XXIX Cracow Epiphany Conference 2023

Search for $H^{\pm} \to \tau\nu$ and fake τ background estimation in the ATLAS experiment

Paper: JHEP 09(2018)139

January 17th, 2023

Monika Juzek (INP PAN)

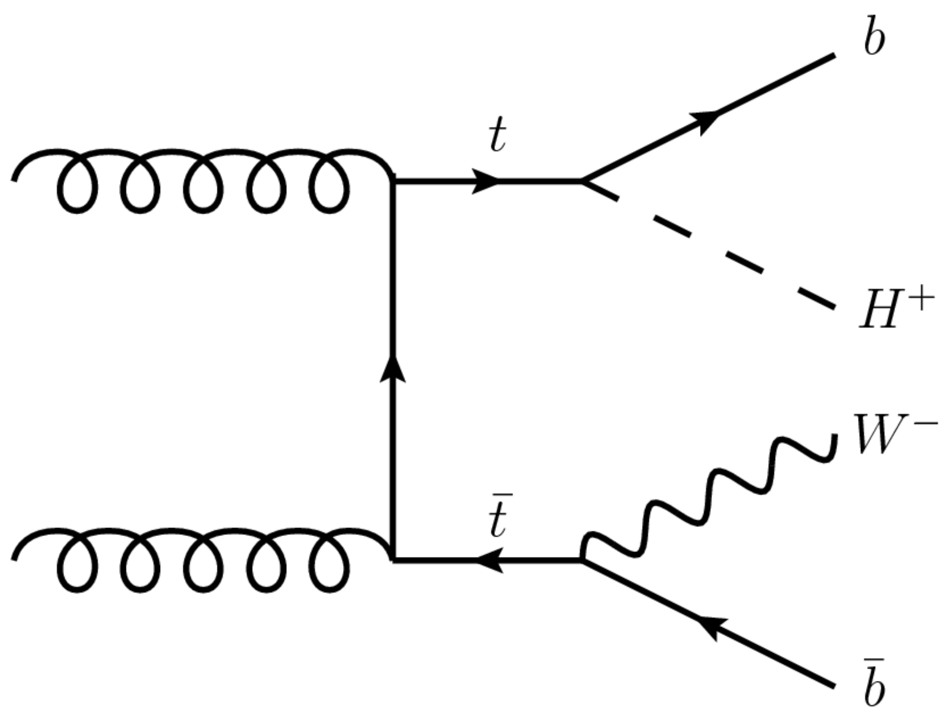
On behalf of the ATLAS Collaboration

THE HENRYK NIEWODNICZAŃSKI
INSTITUTE OF NUCLEAR PHYSICS
POLISH ACADEMY OF SCIENCES

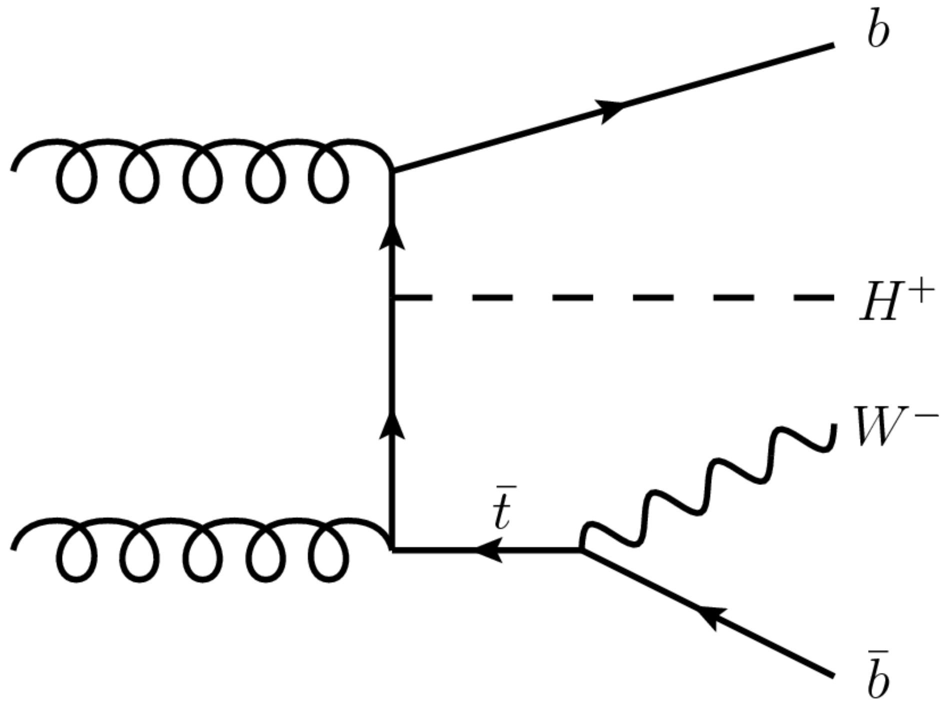
ATLAS EXPERIMENT

Introduction

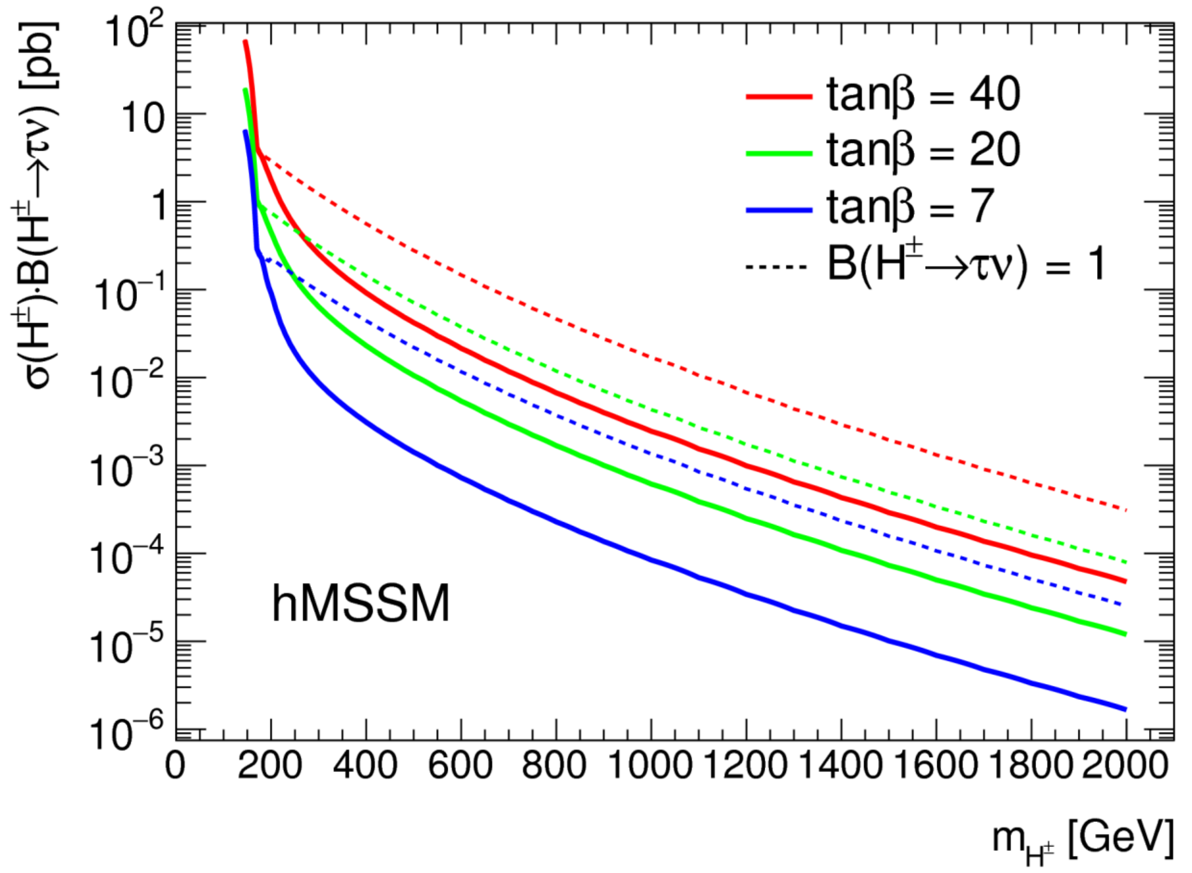
- $H^{\pm}$ bosons are predicted in different extensions of the SM that add a second doublet or triplets to the Higgs scalar sector (2HDM, NMSSM, Triplet, etc.)
- Search for charged $H^{\pm}$ decaying to τν using 36.1 fb^{-1} of $\sqrt{s}$ = 13 TeV pp data collected by the ATLAS experiment at the LHC during the 2015 and 2016 run periods
- Investigated signal mass range: 90 GeV ≤ $m_{H^{\pm}}$ ≤ 2000 GeV
- $H^{\pm}$ production process depends on the mass range but is typically associated with t and b quarks
- Analysis split into two sub-channels based on the decay mode of the associated t:
  - **τ+jets:** most sensitive to high $H^{\pm}$ mass
  - **τ+lepton:** most sensitive to low and intermediate $H^{\pm}$ mass



Low & intermediate $m_{H^{\pm}}$



High $m_{H^{\pm}}$



Event Selection & MVA Strategy

τ+jets channel

- 1 medium τ candidate with $p_{\mathrm{T}}^{\tau} > 40$ GeV
- no loose leptons (e or μ) with $p_T > 20$ GeV
- ≥ 3 jets with $p_T > 25$ GeV, of which at least one is b-tagged
- $E_{\mathrm{T}}^{\mathrm{miss}} > 150$ GeV
- $m_{\mathrm{T}} > 50$ GeV

$$m_{\mathrm{T}} = \sqrt{2p_{\mathrm{T}}^{\tau}E_{\mathrm{T}}^{\mathrm{miss}}(1 - \cos\Delta\phi_{\tau,\mathrm{miss}})},$$

τ+lepton channel

- Exactly 1 ℓ (e or μ), with $p_{\mathrm{T}}^{\ell} > 30$ GeV
- Exactly 1 medium τ with no additional loose or tighter τ leptons in the event
- ℓ and τ with opposite signs
- $E_{\mathrm{T}}^{\mathrm{miss}} > 50$ GeV
- ≥ 1 b-tagged jets

$\Delta\phi_{\tau,miss}$ - the azimuthal angle between the τ candidate and the direction of $E_{\mathrm{T}}^{\mathrm{miss}}$

- **MVA: The BDTs (Boosted Decision Trees) are trained to separate signal from all backgrounds separately for τ+jets and τ+lepton events in 5 $H^{\pm}$ mass ranges**
- BDT score used as a final discriminant in a likelihood fit

ATLAS EXPERIMENT

Background Estimation

- SM backgrounds with true leptons:
  - $t\bar{t}, Wt$
  - V+jets
  - diboson events

estimated using MC

- SM backgrounds with fake τ leptons:
  - fake lepton $\rightarrow \tau$
  - fake jets $\rightarrow \tau$

estimated with data-driven Fake Factor (FF) method

- $t\bar{t}$ events are main background at low and intermediate $H^{\pm}$ masses, while multi-jets events dominate for high $H^{\pm}$ masses

ATLAS EXPERIMENT

Fake Factor Method I

- The fake jets $\rightarrow \tau$ background in the region with the nominal object selection (τ ID) is estimated from events with an orthogonal selection, where the object identification requirement is inversed (anti-τ ID). Such sample is rich in fake τ leptons and depleted in signal.
- **The extrapolation from anti-τ ID to τ ID is done by the fake factor (FF) defined as:**

$$\mathrm{FF} = \frac{N_{\tau}^{\mathrm{CR}}(\mathrm{data}) - N_{\tau}^{\mathrm{CR}}(\mathrm{MC}, \mathrm{true}-\tau)}{N_{\mathrm{anti}-\tau}^{\mathrm{CR}}(\mathrm{data}) - N_{\mathrm{anti}-\tau}^{\mathrm{CR}}(\mathrm{MC}, \mathrm{true}-\tau)}$$

$\longrightarrow$ Events that pass τ identification

$\longrightarrow$ Events that fail the nominal τ identification

parametrization: p_T^{τ} and numer of tracks (1-track or 3-track τ)

- They are extracted from data in one or multiple control regions (CRs) enriched in the fake jets $\rightarrow \tau$ background.
- **In the $H^{\pm} \rightarrow \tau\nu$ analysis FFs are extracted in two CRs:**

   I. **Multi-jets CR** $\rightarrow$ gluon-initiated jets

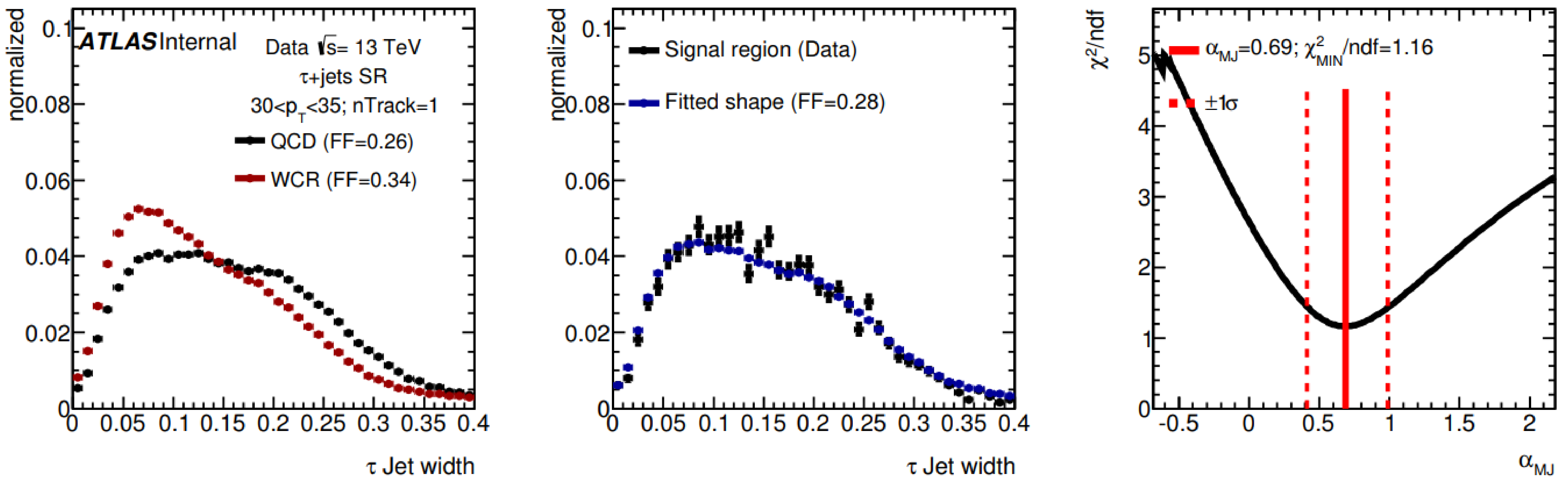
   II. **W+jets CR** $\rightarrow$ quark-initiated jets

Fake Factor Method II

- Different origin of fake jets → τ objects (gluons or quarks) results in different misidentification rate (hence different FFs). To obtain proper jets composition in a signal region, FFs from 2 control regions are combined using the formula:

$$\mathrm{FF}^{\mathrm{comb}}_{\mathrm{SR}}(i) = \alpha_{\mathrm{MJ}}(i) \times \mathrm{FF}_{\mathrm{MJ}}(i) + (1 - \alpha_{\mathrm{MJ}}(i)) \times \mathrm{FF}_{\mathrm{W+jets}}(i)$$

- $\boldsymbol{\alpha_{MJ}}$ **parametr is calculated using a template-fit method**: creates 2 templates for CRs (distributions for discriminant variables) and then fits it to the SR template



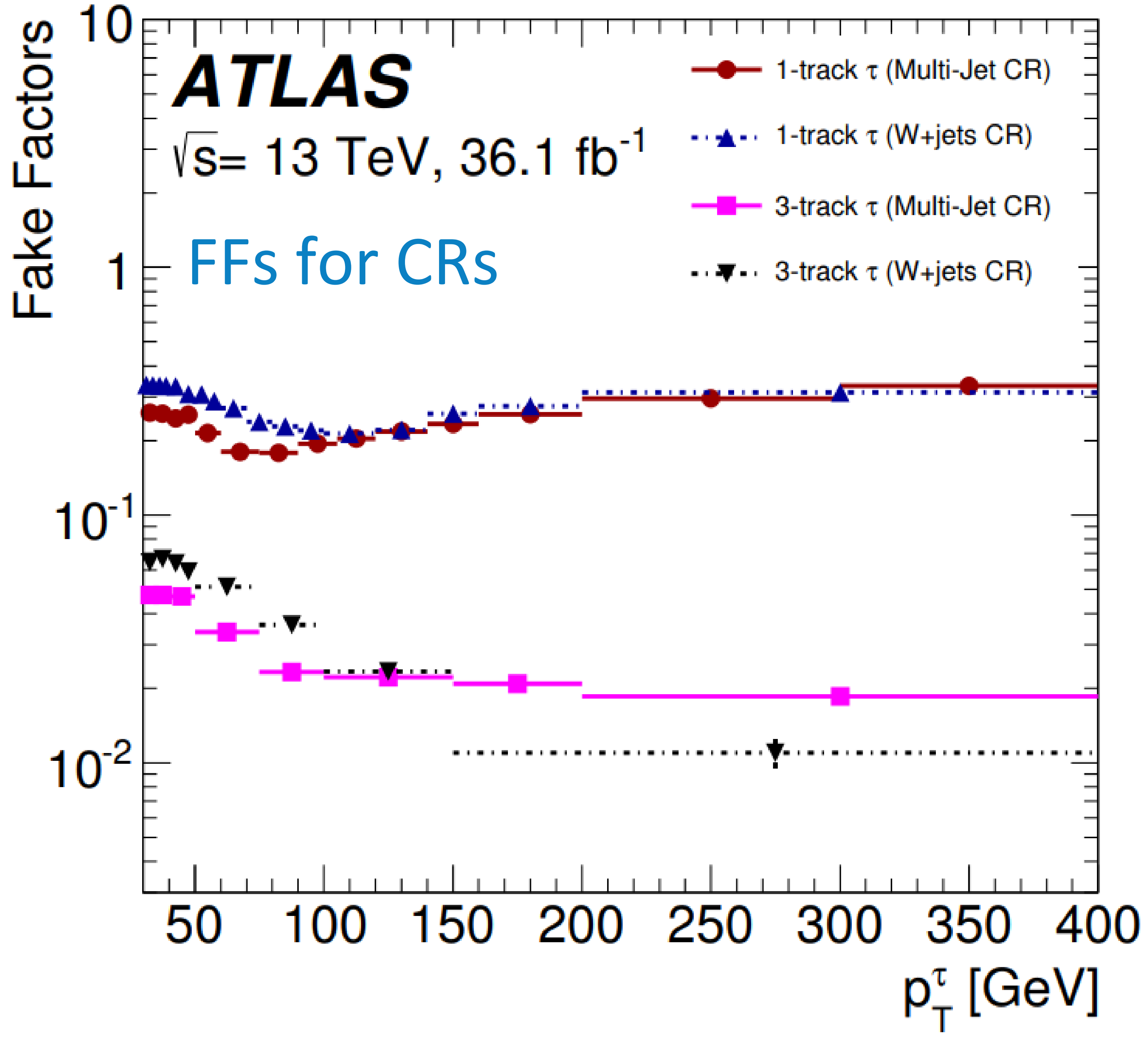
2 discriminant variables:

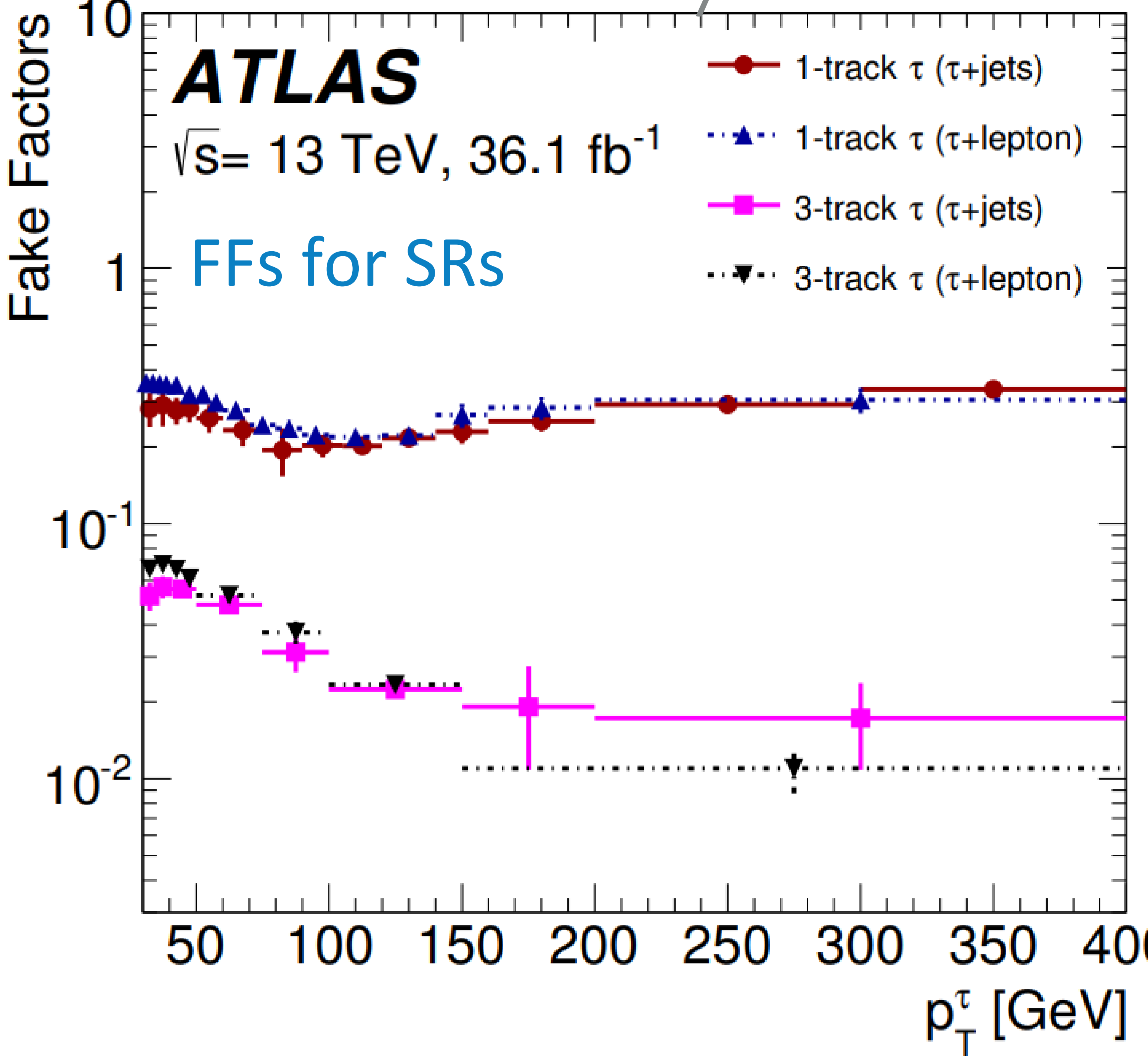
τ jet width and τ BDT score

- **Total number of background events from jets in signal region**:

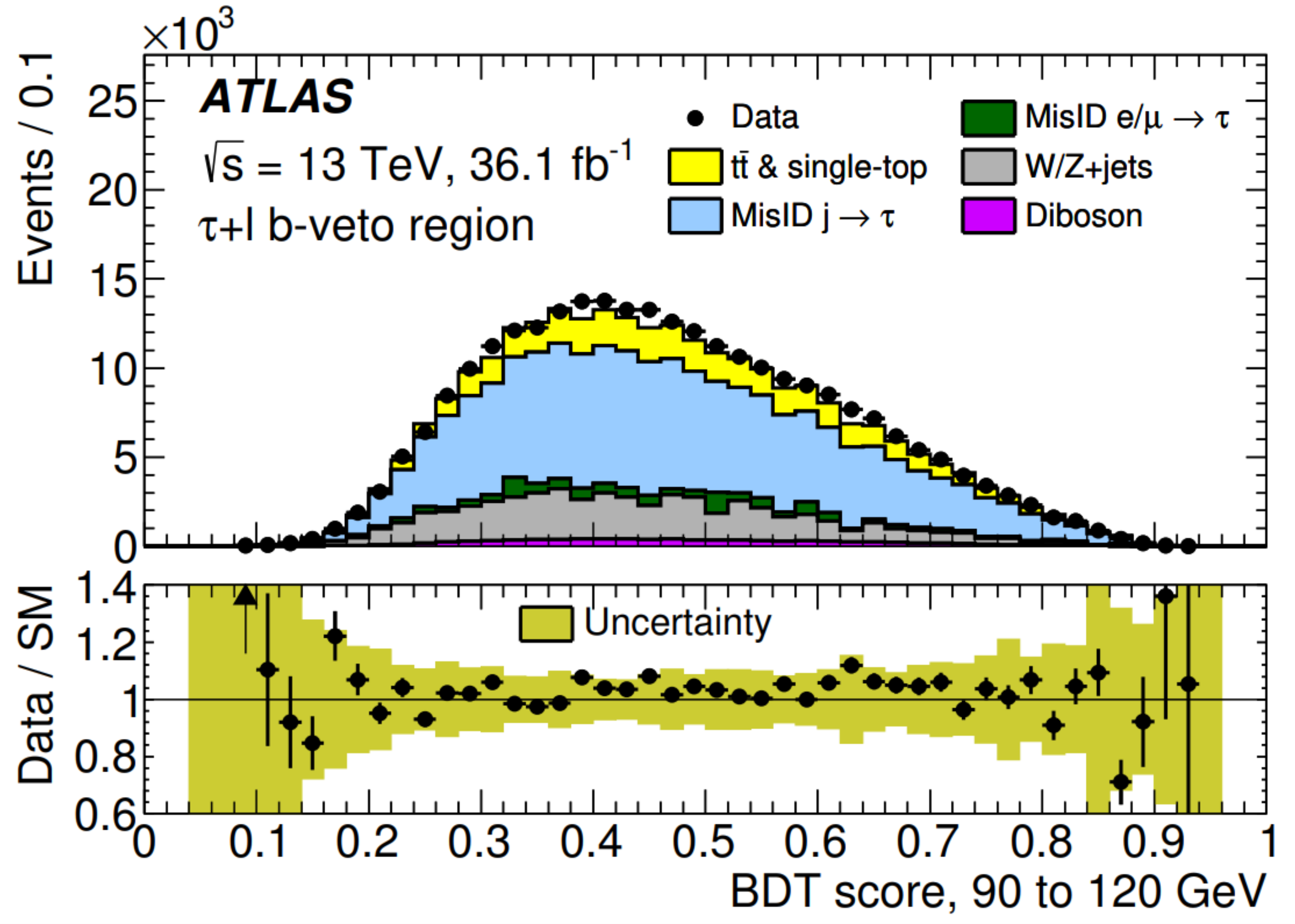
$$N^{\tau}_{\mathrm{fakes}} = \sum_i N^{\mathrm{SR}}_{\mathrm{anti}\text{-}\tau}(i) \times \mathrm{FF}(i),$$

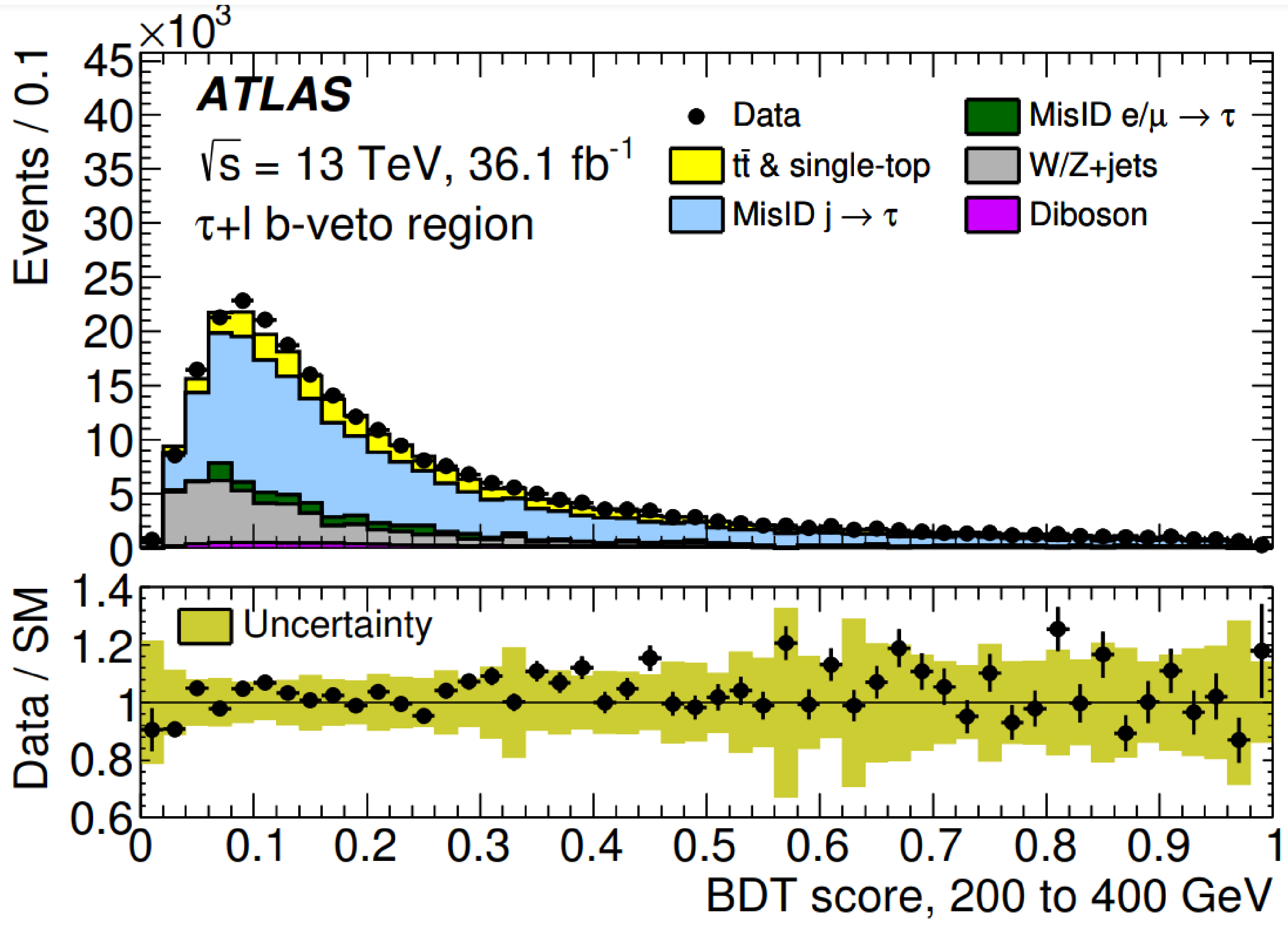
Obtained FFs:





Validation plots:





Systematics of Fake Factor method

I. the lower cut requirement on the τ BDT output score used in the definition of the anti- τ sample

II. the statistical uncertainties in the event yields entering the computation of FFs (each bin of their parameterization and for each control region)

III. the level of contamination of true τ candidates fulfilling the anti- τ selection (varied by 50%)

IV. the statistical uncertainty of the best-fit value of α_{MJ}

V. for the Υ distribution (used in BDT-based τ ID) the uncertainty of the inverse transform sampling method (Smirnov transformation for Υ)

VI. the modelling of heavy-flavour jets mimicking τ candidates

ATLAS EXPERIMENT

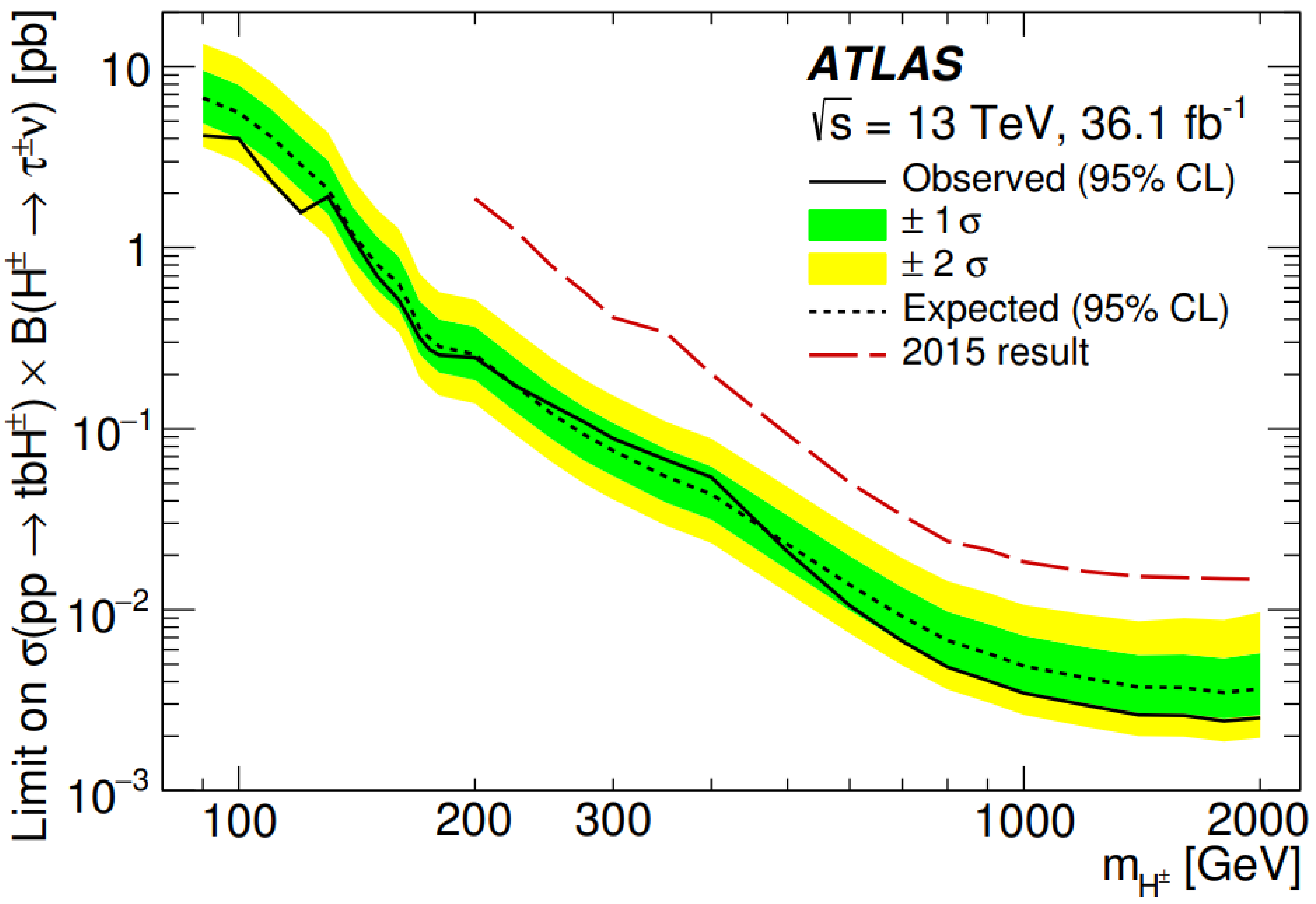
Systematics of Fake Factor method in $H^{\pm} \rightarrow \tau\nu$ analysis

	Source of uncertainty	$\tau_{\text{had-vis}}$+jets Effect on yield	$\tau_{\text{had-vis}}$+lepton Effect on yield
I.	Fake factors: jet composition	1.6%	0.6%
II.	Fake factors: statistical uncertainties	1.6%	1.7%
III.	Fake factors: prompt $\tau_{\text{had-vis}}$ in the anti-τ CR	+5.6% -8.3%	+4.8% -7.2%
IV.	Fake factors: α_{MJ} uncertainty	7%	6.2%
V.	Fake factors: Smirnov transform.	0%	0%
VI.	Fake factors: heavy flavor jet fraction.	5%	5%

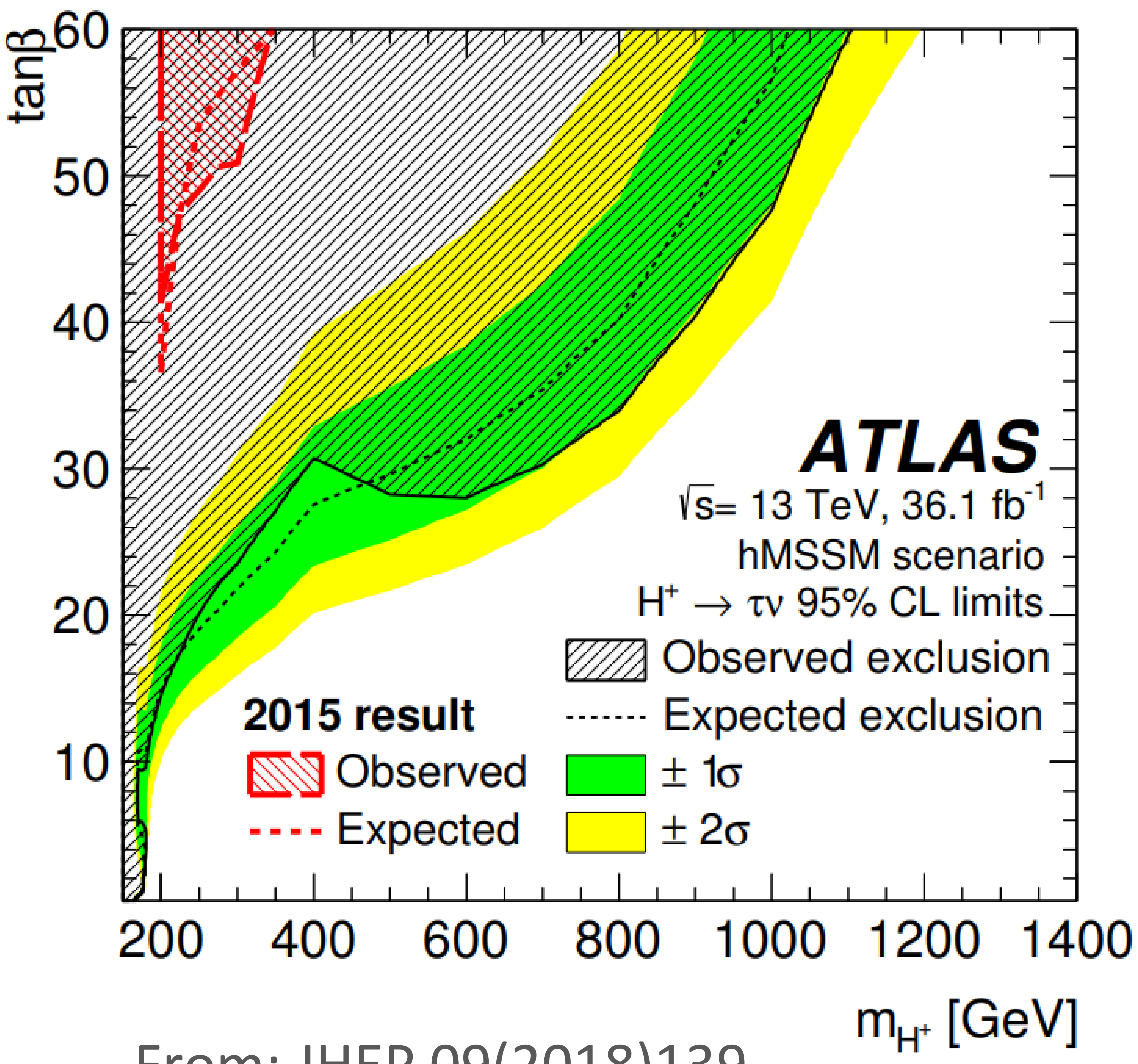
- Estimation of the fake jets $\rightarrow \tau$ background is the main source of systematic uncertainties in the low- and intermediate-mass $H^{\pm}$ search and the second major source (after the signal modelling) for large $H^{\pm}$ masses

ATLAS EXPERIMENT

Model-independent limit



Exclusion limits for hMSSM



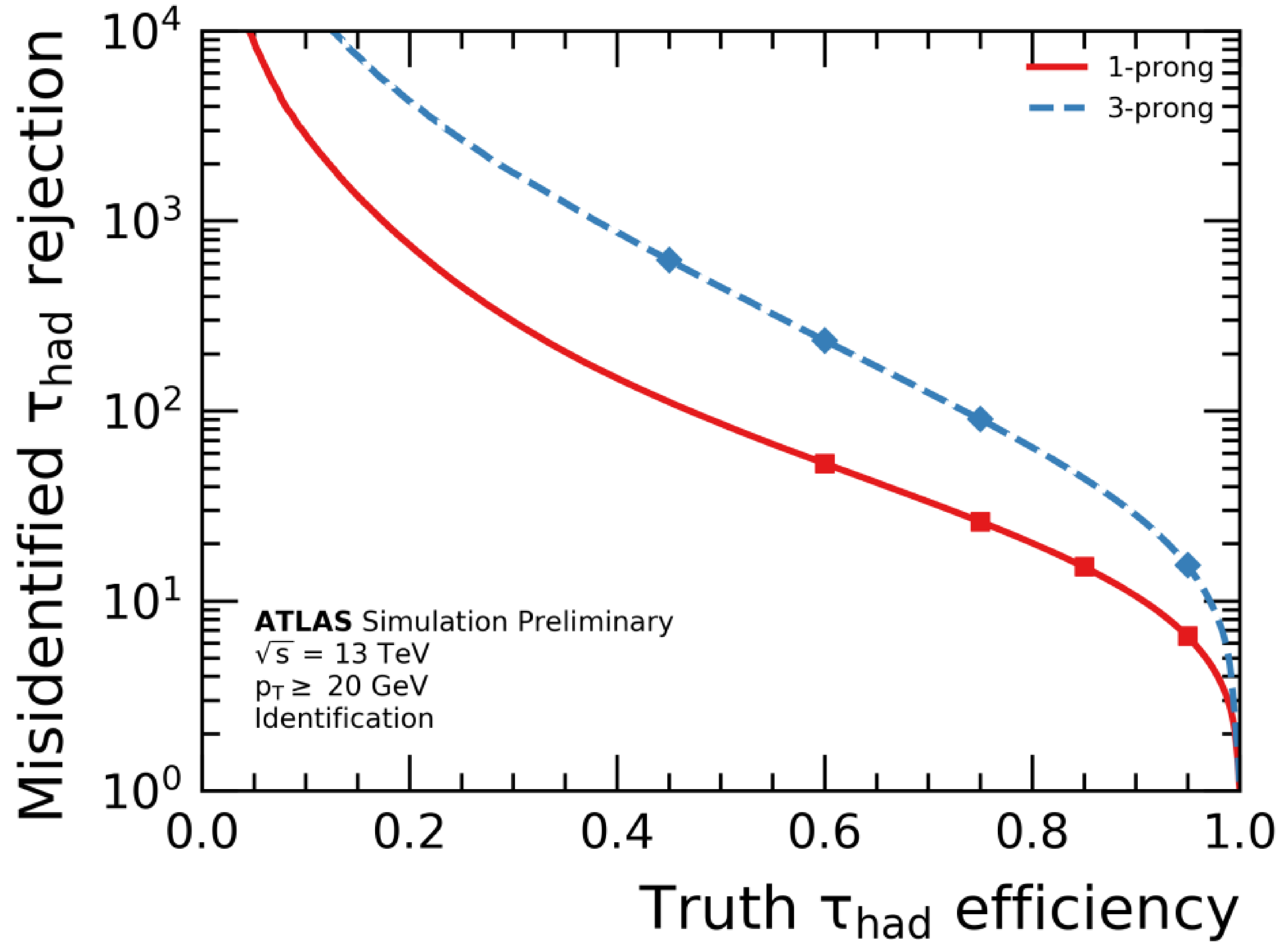
From: JHEP 09(2018)139

- All tan β values are excluded for $m_{H^\pm} \leq 160$ GeV.
- For $\tan\beta = 60$ mass of $H^\pm$ excluded up to 1100 GeV

Future full Run-2 analysis improvements

General:

- Full run-2 analysis covers a wider mass range than before:
  90 GeV ≤ $m_{H^\pm}$ ≤ 2000 GeV → 80 GeV ≤ $m_{H^\pm}$ ≤ 3000 GeV
- Developed a new E_T^{miss} trigger SF approach
- Switch from BDT-based to RNN-based τ identification (identification efficiency improved in the range 75–100%)
- Instead of BDTs the Parametrized Neural Networks (PNN) are trained to separate signal from all backgrounds



From: ATL-PHYS-PUB-2022-044

FF method:

- Optimisation for new RNN-based τ identification
- One discriminant variable for α_{MJ} fit: τ jet width
- Systematic errors updates: evaluation of uncertainty from τ Scale Factors and heavy-flavour FFs

Summary

- The $H^{\pm} \rightarrow \tau\nu$ analysis based on 2015-2016 data collected by ATLAS Collaboration found agreement between data and the background-only hypothesis.
- New exclusion limits have been set
- Data-driven Fake Factor method for fake taus background estimation was presented as a major source of systematic uncertainties
- Currently getting ready to publish results with full Run 2 (2015-2018) data with improved analysis strategy

Thank you for your attention!

ATLAS EXPERIMENT

BACKUP

MVA Strategy

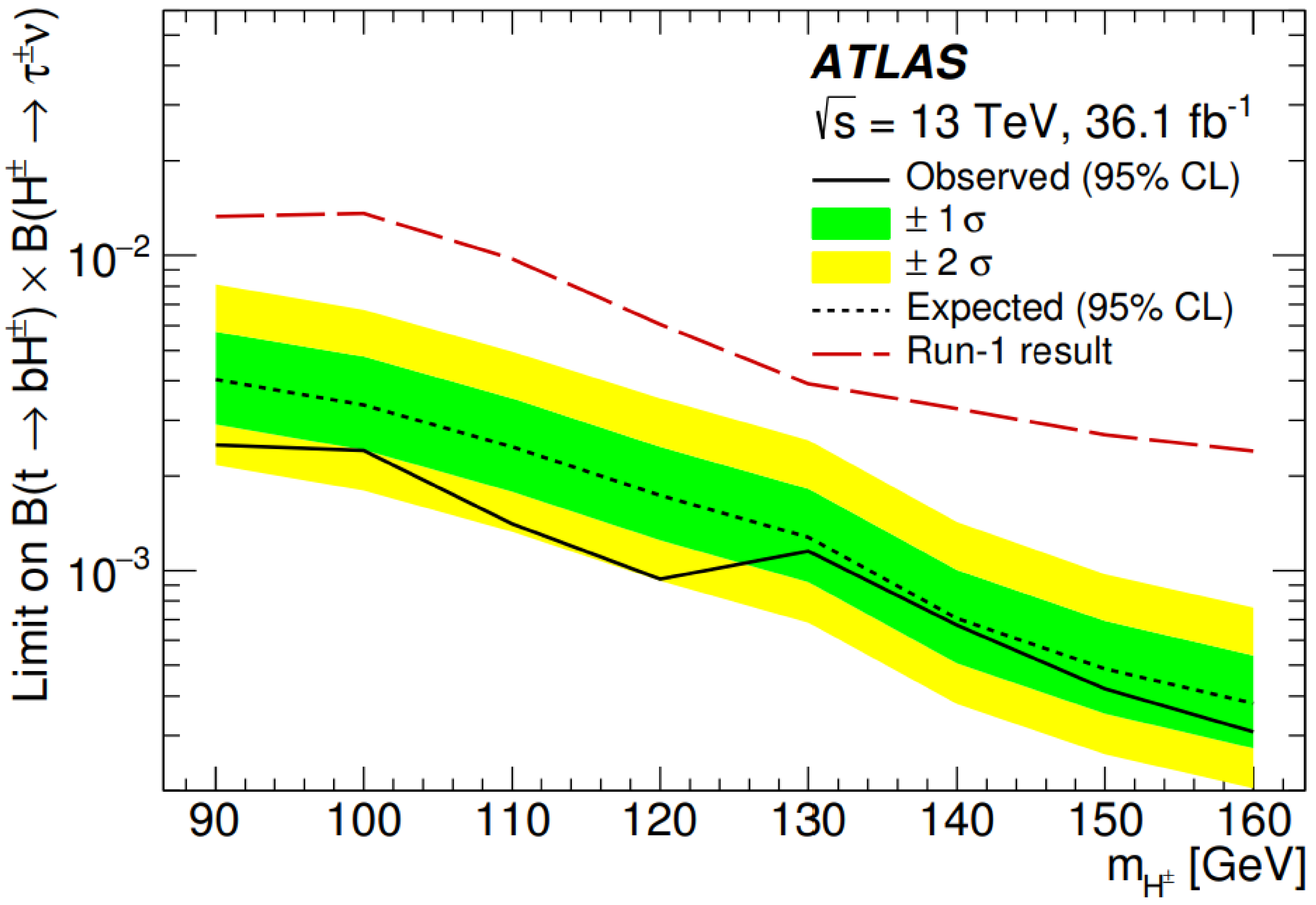
- The BDTs (Boosted Decision Trees) trained to separate signal from all backgrounds separately for τ+jets and τ+lepton events, numer of associated tracks to τ
- The training uses FastBDT library via the TMVA toolkit
- Signal samples are divided into H^+ mass ranges:
  - 80 - 120 GeV
  - 130 - 160 GeV
  - 170 - 190 GeV
  - 200 - 400 GeV
  - 500 - 3000 GeV
- Background modeling and and BDT training kept statistically independent via the k-fold method (k = 5)

MVA input variable	τ+jets	τ+lep
E_T^{miss}	✓	✓
p_T^{τ}	✓	✓
$p_T^{b\text{-jet}}$	✓	✓
p_T^{ℓ}		✓
$\Delta\phi_{\tau,\text{miss}}$	✓	✓
$\Delta\phi_{b\text{-jet},\text{miss}}$	✓	✓
$\Delta\phi_{\ell,\text{miss}}$		✓
$\Delta R_{\tau,\ell}$		✓
$\Delta R_{b\text{-jet},\ell}$		✓
$\Delta R_{b\text{-jet},\tau}$	✓	
$\Upsilon = 2\frac{p_T^{\tau\text{-track}}}{p_T^{\tau}} - 1$	✓*	✓*

(* only used for 1p τ, 90-400 GeV)

	Fold 1	Fold 2	Fold 3	Fold 4	Fold 5	Background
Partition 1	Evaluation	Train	Train	Train	Train	Fold 1
Partition 2	Train	Evaluation	Train	Train	Train	Fold 2
Partition 3	Train	Train	Evaluation	Train	Train	Fold 3
Partition 4	Train	Train	Train	Evaluation	Train	Fold 4
Partition 5	Train	Train	Train	Train	Evaluation	Fold 5

Model-independent limit for $90\ \text{GeV} < m_{H^\pm} < 160\ \text{GeV}$



Definitions of control regions used in FF method

Multi-jets CR	W+jets CR
at least one reconstructed τ_{had} candidate with $p_{\mathrm{T}}^{\tau} > 30$ GeV	at least one reconstructed τ_{had} candidate with $p_{\mathrm{T}}^{\tau} > 30$ GeV
number of jets 1	one lepton (electron or muon)
b-jet veto, electron and muon veto	b-jet veto
$E_{\mathrm{T}}^{\mathrm{miss}} < 80$ GeV	p_{T} of electron or muon > 30 GeV
$m_{\mathrm{T}}(\tau, \mathrm{E}_{\mathrm{T}}^{\mathrm{miss}}) > 50$ GeV	$60 < m_{\mathrm{T}}(l, E_{\mathrm{T}}^{\mathrm{miss}}) < 160 GeV$
transformed τ_{had} BDT score > 0.02	transformed τ_{had} BDT score > 0.02

Υ correction of fake τ candidates

- The FF method does not correctly predict the shape of Υ measured in the SR (distribution of Υ is different for τ and anti-τ candidates)
- Inverse transform sampling method* is employed to model the shape of Υ for misidentified τ candidates in the SRs
- In the CRs where FFs are measured, cumulative distribution functions $F(\Upsilon)$ are calculated from the shapes of Υ (separately for τ and anti-τ candidates)
- Next, in the SRs the shape of Υ predicted for τ candidates is derived from that measured for anti-τ candidates, as follows:

$$\Upsilon_{\tau_{\text{had-vis}}} = F^{-1}_{\tau_{\text{had-vis}}}(F_{\text{anti-}\tau_{\text{had-vis}}}(\Upsilon))$$

*L. Devroye, Non-Uniform Random Variate Generation, published by Springer, 1986

ATLAS EXPERIMENT

Systematic errors as a function of $m_{H^{\pm}}$

Source of systematic uncertainty	Impact on the expected limit (stat. only) in % m_{H^+} = 170 GeV	m_{H^+} = 1000 GeV
Experimental		
luminosity	2.9	0.2
trigger	1.3	<0.1
$\tau_{\text{had-vis}}$	14.6	0.3
jet	16.9	0.2
electron	10.1	0.1
muon	1.1	<0.1
$E_{\text{T}}^{\text{miss}}$	9.9	<0.1
Fake-factor method	20.3	2.7
Υ modelling	0.8	–
Signal and background models		
$t\bar{t}$ modelling	6.3	0.1
W/Z+jets modelling	1.1	<0.1
cross-sections ($W/Z/VV/t$)	9.6	0.4
H^+ signal modelling	2.5	6.4
All	52.1	13.8

ATLAS EXPERIMENT