Estimation of backgrounds from jets misidentified as τ-leptons using the Universal Fake Factor method with the ATLAS detector

Federico Morodei on behalf of the ATLAS Collaboration

Introduction

τ-leptons can decay leptonically (35% branching fraction) or hadronically (65%). In hadronic τ-decays, only the visible decay products are reconstructed ($\tau_{\text{had-vis}}$). In analyses with $\tau_{\text{had-vis}}$ in final state, jets misidentified as $\tau_{\text{had-vis}}$ (fake $\tau_{\text{had-vis}}$) are a sizable background source. Monte Carlo (MC) generators cannot properly model this background and data-driven methods are necessary. The Fake Factor (FF) method is a data-driven technique to estimate the fake $\tau_{\text{had-vis}}$ background in a signal region (SR). Transfer factors (FF) are measured in dedicated control regions (CR) and applied to the antiID SR (i.e. with same cuts as SR but inverted $\tau_{\text{had-vis}}$ identification requirement).

$$N^{\text{fake}}_{\text{ID SR}} = \left(N^{\text{data}}_{\text{antiID SR}} - N^{\text{real}}_{\text{antiID SR}}\right) \times \text{FF} \qquad \text{FF} = \frac{N^{\text{data}}_{\text{ID CR}} - N^{\text{real}}_{\text{ID CR}}}{N^{\text{data}}_{\text{antiID CR}} - N^{\text{real}}_{\text{antiID CR}}}$$

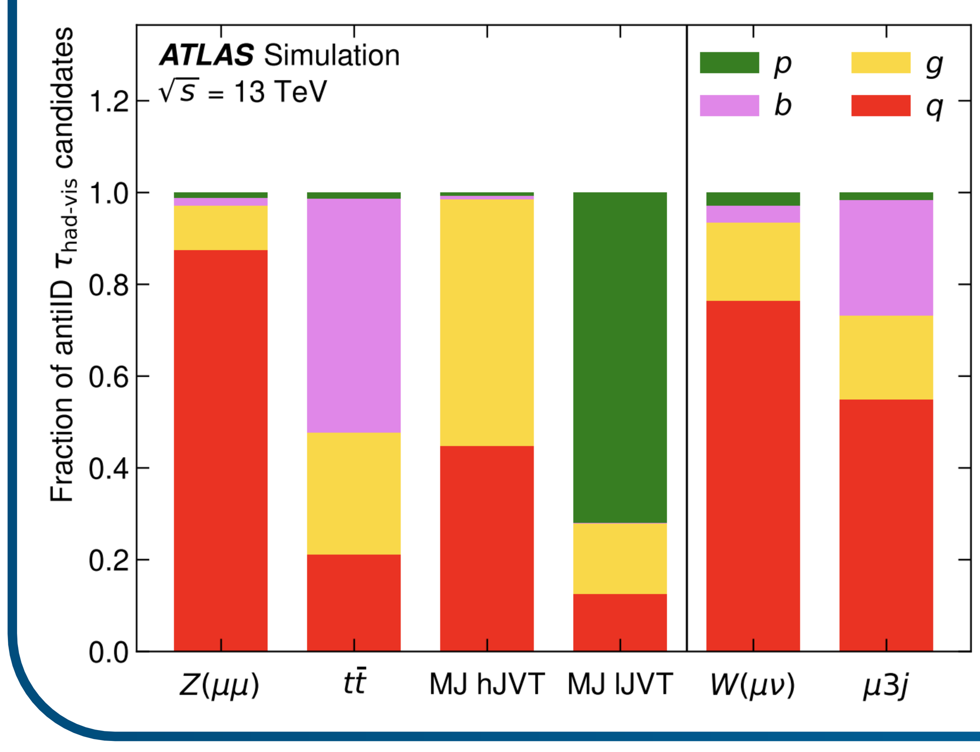
4 sources of fakes $\tau_{\text{had-vis}}$:

- Jets from b-quark (b)
- Jets from light quark (q)
- Jets from gluon (g)
- Jets from pile-up (p)

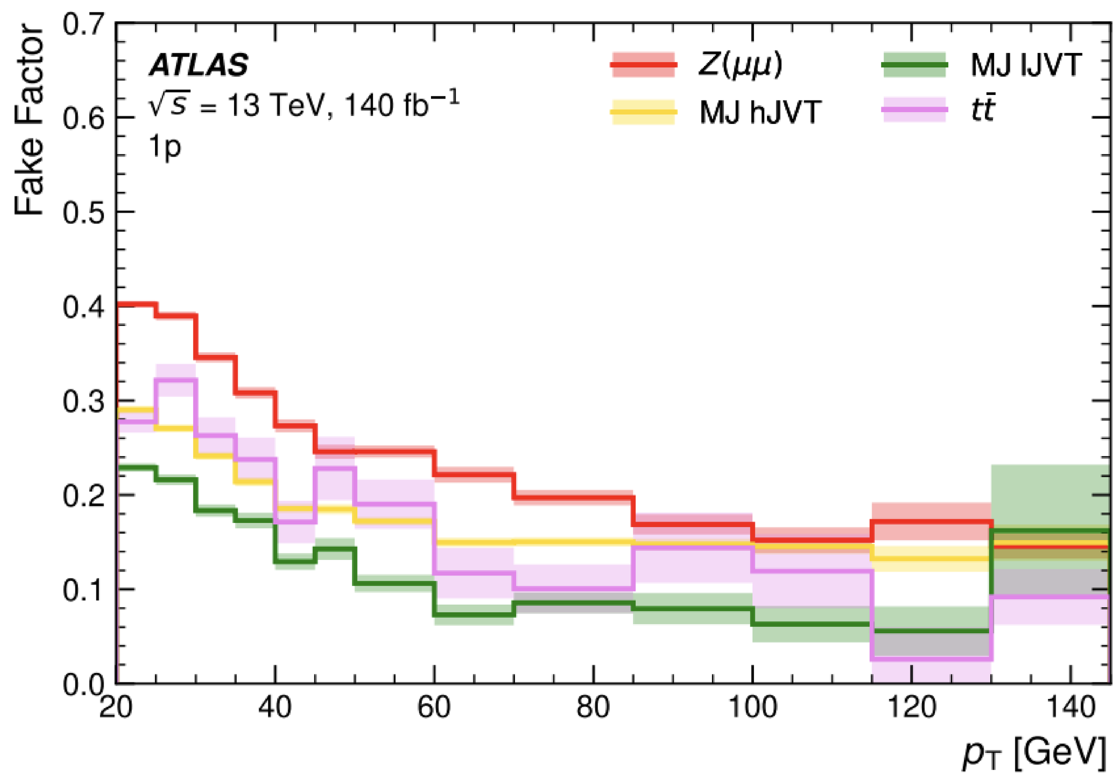
Universal Fake Factor method

The Universal Fake Factor (UFF) method [1] generalizes the FF method. 4 regions are defined, each enriched in one category of fake $\tau_{\text{had-vis}}$:

- $Z(\mu\mu)$ region enriched in q fake $\tau_{\text{had-vis}}$
- $t\bar{t}$ region enriched in b fake $\tau_{\text{had-vis}}$
- High-JVT multijet (MJ hJVT) region enriched in g fake $\tau_{\text{had-vis}}$
- Low-JVT multijet (MJ lJVT) region enriched in p fake $\tau_{\text{had-vis}}$



Fake Factors in the $Z(\mu\mu)$, $t\bar{t}$, MJ hJVT and MJ lJVT regions ▼

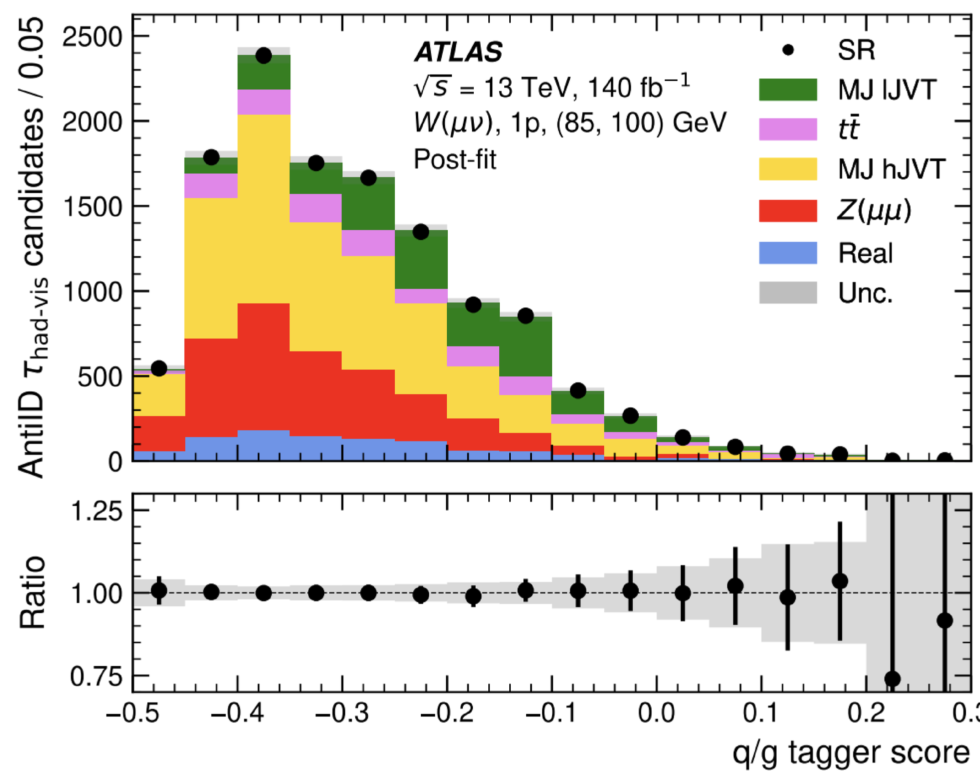


FFs are measured in each of the 4 regions in bins of p_{T} and number of tracks of $\tau_{\text{had-vis}}$. In each bin and each antiID region, a template of ATLAS q/g tagger score [2] is built. The 4 templates are fitted to the analysis antiID SR to estimate the mixture of $Z(\mu\mu)$, $t\bar{t}$, MJ hJVT and MJ lJVT regions contributing to the antiID SR.

The combined FF suitable for the analysis SR is computed as the weighted average of FFs measured in the $Z(\mu\mu)$, $t\bar{t}$, MJ hJVT and MJ lJVT regions, with coefficients extracted from a template fit:

$$\text{FF}_{\text{UFF}} = \mu_{Z(\mu\mu)} \times \text{FF}_{Z(\mu\mu)} + \mu_{t\bar{t}} \times \text{FF}_{t\bar{t}} + \\ + \mu_{\text{MJ hJVT}} \times \text{FF}_{\text{MJ hJVT}} + \mu_{\text{MJ lJVT}} \times \text{FF}_{\text{MJ lJVT}}$$

Template fit result for $W(\mu\nu)$ region as SR ▼



Uncertainties

Statistical uncertainty

Total FF_{UFF} statistical uncertainty given by uncorrelated FF stat uncertainties and correlated uncertainties on fit coefficients:

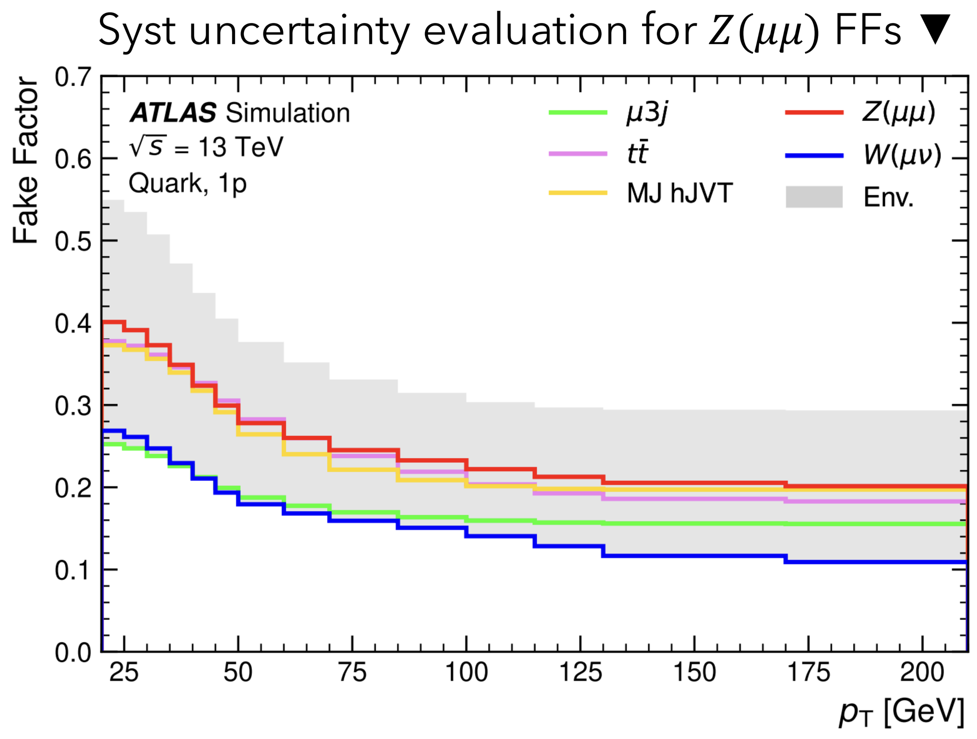
$$\sigma^2_{\text{UFF, stat.}} = \sum_X \mu_X^2 \sigma^2_{\text{FF}_X} + \sum_{X,Y} \text{FF}_X \text{FF}_Y V_{XY}$$

with X and Y running over the 4 regions and V being the Hessian covariance matrix.

Systematic uncertainty

In each region, FFs for $q/g/b/p$ fake $\tau_{\text{had-vis}}$ are evaluated using MC samples. Also two additional regions, $W(\mu\nu)$ and $\mu 3j$ ($W(\mu\nu)$ events with at least 3 jets) are used in this case. A systematic uncertainty is assigned to $Z(\mu\mu)$ FFs using the differences between $Z(\mu\mu)$ MC FFs and the FFs for q fake $\tau_{\text{had-vis}}$ calculated in the additional regions. The same is done for the other FFs using g, b or p fake $\tau_{\text{had-vis}}$. Systematic uncertainties are propagated to FF_{UFF} separately for each source.

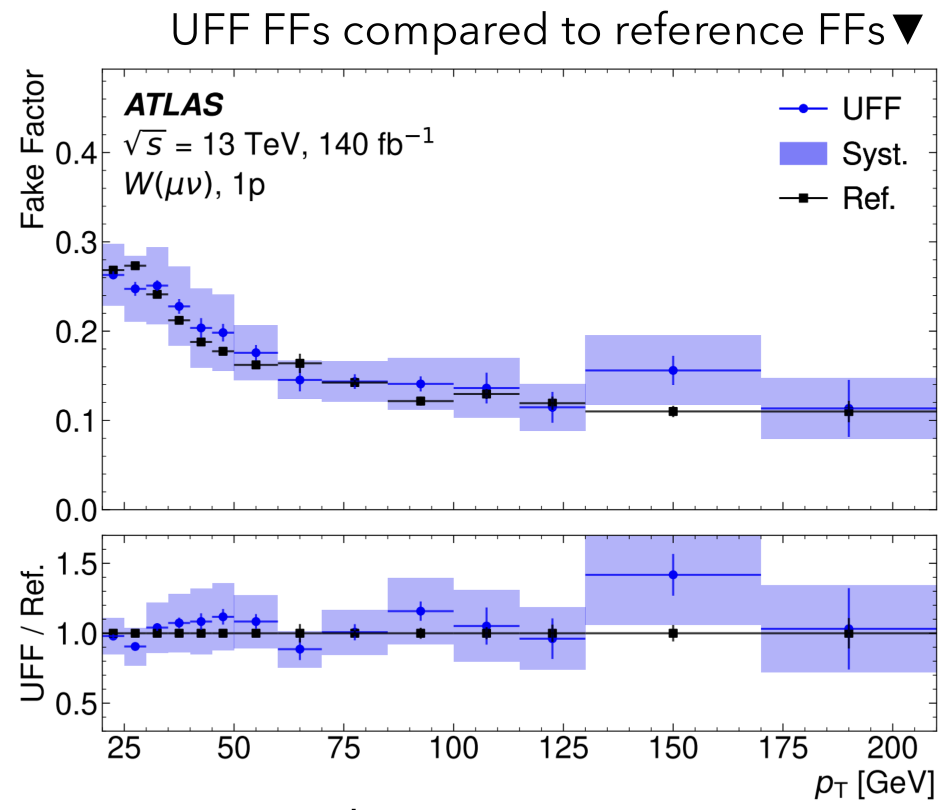
Syst uncertainty evaluation for $Z(\mu\mu)$ FFs ▼

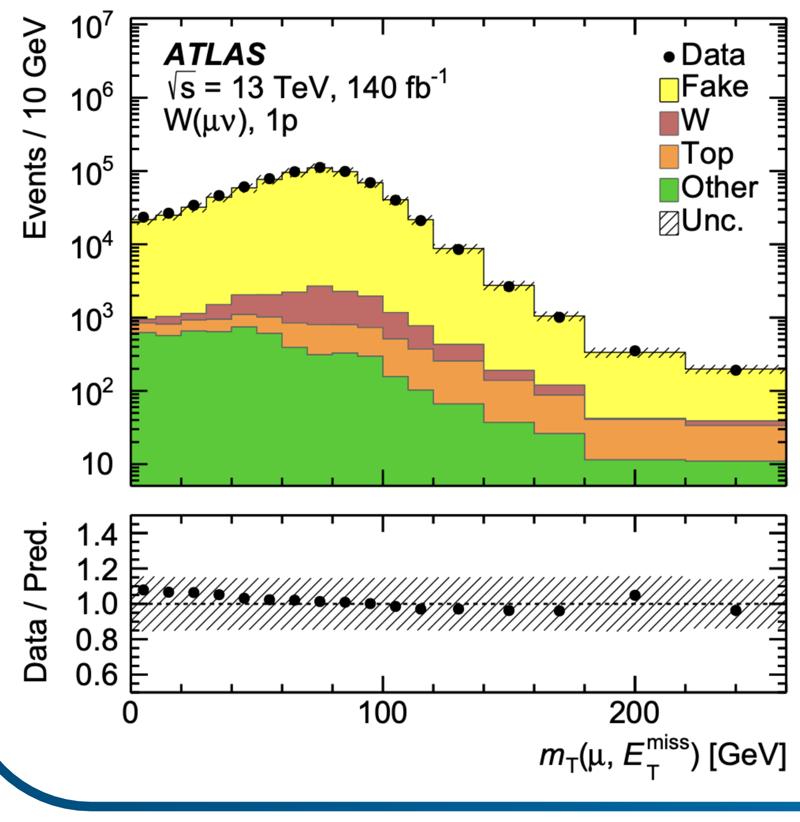


Validation

The $W(\mu\nu)$ region is used to validate the UFF method. A comparison between UFF FFs and reference FFs (determined directly from ID to antiID ratio) is performed, showing agreement within the uncertainty.

UFF FFs compared to reference FFs ▼





◀ Data/prediction comparison in $W(\mu\nu)$ region

The $W(\mu\nu)$ and $\mu 3j$ ID regions are used to compare the total Standard Model prediction (fake $\tau_{\text{had-vis}}$ from UFF method, other processes from MC) with the data. Agreement is observed within the uncertainty.

The UFF method has the main advantage that FFs suitable for any SR can be obtained from a template fit and do not have to be measured in dedicated control regions. Hence, such regions can be exploited to validate the background prediction in the user analysis. The UFF method has been successfully employed by ATLAS analyses with both one [3] and two [4,5] $\tau_{\text{had-vis}}$ in the final state.

[1] arXiv:2502.04156, [2] arXiv: 2308.00716, [3] arXiv: 1807.07915, [4] arXiv: 2305.12938, [5] arXiv:2503.19836