Universal Fake Factor method to estimate hadronic tau backgrounds in ATLAS

arXiv:2502.04156v1

Gabriela Martinovicova

on behalf of the ATLAS Collaboration

ICNFP 2025, 17-31 July 2025

ATLAS EXPERIMENT

CHARLES UNIVERSITY

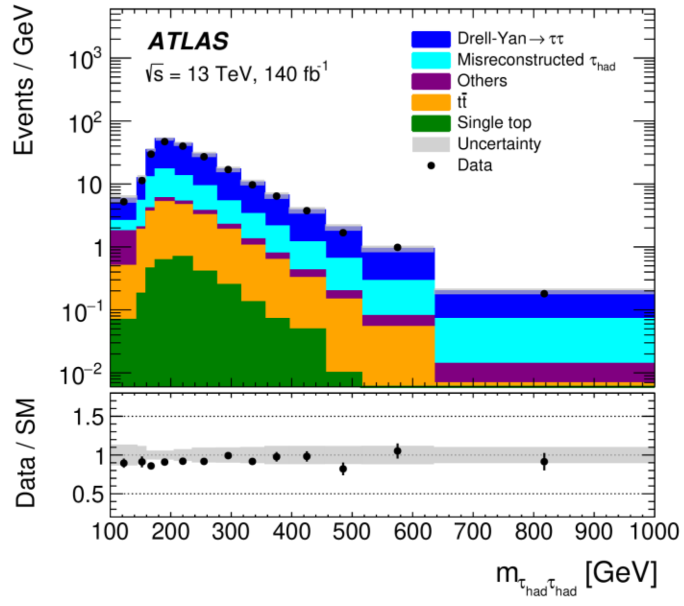
Talk Content

- Introduction of the Universal Fake Factor (UFF) method
- Treatment of UFF systematic uncertainties
- Validation in topical physics regions

Motivation for Fake Tau background estimation

- Hadronic taus τ_{had} appear in many key physics processes:
  - They are crucial for both searches and precision measurements
  - e.g., Higgs analyses ($H -> \tau\tau, HH -> bb\tau\tau$), Standard Model measurements, new physics searches (SUSY, leptoquarks,...)
- Neutrinos are no detected → only hadrons can be reconstructed as visible decay products
- A major challenge:
  - **jet misreconstructed as τ-lepton** → dominant background
  - Fake background not reliably modelled by MC, data driven estimation methods are preferred

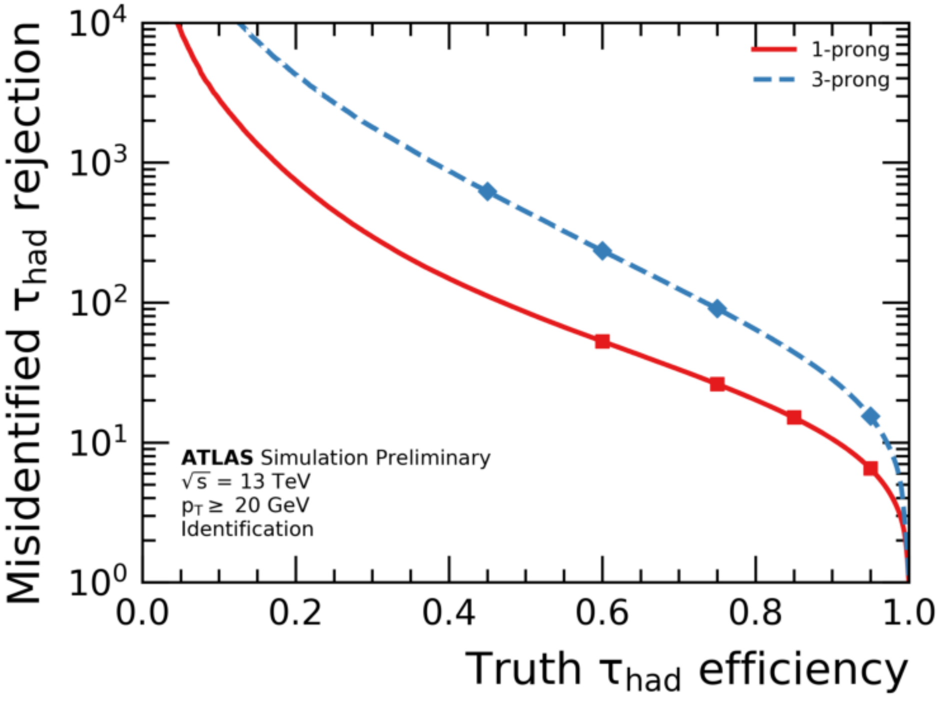
arXiv:2305.12938



Fake τ_{had} Suppression with RNN-based Tau Identification Algorithm

ATL-PHYS-PUB-2022-044

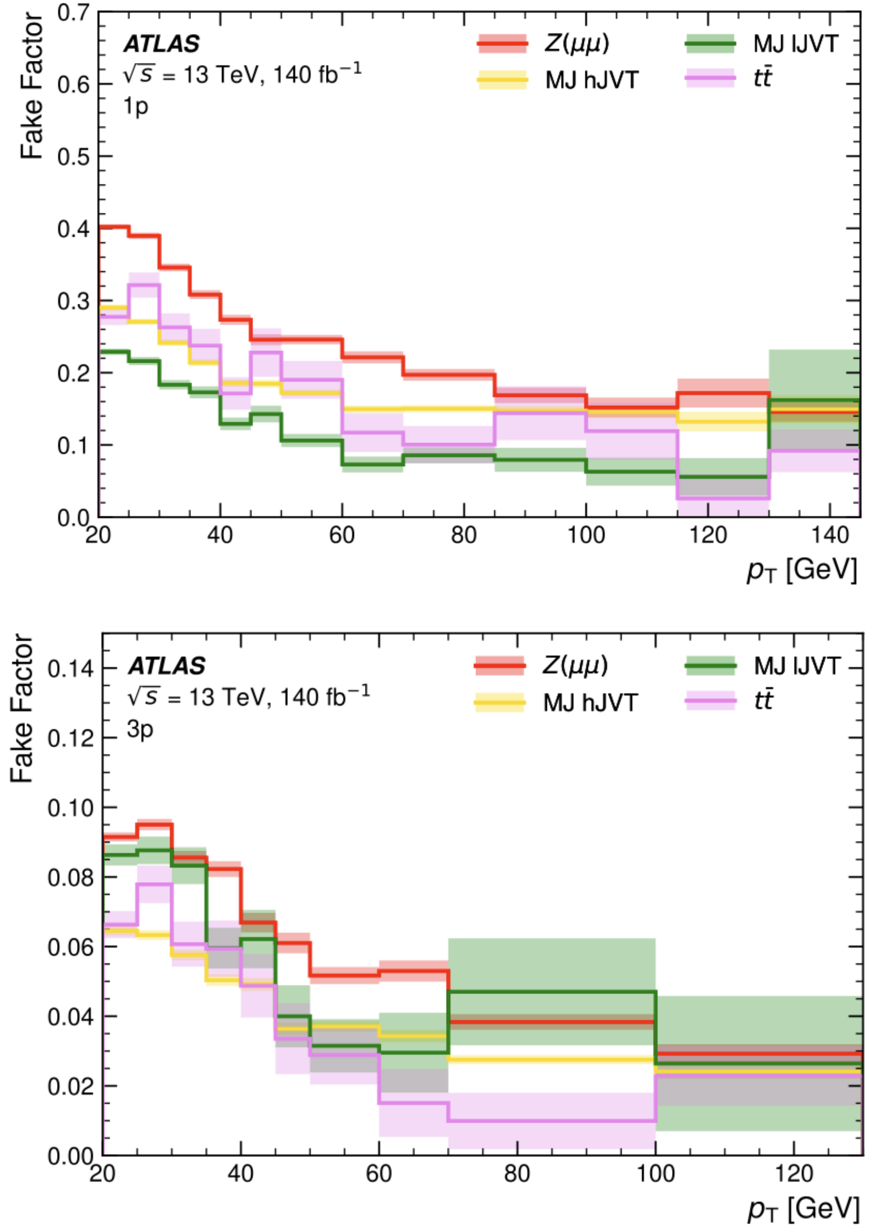
- It combines tracking and calorimeter data for better signal–background separation.
- The ID is highly optimized to suppress fake τ_{had} while keeping real τ_{had}.
- Despite these excellent identification capabilities, some fake τ_{had} inevitably pass the selection criteria



Universal Fake Factor method in a nutshell

FAKE FACTOR METHOD

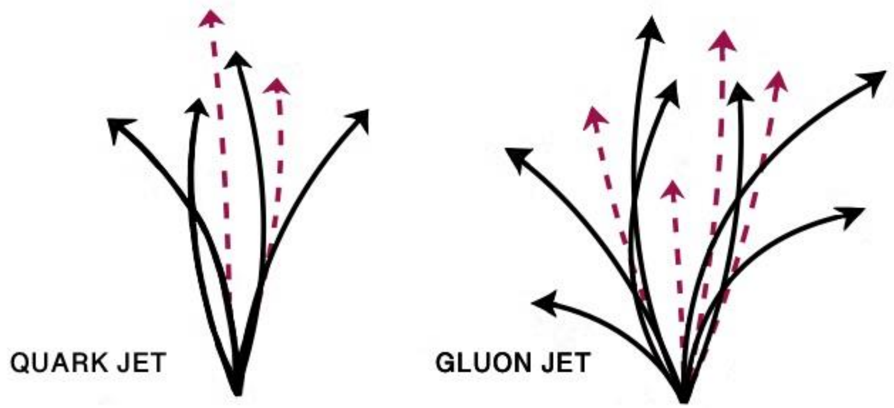
- Data-driven method of the fake τ_{had} background determination
- $N_{\text{ID}}^{SR} = N_{\text{antiID}}^{SR} * FF$
- DIRECT Fake Factor $FF = \frac{N_{\text{ID}}}{N_{\text{antiID}}}$
  - Can/Must be determined in a CR and as function of τ_{had} kinematic
- FF dependent on fake τ_{had} origin



ATLAS EXPERIMENT

Universal Fake Factor Method

- FFs depend on the fake τ_{had} origin
  - Light-quark-initiated jets (*q*)
  - Gluon-initiated jets (*g*)
  - b-quark-initiated jets (*b*)
  - Pile-up jets (*p*)
- need to find a CR with the same composition of *q*-/*g*-/*b*-/*p*-jets originating fake τ_{had} as the SR



UFF method

- determines the fractions μ_x of fake types (*q, g, b, p*) in the anti-ID SR
- using pure *q/g/b/p*-initiated fake taus regions
  - $N^{SR}_{\text{passID}} = N_q * FF_q + N_g * FF_g + N_b * FF_b + N_p * FF_p = N^{SR}_{\text{failID}} * FF_{\text{comb}}$
  - $FF_{\text{comb}} = \mu_q * FF_q + \mu_g * FF_g + \mu_b * FF_b + \mu_p * FF_p$
  - impossible to find regions containing only one type of fake

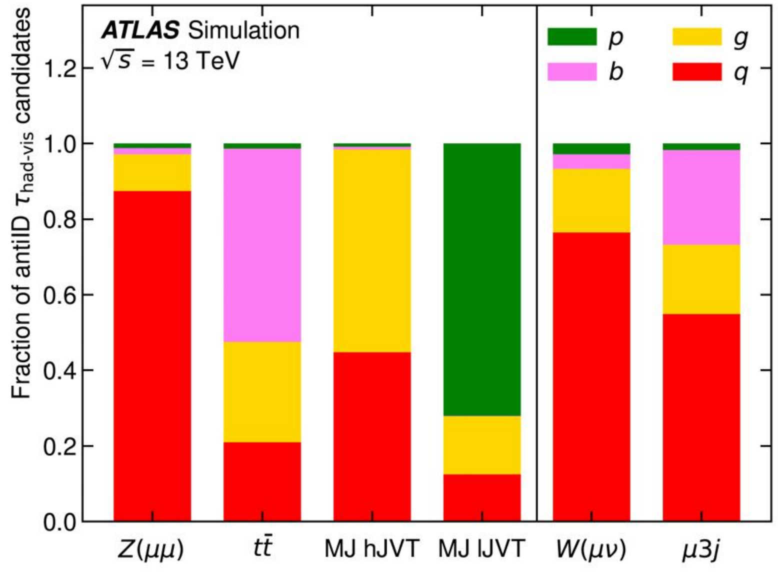
Universal Fake Factor Method

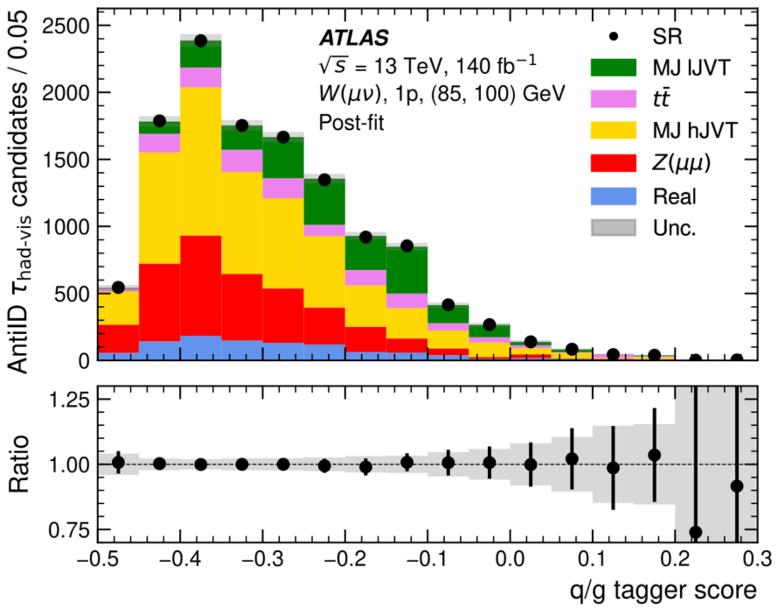
Experimentally:

- Use fake-enriched regions to estimate *q/g/b/p* components for the UFF method
  - $Z(\mu\mu)$ region enriched in q-jet fake τ_{had}
  - MJ hJVT region enriched in g-jet fake τ_{had}
    - Jets consistent with tracks pointing to the primary vertex → high JVT score
  - MJ lJVT region enriched in pile-up fake τ_{had}
    - Jets from pile-up vertices (extra collisions) → low JVT score
  - $t\bar{t}$ regions enriched in b-quark jet fake τ_{had}

$$UFF\ FF = FF_{comb} = FF_{Z(\mu\mu)} * \mu_{Z(\mu\mu)} + FF_{MJ\ hJVT} * \mu_{MJ\ hJVT} + FF_{MJ\ lJVT} * \mu_{MJ\ lJVT} + FF_{t\bar{t}} * \mu_{t\bar{t}}$$

- The fractions μ_x are derived from a fit of anti-ID templates (data) to the anti-ID user spectrum (data)

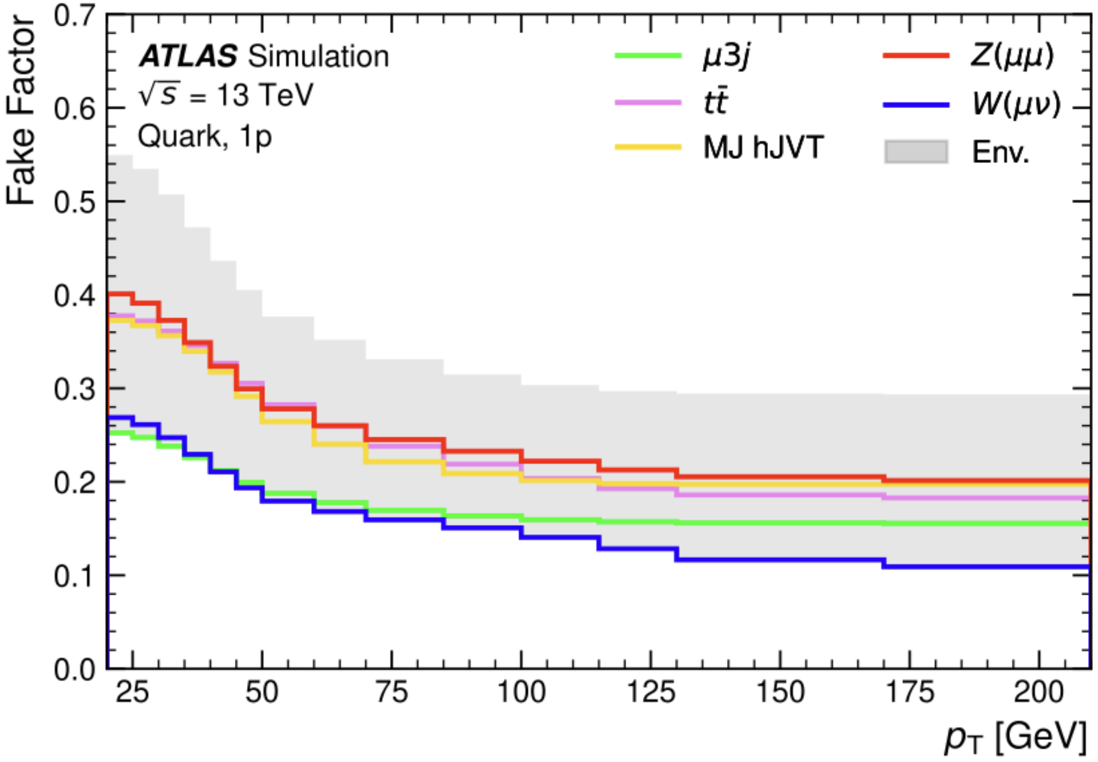




Universal Fake Factor method systematic uncertainties estimate and validation

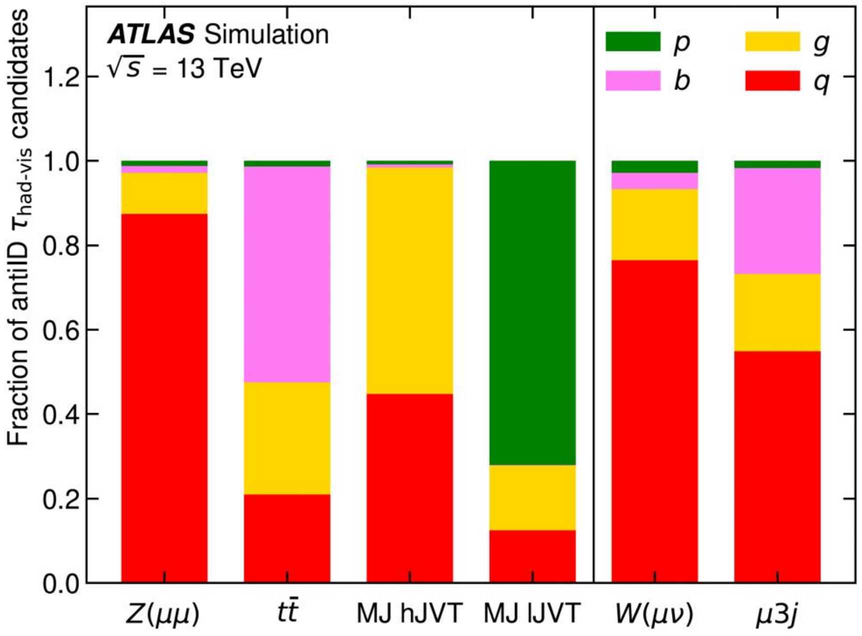
Systematic uncertainty estimate

- UFF assumes *q/g/b/p* fake factors are independent from the production process
- Tested using MC simulation comparing the FFs for $q/g/b/p$ fake τ_{had} in different regions
  - Five regions representing common fake-production processes and kinematic regimes
- Systematic uncertainty given by the difference between FFs measured in these regions



ATLAS EXPERIMENT

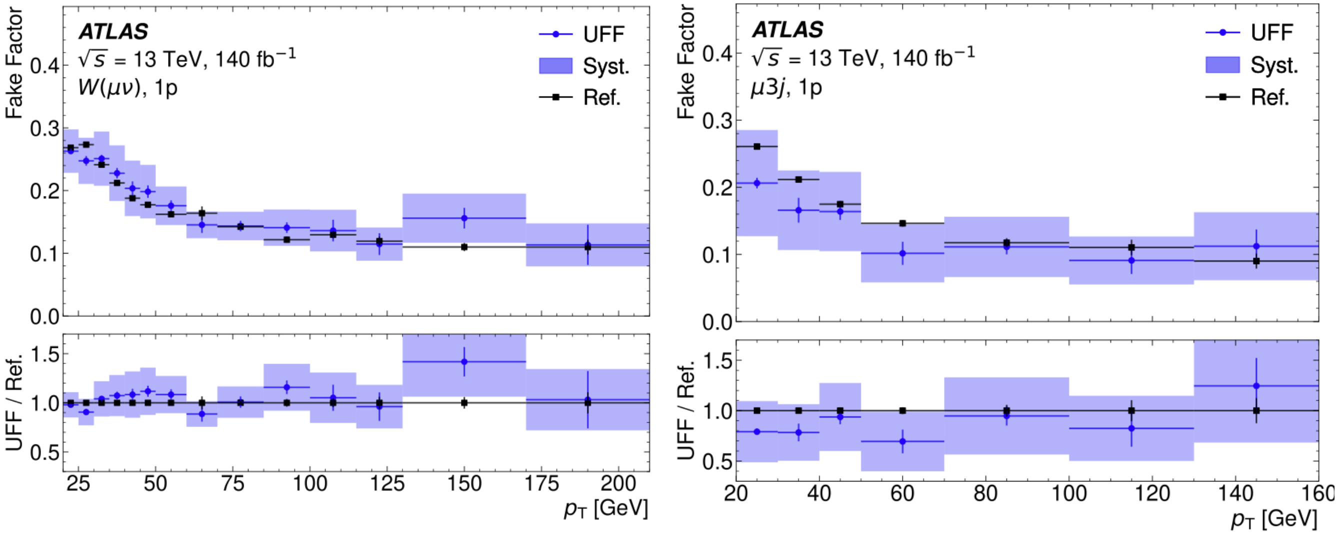
Validation in $W(\mu\nu)$ *and* $\mu 3j$ *regions*

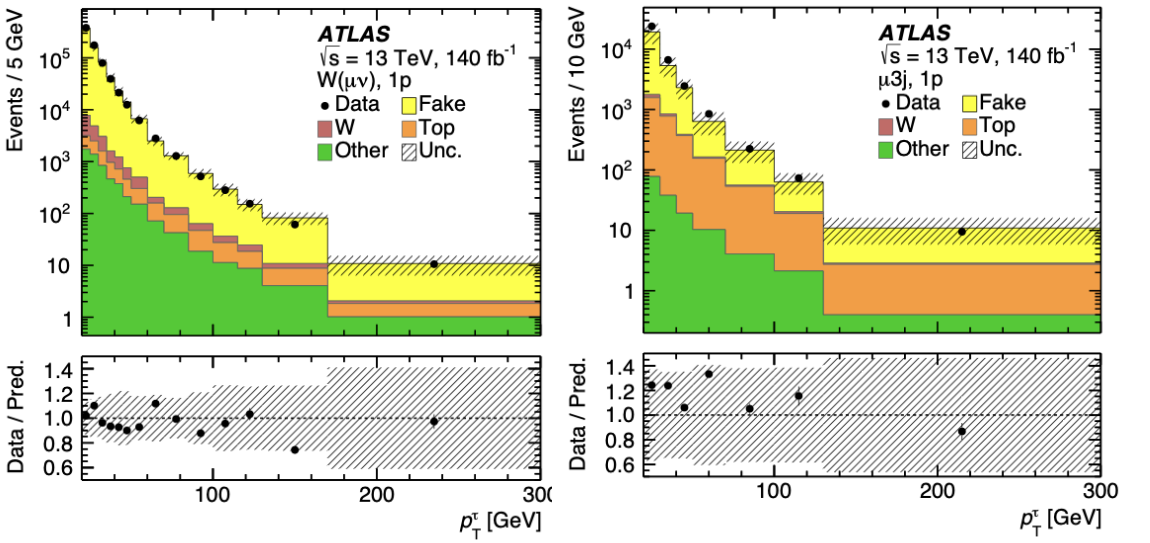


- $W(\mu\nu)$ *region*
  - exactly one muon
  - exactly one hadronic tau
  - same sign events
  - enriched in q-fake τ_{had} associated with W-production
- $\mu 3j$ *region*
  - selection same as for $W(\mu\nu)$
  - at least three jets
  - at least one of them must be b-tagged
  - enriched in $t\bar{t}$ semileptonic events, containing q-fake τ_{had} from the decay of W bosons

ATLAS EXPERIMENT

Validation in $W(\mu\nu)$ and $\mu 3j$ regions





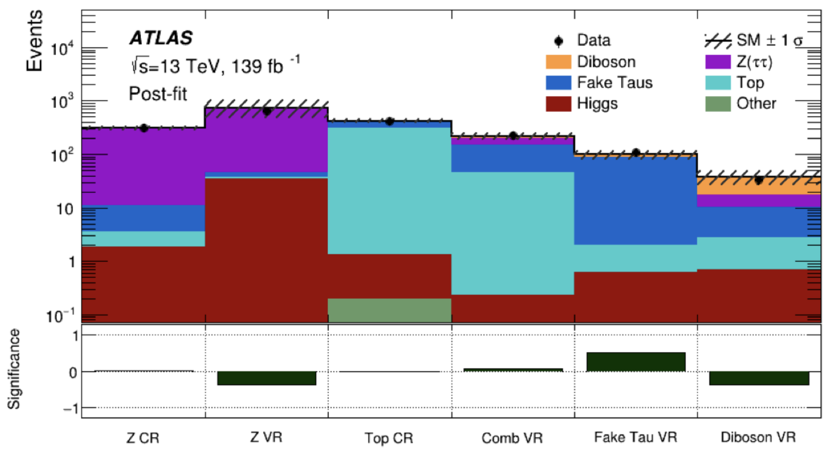
- Possible to determine the FFs directly – reference FFs
  - As the ratio of the ID to anti ID event yield
- UFF FFs agree with reference FFs within uncertainties in both validation regions

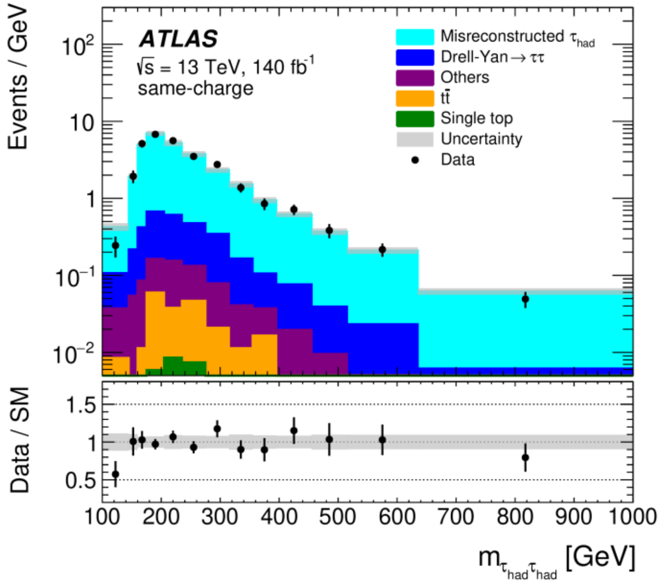
- Comparison of predicted fake background spectra with data
  - Prediction agrees with data within uncertainty bands in both validation regions

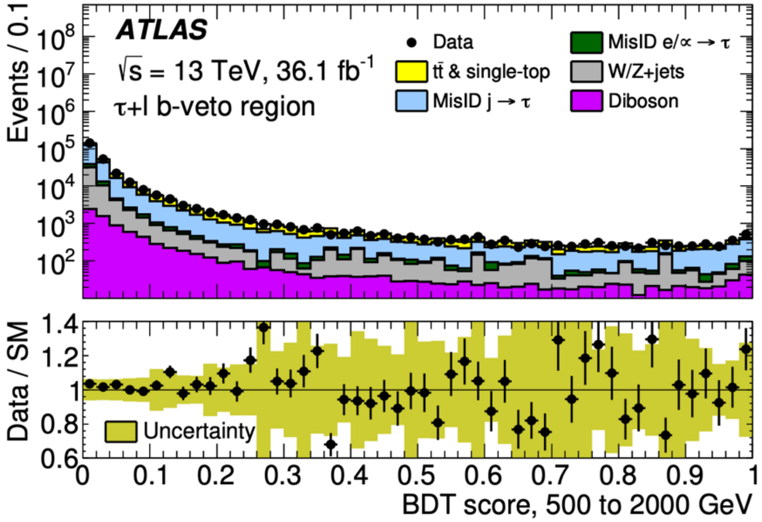
ATLAS EXPERIMENT

Validation and usage of the method in published ATLAS analyses

- Search for dark matter produced in association with a Higgs boson decaying to tau leptons
  - arXiv:2305.12938
- High-mass $pp \to \tau\bar{\tau}$ cross-section measurement
  - arXiv:2503.19836
- Search for charged Higgs bosons decaying via $H^{\pm} \to \tau^{+}\nu_{\tau}$ in the τ+jets and τ+lepton final states
  - arXiv:1807.07915







Conclusion

- Presented a new technique to determine FFs
- The UFF method can evaluate FFs for any signal region
- Systematic uncertainty is estimated by comparing FFs for *q/g/b/p* fakes in different regions
  - Using MC samples
- Detailedly described validation in $W(\mu\nu)$ and $\mu 3j$ regions
  - UFF FFs agree with reference FFs within uncertainty
  - Fake-background prediction agrees with the data, withing the uncertainties
- Preliminary version of the UFF method used in multiple analyses

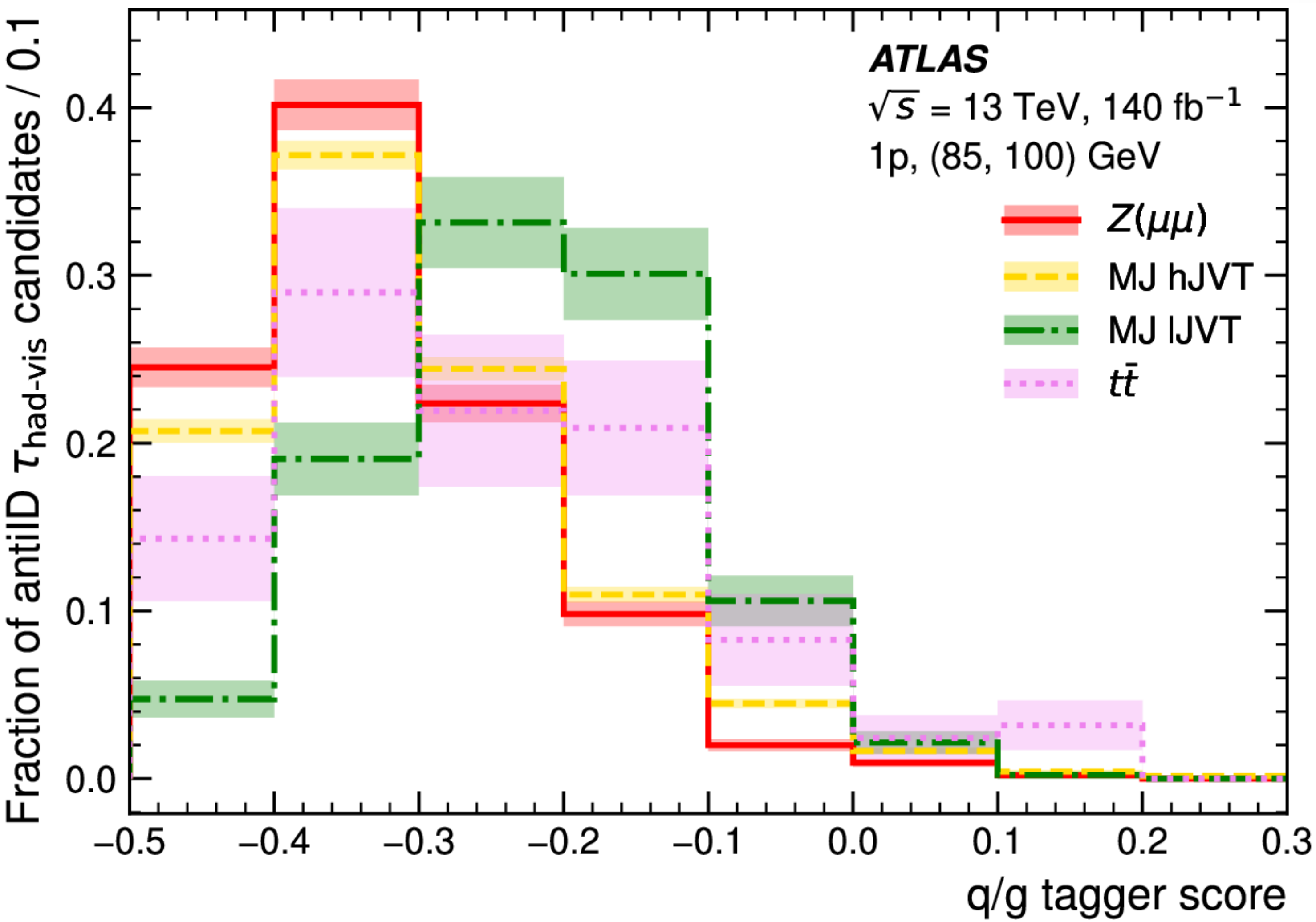
BACKUP

$Z(\mu\mu)$	$t\bar{t}$	MJ	
$N_\mu = 2$	$N_\mu = 1$	$N_\mu = 0$	
$N_e = 0$	$N_e = 1$	$N_e = 0$	
$N_{b\text{-jet}} = 0$	$N_{b\text{-jet}} \geq 1$	$N_{b\text{-jet}} = 0$	
$N_\gamma = 0$	$N_\gamma = 0$	$N_\gamma = 0$	
Single-μ trigger	Single-e trigger	Single-jet trigger	
	Tight LLH e ID		
Tight μ ID	*Tight* μ ID		
$p_T^{\mu_1} > 28$ GeV	$p_T^e > 27$ GeV	$p_T^{j_1} > 450$ GeV	
$p_T^{\mu_2} > 15$ GeV	$p_T^\mu > 10$ GeV		
		$\lvert\eta_{j_1}\rvert < 3$	
		$\lvert\eta_{j_2}\rvert < 3$	
$q_{\mu_1} q_{\mu_2} = -1$	$q_e q_\mu = -1$		
$m_{\mu\mu} \in (76, 106)$ GeV		$\Delta\phi(j_1, j_2) > 2.7$	
		$p_T^\tau < \left(p_T^{j_1} + p_T^{j_2}\right) / 8$	
		$N_j = 2$ OR $p_T^{j_3} < p_T^\tau$	
Tight $\tau_{\text{had-vis}}$ eBDT	*Tight* $\tau_{\text{had-vis}}$ eBDT	*Tight* $\tau_{\text{had-vis}}$ eBDT	
		MJ hJVT	MJ lJVT
$\text{JVT}_{\tau_{\text{had-vis}}} > 0.8$	$\text{JVT}_{\tau_{\text{had-vis}}} > 0.8$	$\text{JVT}_{\tau_{\text{had-vis}}} > 0.8$	$\text{JVT}_{\tau_{\text{had-vis}}} < 0.8$

Process	Generator		PDF set		UE tune	ME order	XS order
	ME	PS	ME	PS			
V + jets QCD	SHERPA 2.2.1		NNPDF3.0NNLO		Sherpa	NLO/LO	NNLO
V + jets EW	SHERPA 2.2.1		NNPDF3.0NNLO		Sherpa	LO	LO
$t\bar{t}$	POWHEG BOX v2	PYTHIA 8.230	NNPDF3.0NLO	NNPDF2.3LO	A14	NLO	NNLO
Single top	POWHEG BOX v2	PYTHIA 8.230	NNPDF3.0NLO	NNPDF2.3LO	A14	NLO	NLO
$t\bar{t}V$	MG5 v2.3.3	PYTHIA 8.210	NNPDF3.0NLO	NNPDF2.3LO	A14	NLO	NLO
VV	SHERPA 2.2.1,2.2.2		NNPDF3.0NNLO		Sherpa	NLO/LO	NNLO
Dijet	PYTHIA 8.230		NNPDF2.3LO	NNPDF2.3LO	A14	LO	LO
V + jets QCD	POWHEG BOX v1	PYTHIA 8.186	CT10NLO	CTEQ6L1	AZNLO	NLO	NLO
Dijet	POWHEG BOX v2	PYTHIA 8.245	CT18NNLO	NNPDF2.3LO	A14	NLO	NLO

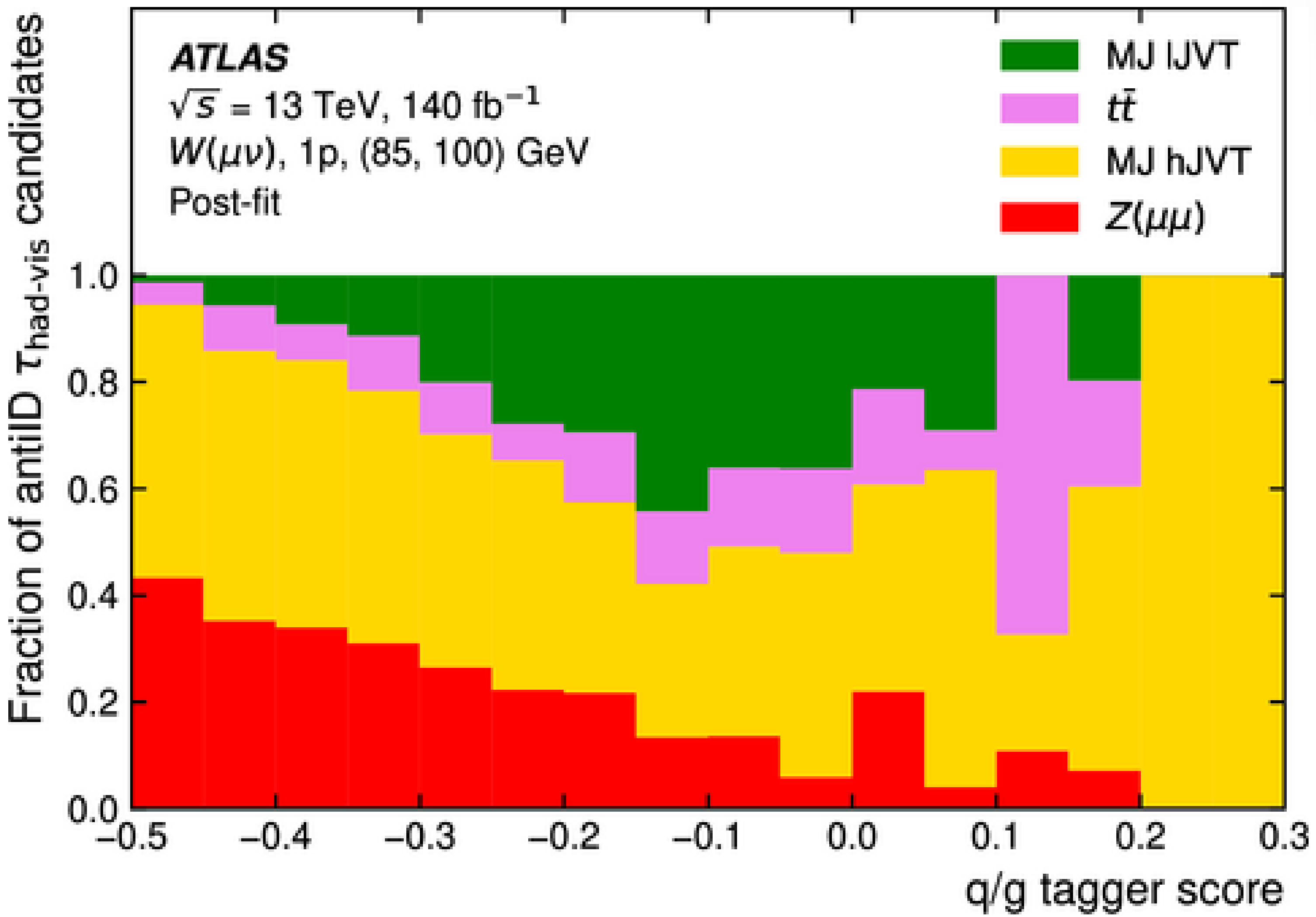
- List of MC generators, their settings, and the αs-order of the cross-section (XS) calculation

ATLAS EXPERIMENT



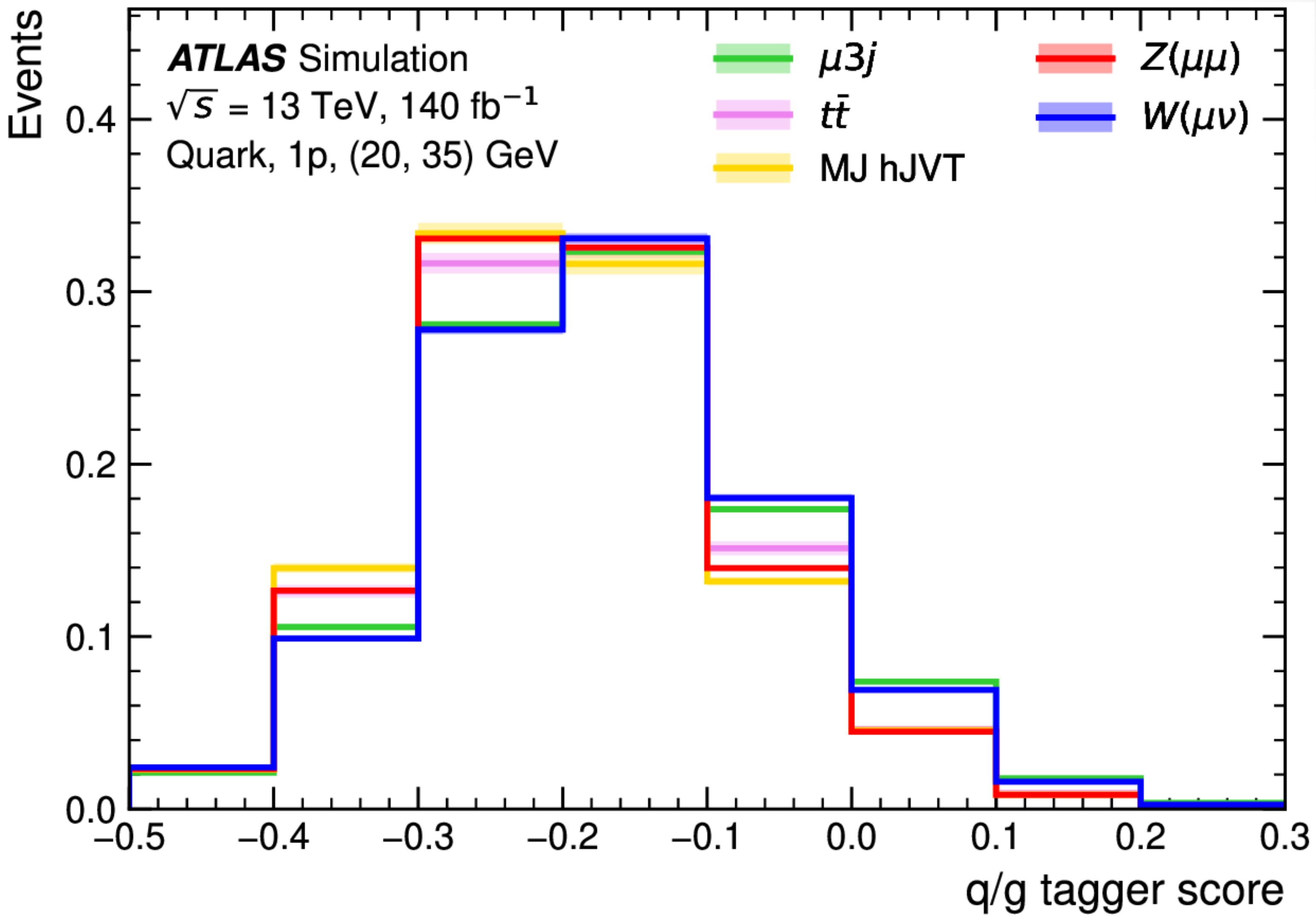
- Templates of the q/g tagger score distributions

arXiv:2502.04156v1



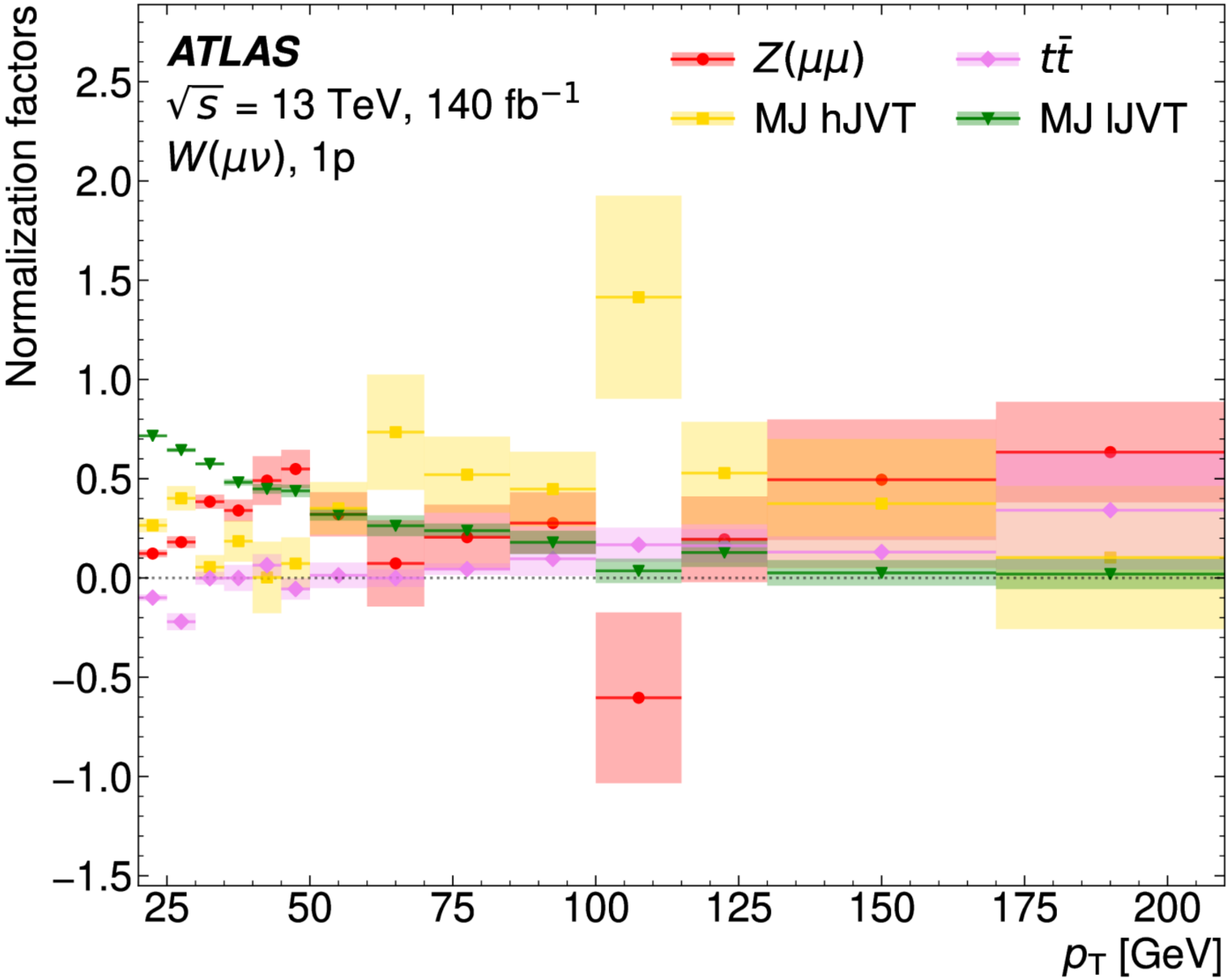
- Results of the template fit

arXiv:2502.04156v1



- Spectra of the q/g tagger score determined for q fakes in the Z(μμ), $t\bar{t}$, MJ hJVT, W(μν), and μ3j antiID subregions using MC

arXiv:2502.04156v1



- Dependence on p_{T} of the $Z(\mu\mu)$ (red), MJ hJVT (yellow), MJ lJVT (green) and $t\bar{t}$ (violet) post-fit normalization factors

arXiv:2502.04156v1