Performance of tau lepton identification in ATLAS with 7 TeV data

Andrzej Zemła

On behalf of the ATLAS Collaboration

Cracow Epiphany Conference
10-12 January Cracow

Overview

- Datasets and event selection
- Tau reconstruction in ATLAS
- Tau identification in ATLAS
- Tau identification variable distributions
- Cut based identification performance
- Systematic uncertainties
- Comparison with multivariate methods
- Summary

Data & event selection

Data

- collected with the ATLAS detector till mid of July 2010, corresponding to integrated luminosity of 244 nb^{-1}
- fraction of candidates from real τ leptons is negligible in this dataset but identification variables for QCD jets reconstructed as τ candidates can be compared to predictions from Monte Carlo simulations

MC

- di-jet Pythia samples with p_T of the leading outgoing partons in range 8-280 GeV

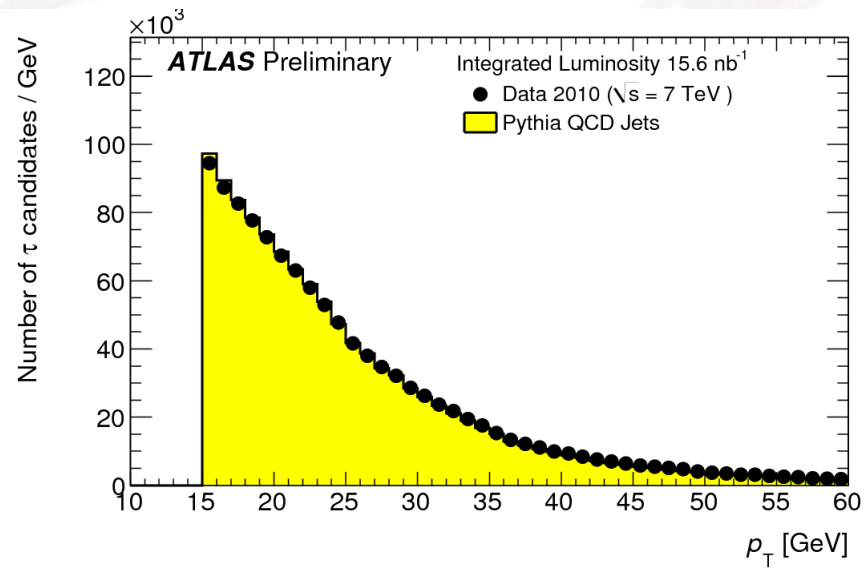
Event selection

- basic beam and data quality requirement for tracker and calorimeters
- dedicated Level1 τ trigger
- cleaning for out-of-time cosmic events or sporadic noise effects in the calorimeters
- at least one vertex reconstructed with more than four tracks
- at least two τ candidates with $|\eta| < 2.5$ and one with $p_T > 30$ GeV and other with $p_T > 15$ GeV
- $\Delta\phi$ between two τ candidates to be at least 2.7 rad (ϕ - the angle in the plane transverse to the beam pipe).

Tau reconstruction in ATLAS

- **There are two algorithms for τ reconstruction in ATLAS experiment: track-seeded and calorimeter-seeded**
- **τ candidates reconstructed by *calo-seeded* algorithm were used in this analysis**
  - ***Calorimeter-seeded*** candidates consist of calorimeter jets reconstructed with *anti-k_t* algorithm (D = 0.4) starting from topological clusters
    - candidate p_T > 10 GeV
    - energy calibrated using global cell energy-density weighting (GCW) calibration scheme

Reconstruction of tau candidate provides very little rejection against QCD jet backgrounds – identification methods are needed

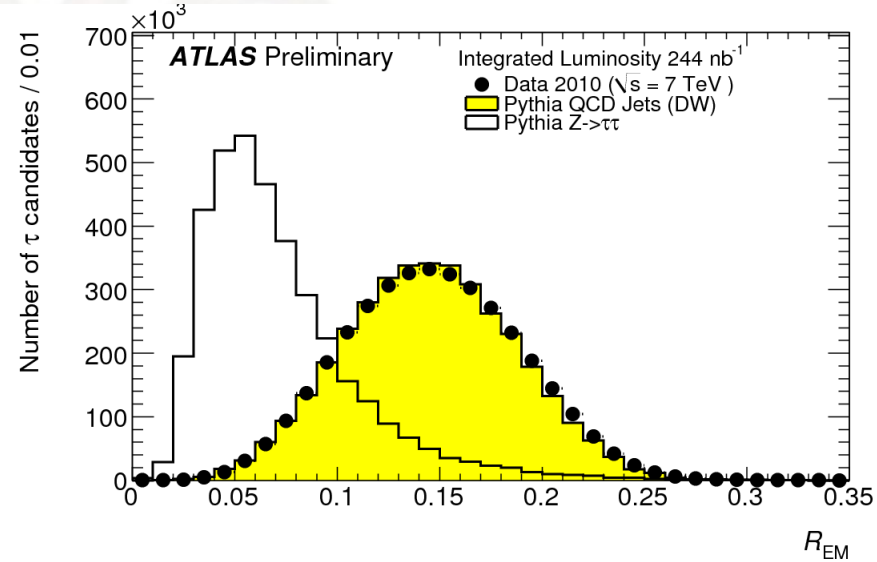


Transverse momentum of tau candidates.

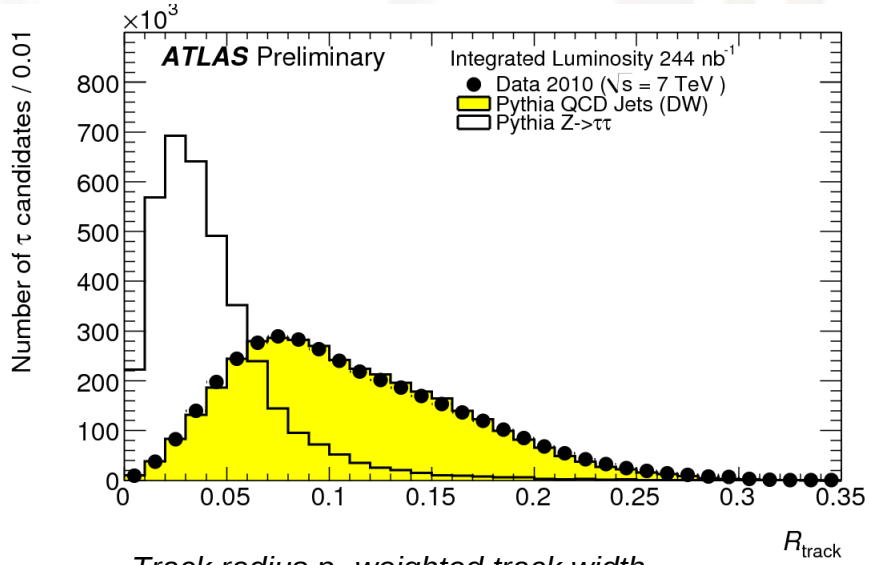
On all plots the number of tau candidates in MC samples are normalised to the number of tau candidates selected in data.

Tau identification in ATLAS

- identification methods for τ candidates include selections based on *simple cuts* and multivariate methods: *boosted decision trees (**BDT**) projective likelihood (**LL**)*
- there are several different classes of discriminating variables employed in the tau ID algorithms, can be grouped to:
  - Shower width (shower radius and isolation)
  - Particle multiplicity (e.g. number of tracks, clusters)
  - Shower composition (e.g. EM fraction)
- relatively long list of variables which could be considered to use in ID methods is aviable
- for early data chosen variables which are robust and relatively uncorrelated with each other

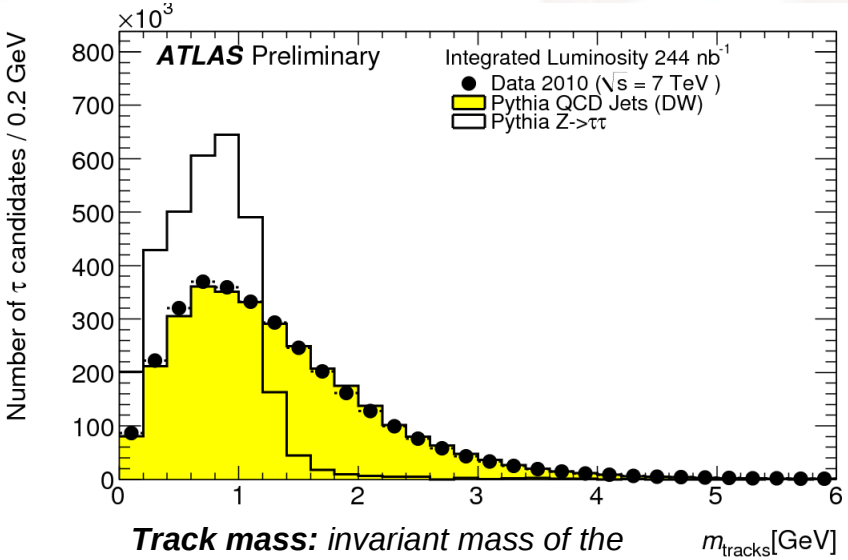


EM radius: transverse energy weighted shower width in the electromagnetic calorimeter

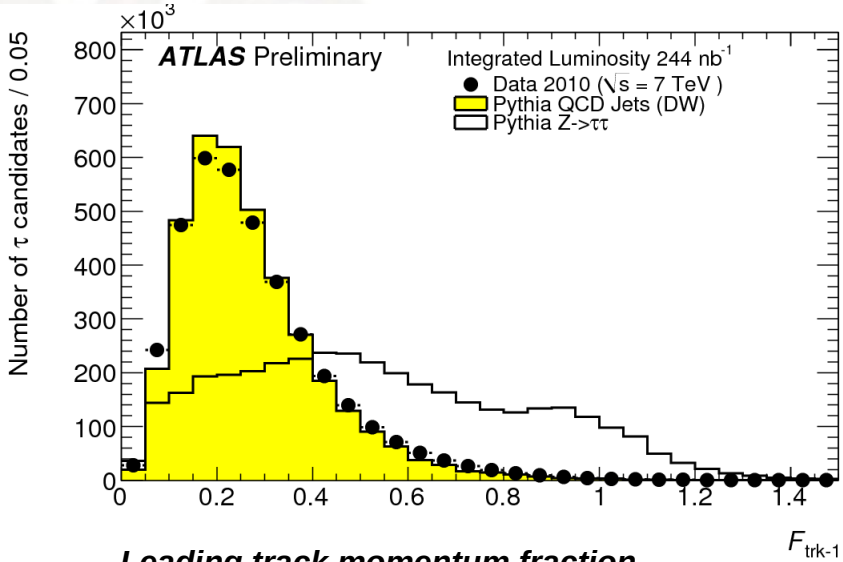


Track radius p_T weighted track width

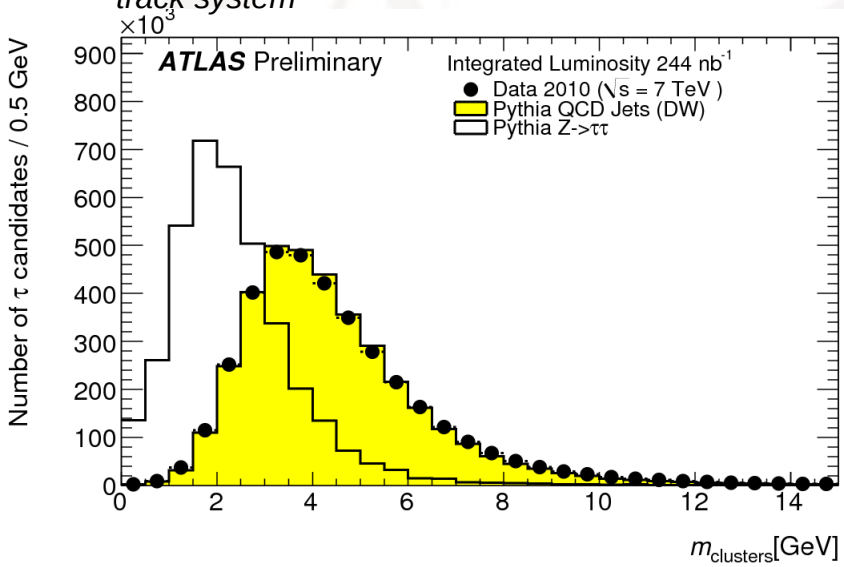
Tau ID variables distribution



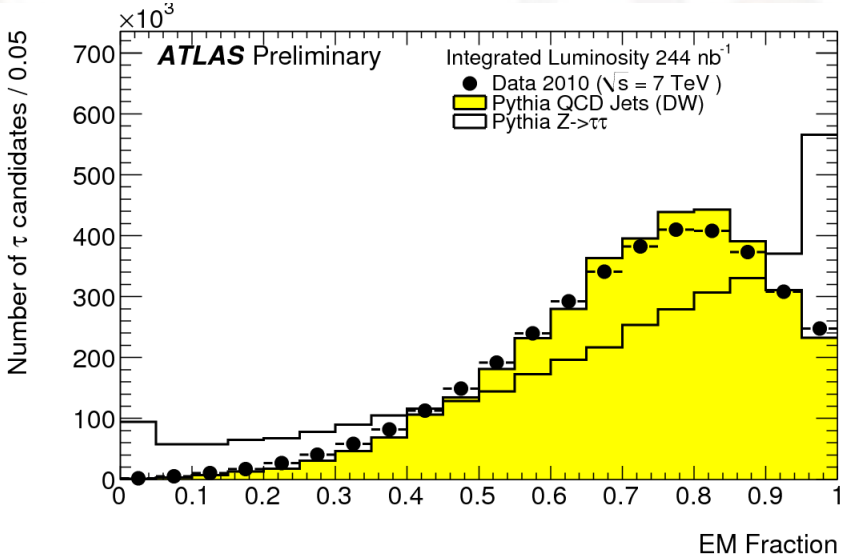
Track mass: *invariant mass of the track system*



Leading track momentum fraction



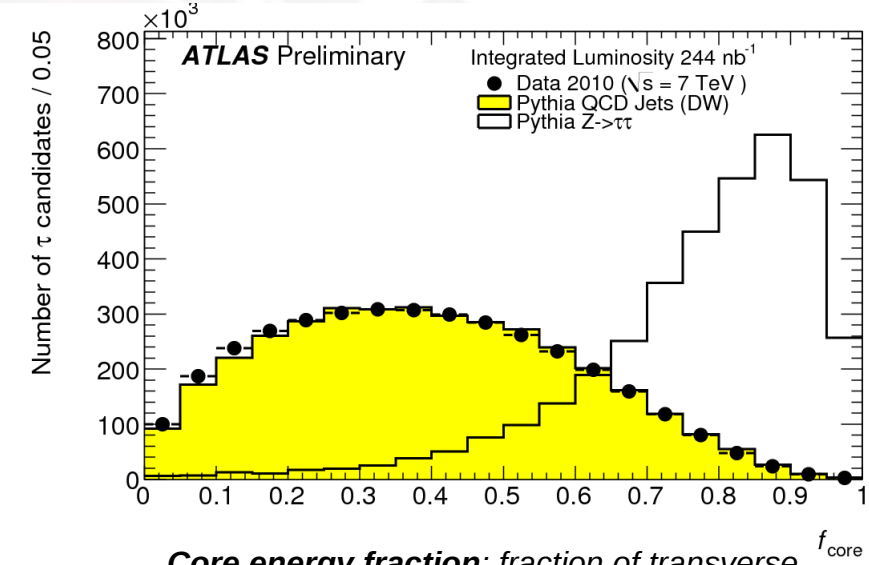
Cluster mass*: invariant mass computed from associated topoclusters.*



Electromagnetic fraction*: fraction of global cell weighting (GCW) calibrated transverse energy of the t candidate deposited in the EM calorimeter*

Tau ID variables distribution cont'd

- for *cut based optimisation* only three of them are used: R_{EM}, R_{track}, $f_{trk,1}$
- *boosted decision trees* uses all of presented variables
- *projective likelihood* method uses all instead of f_{core}



***Core energy fraction**: fraction of transverse energy in the core ($\Delta R < 0.1$) of the τ candidate*

- **All variables well described by MC predictions**
- **Statistical errors on the MC negligible**
- ***Identification variables* used for these methods have shown good tau/jet separation potential in MC studies**

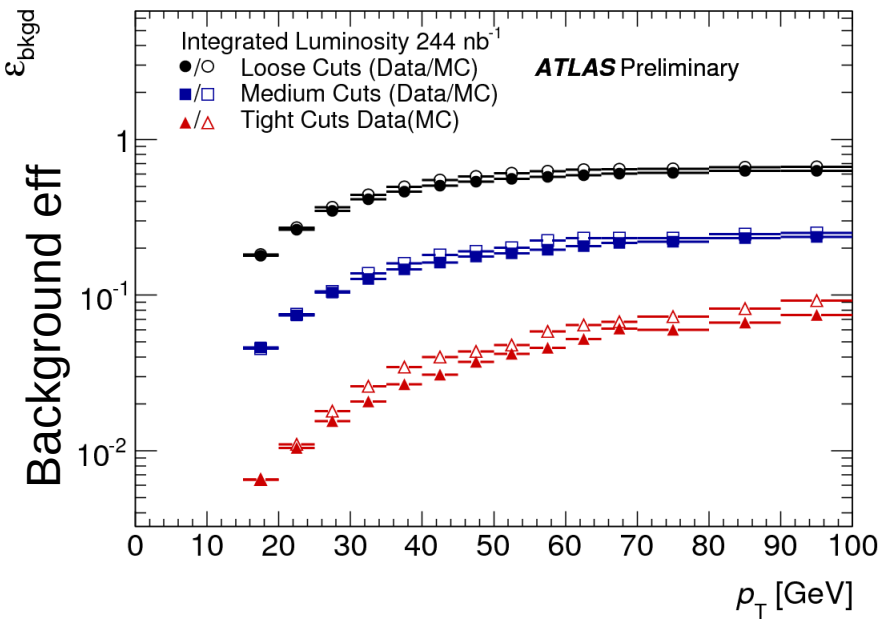
Cut based identification performance

Tau ID performance is defined by:

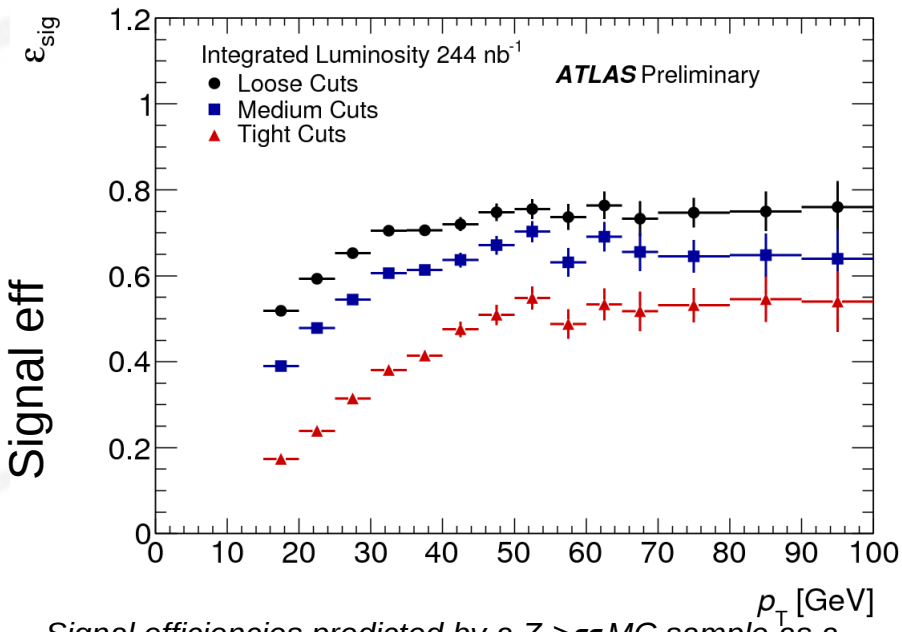
- Signal efficiency $\epsilon_{sig} = \frac{N^{\tau}_{pass,match}}{N^{\tau}_{match}}$
- Background efficiency $\epsilon_{bkgd} = \frac{N^{bkgd}_{pass}}{N^{bkgd}_{total}}$

Cut based identification:

- Uses three variables: R_{EM}, R_{track}, $f_{trk,1}$
- Tuned for three efficiency tresholds
  - loose (eff 30%), medium (50%) and tight (60%)
- Different cuts applied for tau candidates with n_{track} =1 and n_{track} > 1



Background efficiencies for data and MC as function of the reconstructed taus p_T



Signal efficiencies predicted by a Z->ττ MC sample as a function of the reconstructed taus p_T

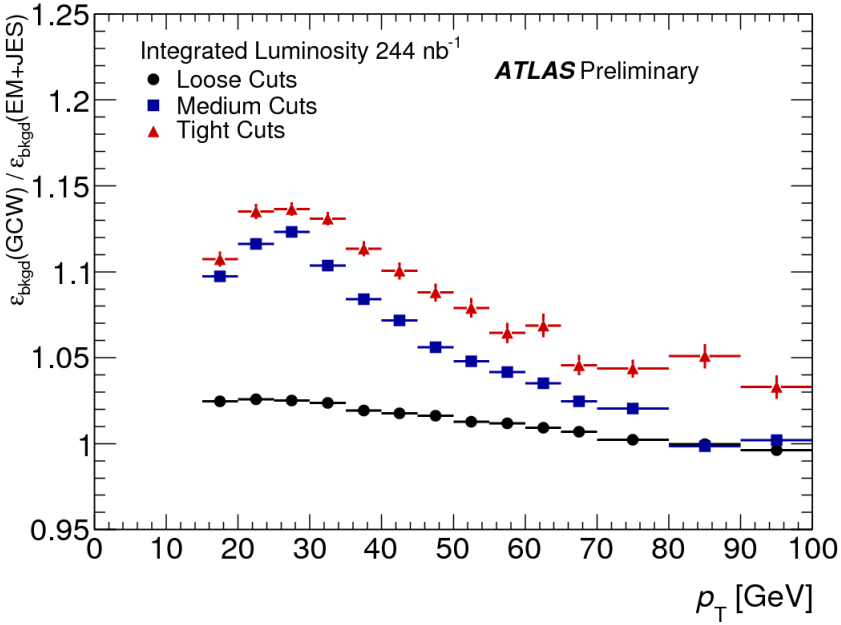
Systematic uncertainties

Systematic uncertainties from the transverse momentum calibration for τ candidates

Variation of the background efficiency was studied by comparing the calibration of τ candidates using global cell energy weighting (GCW) scheme with the simple p_T and η dependent calibration (EM+JES) scheme.

This calibration affects the reconstruction of three of the seven identification variables: $m_{clusters}$, f_{EM} and $f_{trk,1}$

Only $f_{trk,1}$ is used in cut based identification



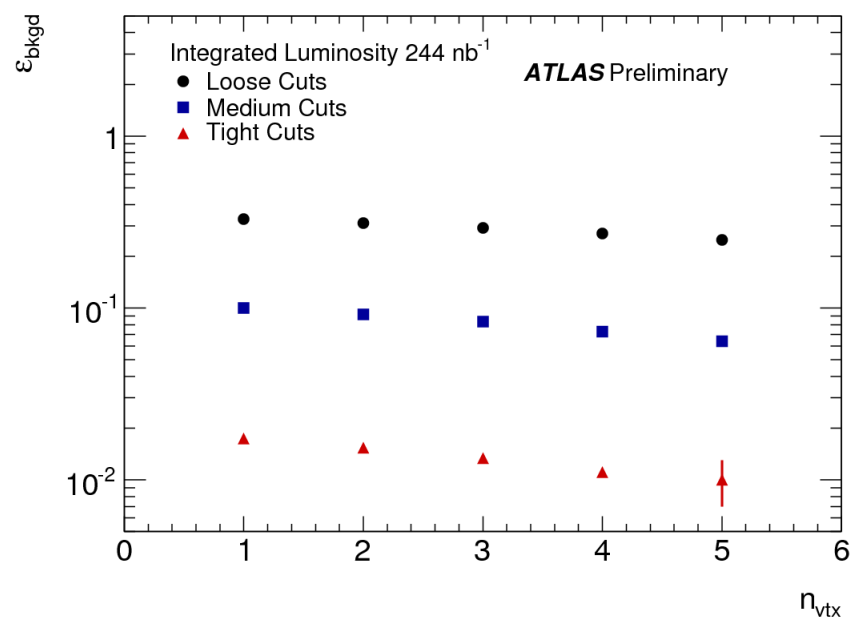
Ratio of background efficiencies using EM+JES and GCW calibrations as a function of reconstructed τ p_T

Systematic uncertainties ..

- **Systematic uncertainties form pile-up effects due to varying beam conditions**

Increase of beam intensities during data taking lead to different pile-up conditions that affect the distributions of the ID variables.

Since number of vertices n_{vtx} is highly correlated with pile-up activity, the background efficiency was evaluated as a function of n_{vtx}



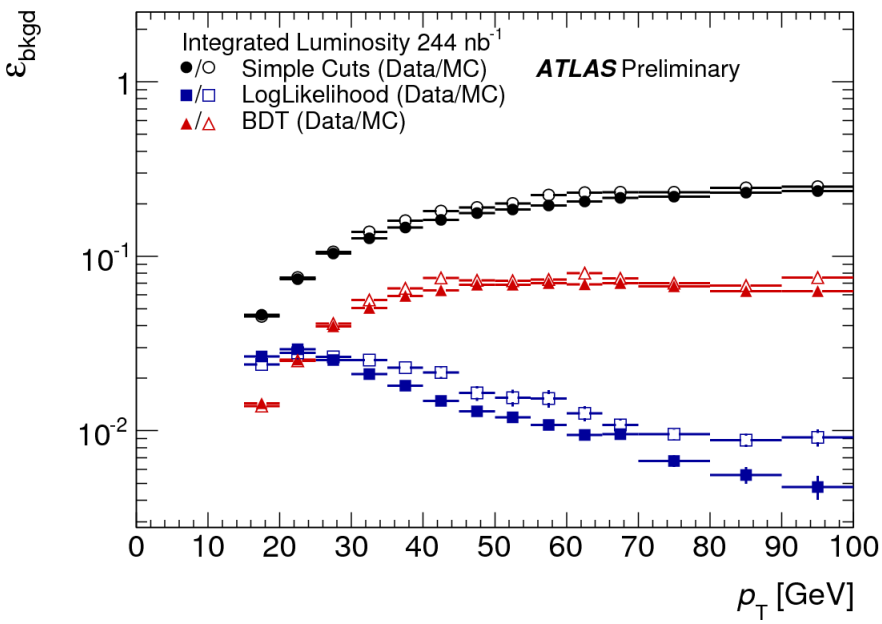
Background efficiencies as a function of n_{vtx}

- **A systematic uncertainty is determined by taking the mean difference of the background efficiency for candidates in events with n_{vtx} = 1 and n_{vtx} > 1 with the background efficiencies obtained from the entire sample.**

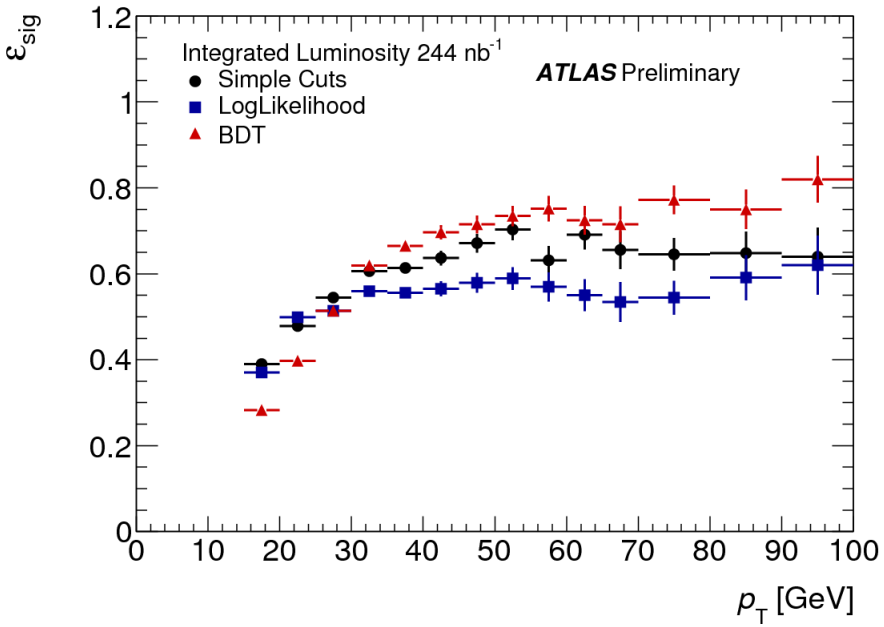
- **Other sources of systematic uncertainties such as beam spot variations, the impact of calorimeter noise, and detector alignment effects are still investigated..**

Tau ID performance: comparison with multivariate methods

- multivariate τ identification discriminants such as boosted decision trees BDT and projective likelihood LL methods were used



Background efficiencies *in data and MC as a function of pT with the* ***medium*** *selection for cut-based, BDT and LL identification*

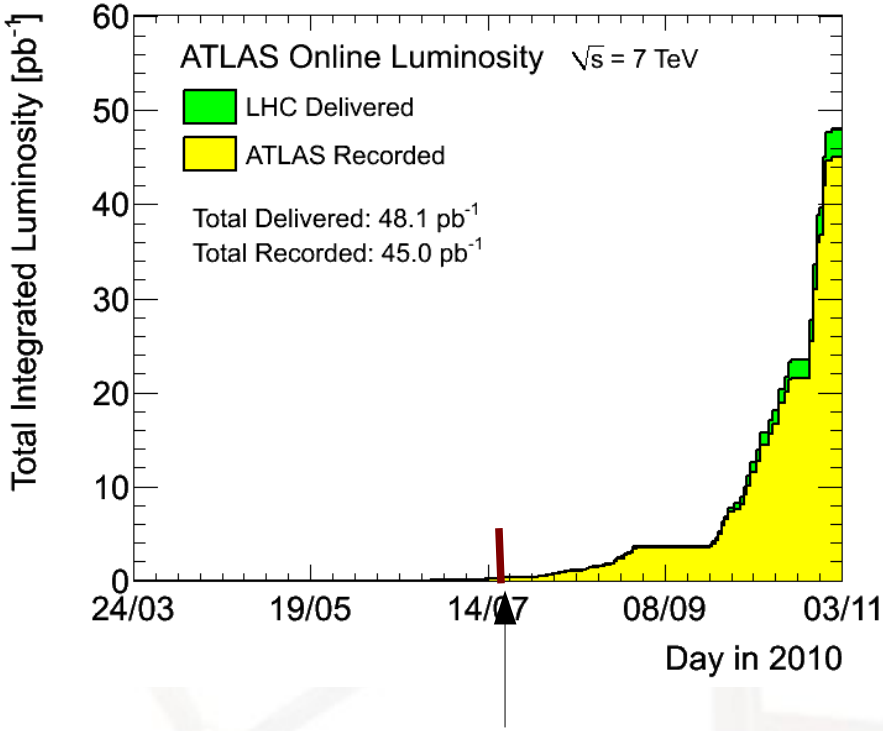


Signal efficiencies *from MC as a function of the transverse momentum of tau candidates with the* ***medium*** *selection for cut-based, BDT, and LL identification.*

BDT and LL identification algorithms increase background rejection power significantly better than cut-based method

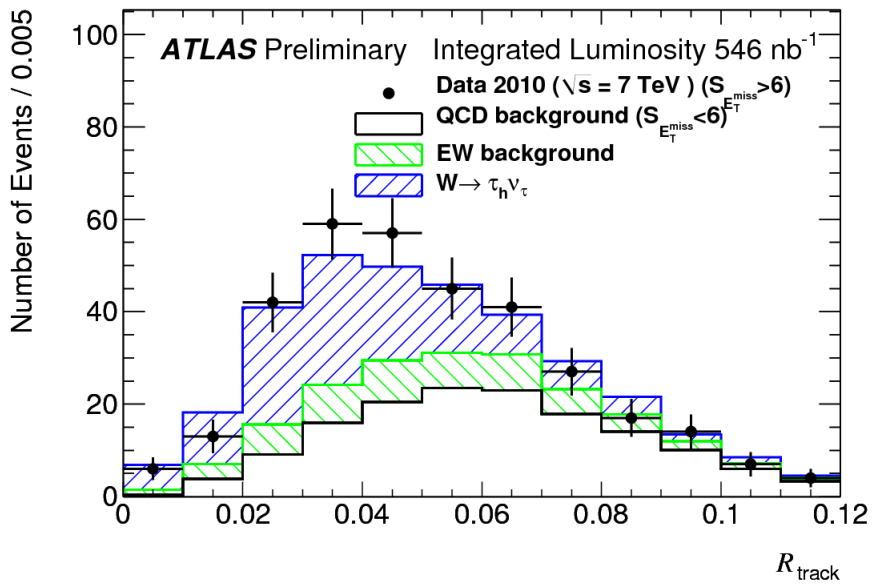
Summary

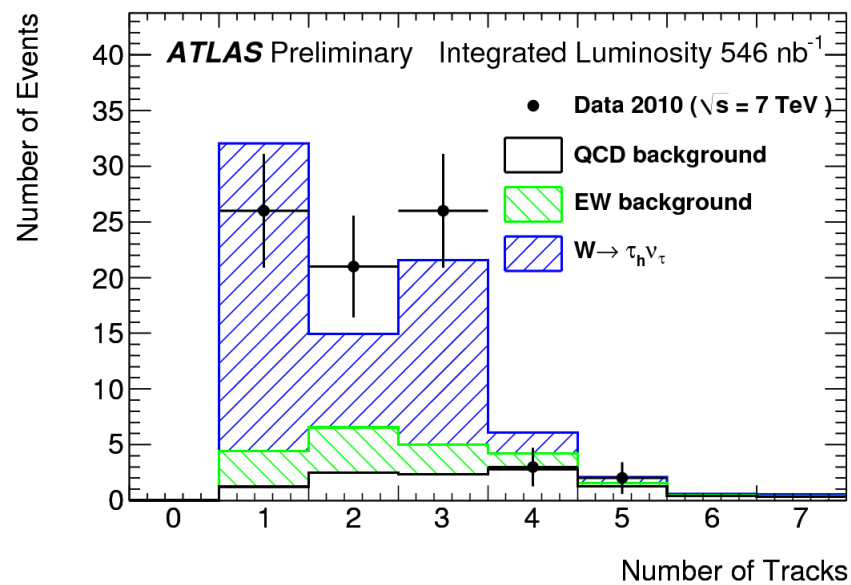
- All variables well described by MC predictions
- Discriminating variables show good separation power between τ leptons and fake τ candidates from QCD jets
- Background efficiency for three cut-based selections was measured - good agreement with MC predictions
- Data and MC predictions show that BDT and LL identification algorithms increase background rejection power significantly better then cut-based ID
- Performance of the algorithms on real τ leptons to come soon. Many studies ongoing on real taus coming from W and Z decays



presented results

Backup

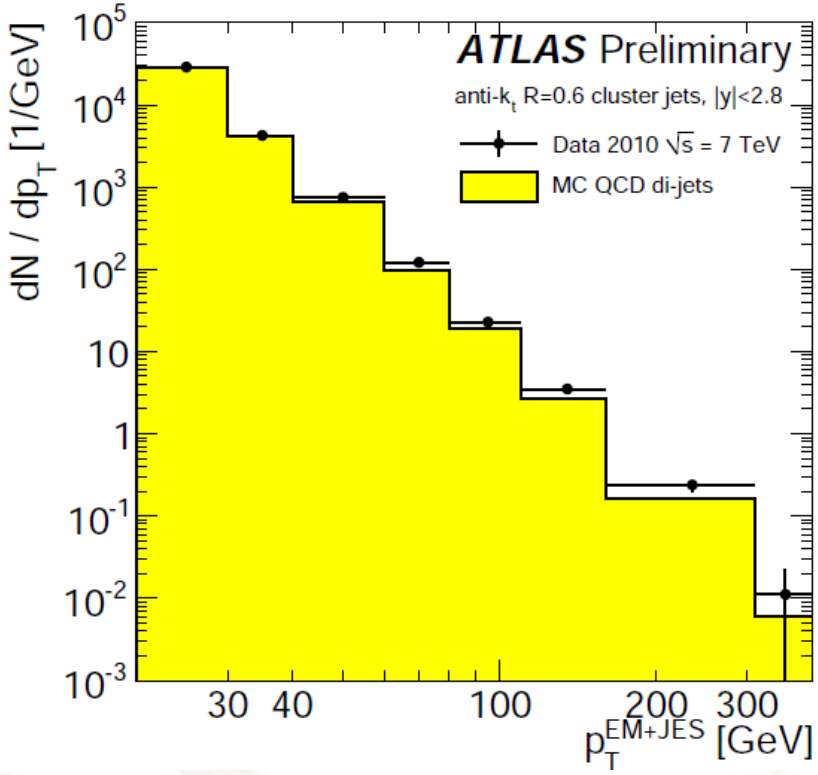

ATLAS Preliminary
Integrated Luminosity 546 nb^{-1}
Data 2010 ($\sqrt{s}$ = 7 TeV) ($S_{E_T^{miss}}$>6)
QCD background ($S_{E_T^{miss}}$<6)
EW background
$W\rightarrow \tau_h \nu_\tau$
Number of Events / 0.005
R_{track}


ATLAS Preliminary
Integrated Luminosity 546 nb^{-1}
Data 2010 ($\sqrt{s}$ = 7 TeV)
QCD background
EW background
$W\rightarrow \tau_h \nu_\tau$
Number of Events
Number of Tracks

EM+JES calibration

Simple p_T and h-dependent calibration scheme (EM+JES calibration).

This simple calibration scheme corrects for the non-linear correlation between the energy reconstructed in the calorimeter and the energy of the particles forming jets. Jets are found from clusters or towers at the electromagnetic scale and the calibration constants are applied as a function of the uncalibrated jet p_T and η. The calibration constants can be calculated from the Monte Carlo simulation, or from data using γ+jet and di-jet balance techniques.

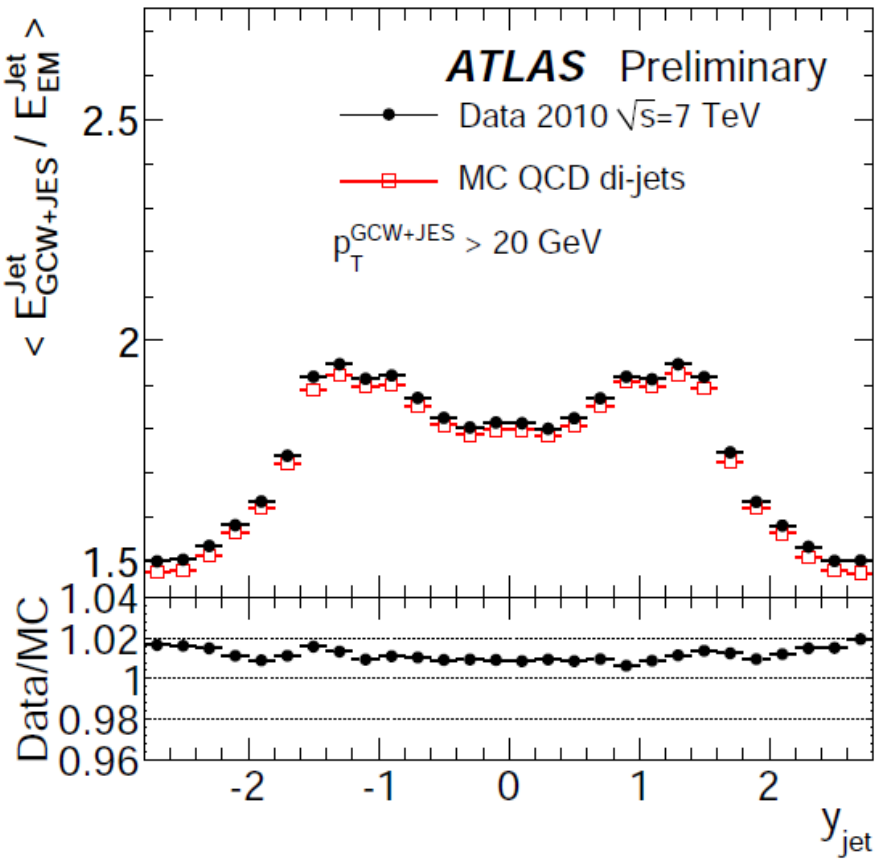


Jet p_T distribution using the EM+JES calibration scheme

GCW calibration

Global cell energy-density weighting calibration scheme (GCW calibration).

This calibration scheme attempts to compensate for the different calorimeter response to hadronic and electromagnetic energy depositions. The hadronic signal is characterized by low cell energy densities and, thus, weighted up. The weights, which depend on the cell energy density and the calorimeter layer only, are determined by minimizing the energy fluctuations between the reconstructed and particle jets in Monte Carlo simulation. The weights also compensate for energy losses in the dead material. Jets are found from uncalibrated clusters or towers, then cells are weighted and a final p_T and η-dependent correction is added to ensure that the jet energy is properly reconstructed.



the ratio of jet energies before and after the application of the GCW+JES calibration scheme

Tau identification variables

- $m_{clusters}$ **cluster mass**: invariant mass computed from associated topoclusters
- m_{tracks} **track mass**: invariant mass of the track system
- R_{track} **track radius**: p_T weighted track width $$R_{track} = \frac{\sum_i^{\Delta R_i < 0.2} p_{T,i} \Delta R_i}{\sum_i^{\Delta R_i < 0.2} p_{T,i}}$$
- $f_{trk,1}$ **leading track momentum fraction** $$f_{trk,1} = \frac{p_{T,1}^{track}}{p_T^{\tau}}$$
- R_{EM} **electromagnetic radius**: transverse energy weighted shower width in the eletromagnetic (EM) calorimeter $$R_{EM} = \frac{\sum_i^{\Delta R_i < 0.4} E_{T,i}^{EM} \Delta R_i}{\sum_i^{\Delta R_i < 0.4} E_{T,i}^{EM}}$$
- f_{core} **core energy fraction**: fraction of transverse energy in the core ($\Delta R <$ 0.1) of the τ candidate
- f_{EM} **electromagnetic fraction**: fraction of GCW calibrated transverse energy of the t candidate deposited in the EM calorimeter