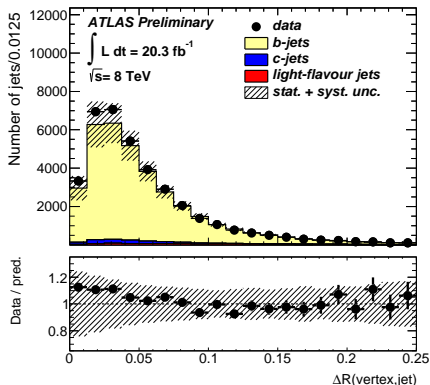
Here are the p_T and η distributions of the probe jets after all corrections and event selection in data and MC:



There is a slight excess in data, but the shapes are very similar and data/MC agree within uncertainties.

Kinematic Distributions II ATLAS-CONF-2016-001

Here are the $\Delta R(\text{secondary vertex, jet})$ distribution for probe jets which contain secondary vertices and the ΔR^{\min} (the minimum ΔR between the probe jet and other jets) distribution for all probe jets.



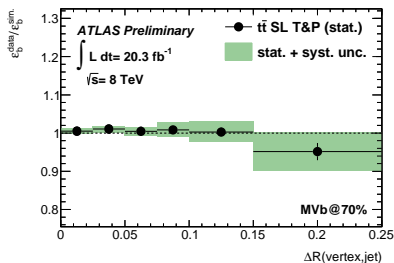
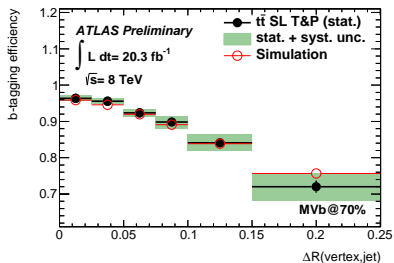
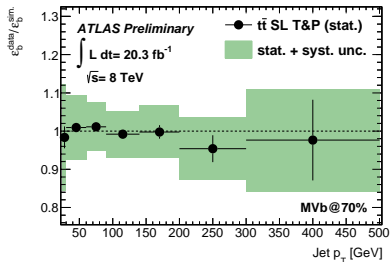
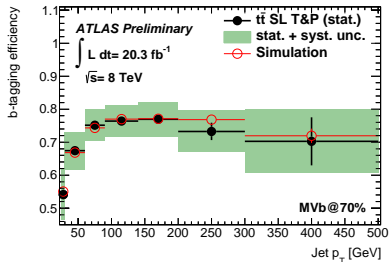
There is a slight excess in data, but the shapes are very similar.

Systematics

- ▶ The major systematics are those which alter the flavour fraction (e.g. the MC generator systematic, which varies little with p_T).
- ▶ The Jet Energy Scale and Jet Resolution systematics are also quite large, but decrease with p_T .

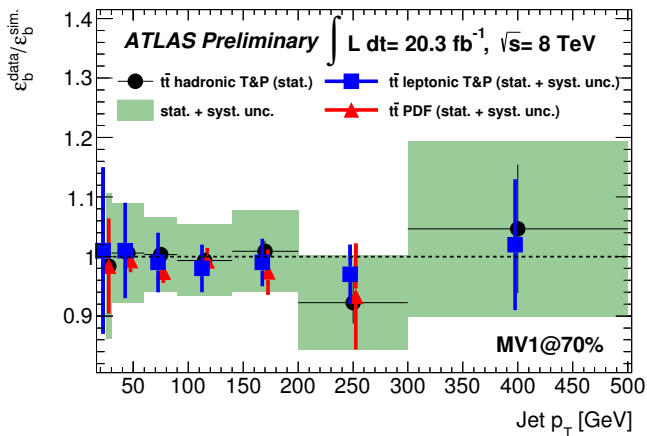
p_T [GeV]	25-30	30-60	60-90	90-140	140-200	200-300	300-500	
Mis-tag ϵ_c	1.5	0.9	0.7	0.8	0.8	0.9	0.5	
Mis-tag ϵ_{light}	1.2	0.4	0.3	0.2	0.2	0.2	0.8	
Mis-tag ϵ_{fake}	0.6	0.2	0.2	0.1	0.0	0.0	0.1	
Multijet normalisation	0.9	0.6	0.5	0.4	0.3	0.0	0.1	
$t\bar{t}$ normalisation	0.8	0.6	0.5	0.5	0.4	0.3	0.2	
Z+jets normalisation	0.4	0.3	0.2	0.2	0.1	0.2	0.0	
Single-top normalisation	0.3	0.3	0.2	0.2	0.3	0.2	0.6	
Luminosity	0.3	0.2	0.1	0.1	0.1	0.1	0.0	
Jet energy scale	6.2	2.1	0.8	0.4	0.4	0.2	1.9	lower with p_T
Jet energy resolution	8.6	4.1	2.5	2.3	1.6	2.5	1.5	
ISR/FSR	5.7	4.8	4.1	3.7	3.8	2.7	1.8	
Top-quark mass	0.8	0.4	0.0	0.4	0.2	0.5	1.5	
Top-quark p_T reweighting	0.2	0.0	0.1	0.6	1.5	2.7	2.4	
$t\bar{t}$ generator	5.9	4.2	3.6	4.0	4.7	5.5	6.9	constant with p_T
Fragmentation	2.1	2.2	1.0	1.0	1.1	2.9	2.0	
Hadronic calorimeter modelling	0.7	0.1	0.3	0.2	0.2	0.4	2.5	
Data statistics	2.9	0.8	0.8	0.9	1.8	3.5	10.6	
Total systematic	13.8	8.3	6.2	6.1	6.6	7.8	8.7	
Total uncertainty	14.1	8.3	6.3	6.2	6.9	8.5	13.7	
Scale factor	0.98 ± 0.14	1.01 ± 0.08	1.01 ± 0.06	0.99 ± 0.06	1.00 ± 0.07	0.95 ± 0.08	0.98 ± 0.13	

MVb Results ATLAS-CONF-2016-001



the b -tagging performance does degrade with increasing $\Delta R(\text{secondary vertex}, \text{jet})$, but the data is well-modeled in MC

the SL T&P calibration technique allows access to a higher jet p_T (~ 500 GeV) than the previous technique ($t\bar{t}$ PDF)



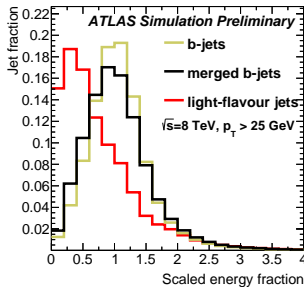
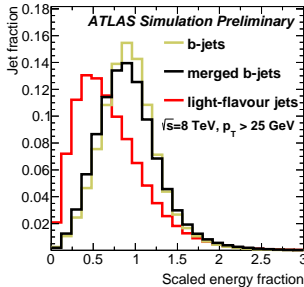
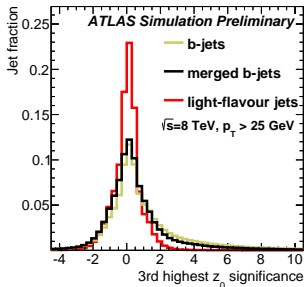
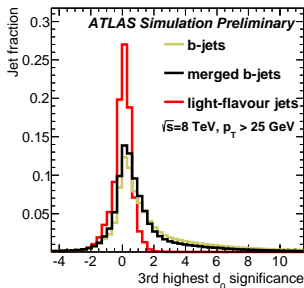
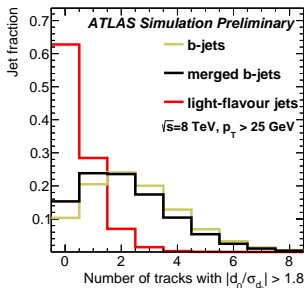
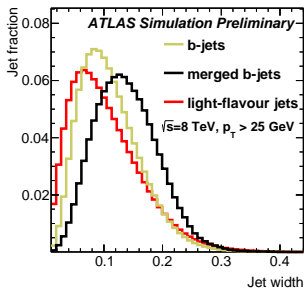
this is very useful for analyses using high p_T b -jets

Summary

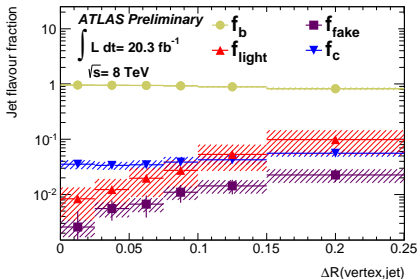
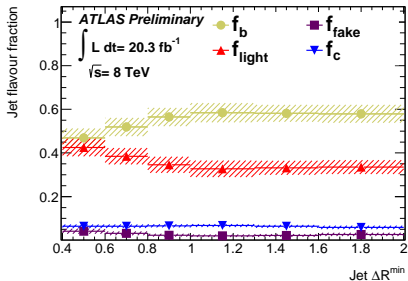
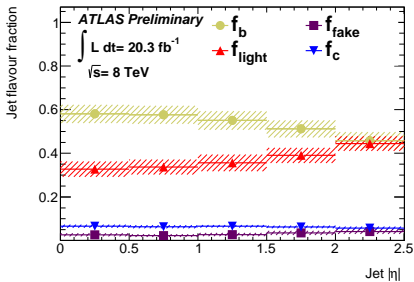
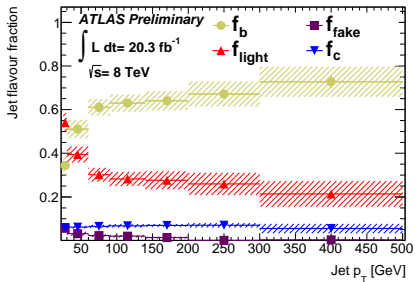
- ▶ b -tagging in dense environments will become increasingly important as searches push sensitivity into boosted regions
- ▶ the MVb algorithm developed by ATLAS is shown to be powerful in improving b -tagging performance in dense environments
- ▶ the systematic uncertainties of MVb are well understood and MVb results in data and simulation are consistent within uncertainties
- ▶ MVb scale factors at the 70% working point have been derived and are consistent with unity within uncertainties

Backup MVb Variables

ATL-PHYS-PUB-2014-014

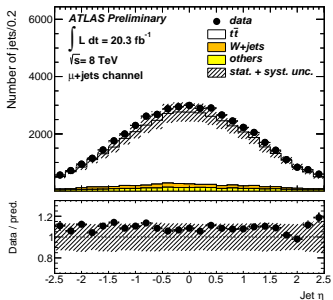
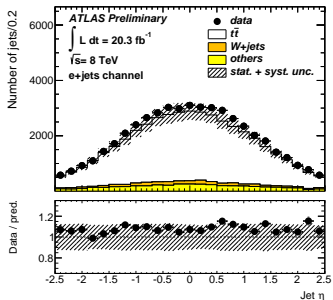
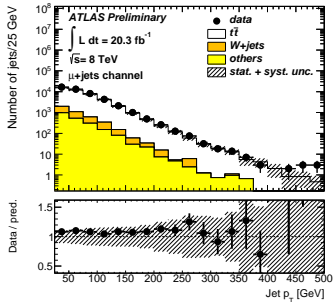
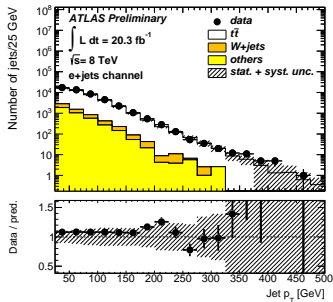


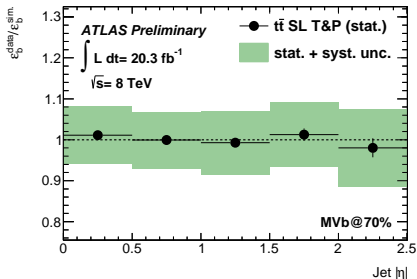
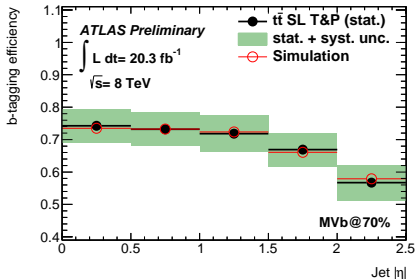
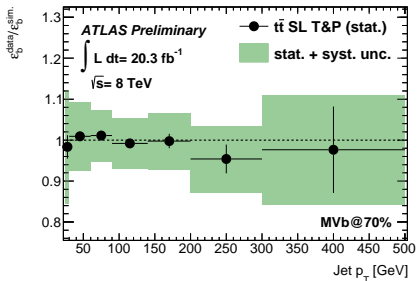
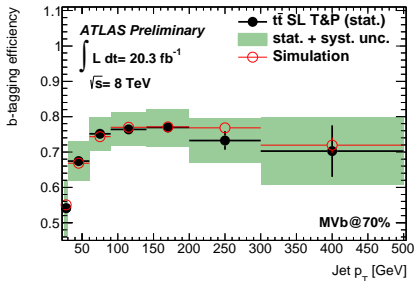
Backup Flavour Fraction ATLAS-COIN-2016-001



Backup Kinematic Distributions

ATLAS-CONF-2016-001





Backup Results II ATLAS-CONF-2016-001

