SEARCH FOR $H^{\pm} \to \tau\nu$ AND FAKE τ BACKGROUND ESTIMATION IN THE ATLAS EXPERIMENT[*] [**]

Monika Juzek

on behalf of the ATLAS Collaboration

Institute of Nuclear Physics Polish Academy of Sciences, Kraków, Poland

Received 29 March 2023, accepted 19 April 2023,
published online 6 September 2023

A significant part of the particle physics programme at the Large Hadron Collider (LHC) at CERN focuses on searching for potential signs of processes beyond the Standard Model. An example of such phenomena would be the discovery of the charged Higgs boson decaying into τ lepton and neutrino that is predicted by theories such as Supersymmetry. Identification of hadronically decaying τ leptons plays a crucial role in τ-related studies and misidentification of such decays can lead to extra background events. This paper presents an overview of the $H^{\pm} \to \tau\nu$ analysis based on 2015+2016 data recorded with the ATLAS detector at the LHC and shows the usage of the data-driven Fake Factor method for fake τ background modelling. The systematics associated with the method is also discussed.

DOI:10.5506/APhysPolBSupp.16.7-A30

1. Introduction and motivation

The Higgs boson was discovered in 2012 by the ATLAS [1] and CMS [2] experiments at the LHC at CERN [3, 4]. It immediately raised a question of whether it is the Higgs particle predicted by the Standard Model, as some Supersymmetry theories (like two-Higgs-doublet models — 2HDM), representing Beyond Standard Model (BSM) physics, are predicting more Higgs bosons to exist such as a charged Higgs boson $H^{\pm}$ [5]. $H^{\pm}$ production process depends on the mass range but is typically associated with t and b

[*] Presented at the 29$^{\text{th}}$ Cracow Epiphany Conference on *Physics at the Electron–Ion Collider and Future Facilities*, Cracow, Poland, 16–19 January, 2023.

quarks. In this analysis, the focus is put on the $H^{\pm} \to \tau\nu$ decay channel, which is one of the main dominant decay modes of $H^{\pm}$ below the $t\bar{t}$ threshold and remains significant for higher masses at high $\tan(\beta)$ values. A crucial issue in searches using τ leptons is the proper estimation of backgrounds arising from jets misidentified as hadronically decaying τ leptons.

2. Search for a charged Higgs boson

The $H^{\pm} \to \tau\nu$ analysis is based on 36.1 fb^{-1} pp collision data of $\sqrt{s} =$ 13 TeV collected by the ATLAS experiment at the LHC in 2015 and 2016. The investigated signal mass range of $H^{\pm}$ is 90–2000 GeV. Only hadronically decaying τs are taken into account, but two different channels are created depending on the decaying mode of the W into hadrons (τ-jet) or leptons (τ-lep). The BDTs (Boosted Decision Trees) are trained to separate signal from all backgrounds separately for $\tau+$jets and $\tau+$lepton events in five $H^{\pm}$ mass ranges. BDTs are also used for the identification of hadronically decaying τ leptons. Full description of the $H^{\pm} \to \tau\nu$ analysis strategy and event selection is described in [6].

3. Background estimation

An important source of the background in all τ-related physics analyses is the object misidentification. The quark- or gluon-initiated jets can genuinely fake hadronically decaying τ leptons and fulfill the selection criteria of the signal region (SR). The fake τ background is not well modelled by the MC simulation, hence it has to be measured by the data-driven Fake Factor method.

The fake jets $\to \tau$ background in the region with the nominal object selection (τ-ID) is estimated using events from the inversed object identification requirements (anti-τ-ID). Such a sample is rich in fake τ leptons and depleted in a signal. The extrapolation from anti-τ-ID to τ-ID is done by the fake factor (FF), calculated in a dedicated control region (CR) as the ratio between the number of the τ candidates that fulfill the τ-ID criteria ($N^{\mathrm{CR}_{\tau-\mathrm{id}}(\mathrm{data})}$), to the number of corresponding candidates failing the identification criteria ($N^{\mathrm{CR}_{\mathrm{anti}-\tau-\mathrm{id}}(\mathrm{data})}$), according to the formula

$$\mathrm{FF} = \frac{N^{\mathrm{CR}_{\tau-\mathrm{id}}(\mathrm{data})} - N^{\mathrm{CR}_{\tau-\mathrm{id}}(\mathrm{MC,true}-\tau)}}{N^{\mathrm{CR}_{\mathrm{anti}-\tau-\mathrm{id}}(\mathrm{data})} - N^{\mathrm{CR}_{\mathrm{anti}-\tau-\mathrm{id}}(\mathrm{MC,true}-\tau)}}\,. \tag{1}$$

The contribution from true τ events in either category is subtracted using the MC simulation. The Fake Factors are usually parametrized as a function of p_{T}^{τ} and the number of charged particles from the τ decay (1-track or 3-track τ).

In the $H^{\pm} \to \tau\nu$ analysis, they are extracted from data in two control regions: Multi-jets (mainly gluon-initiated jets) and $W+$jets CRs (enriched in quark-initiated jets). The definition of both CRs is shown in Table 1. The transverse mass m_{T} of the τ (analogously for ℓ, representing electron or muon) and missing transverse momentum ($E_{\mathrm{T}}^{\mathrm{miss}}$) are defined by

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

The $E_{\mathrm{T}}^{\mathrm{miss}}$ vector is reconstructed as the negative vector sum of transverse momenta of the fully calibrated objects reconstructed from the *pp* collision [7].

Table 1. Selection criteria for FF Control Regions in $H^{\pm} \to \tau\nu$ analysis [6].

Multi-jet CR	W+jets CR
at least one τ candidate with $p_{\mathrm{T}}^{\tau} > 30$ GeV	at least one τ candidate with $p_{\mathrm{T}}^{\tau} > 30$ GeV
number of jets at least 2	one electron or muon
b-jets veto, electron and muon veto	b-jets veto
$E_{\mathrm{T}}^{\mathrm{miss}} < 80$ GeV	p_{T} of electron and muon > 30 GeV
$m_{\mathrm{T}}\left(\tau, E_{\mathrm{T}}^{\mathrm{miss}}\right) < 50$ GeV	$60 < m_{\mathrm{T}}\left(\ell, E_{\mathrm{T}}^{\mathrm{miss}}\right) < 160$ GeV
τ identification score > 0.02	τ identification score > 0.02

A different origin of the falsely identified jets as the τ leptons (gluons or quarks) results in a different misidentification rate, hence different FFs. To obtain proper jets composition in the signal region, the FFs from two control regions are combined using the formula

$$\mathrm{FF} = \alpha_{\mathrm{MJ}} \times \mathrm{FF}_{\mathrm{MJ}} + (1 - \alpha_{\mathrm{MJ}}) \times \mathrm{FF}_{W+\mathrm{jet}} . \tag{3}$$

The α_{MJ} parameter is calculated using the template-fit method. The first step is creating two templates for CRs (distributions of the discriminant variables) and then fitting them to the SR distribution by minimizing the χ^2 statistics. Depending on the number of prongs in the τ decay, two discriminant variables are used: the τ jet width and the τ identification BDT score.

Then, the total number of background events from jets in the signal region ($N^{\mathrm{SR}_{\tau-\mathrm{fakes}}}$) is computed as

$$N^{\mathrm{SR}_{\tau-\mathrm{fakes}}} = \left(N^{\mathrm{SR}_{\mathrm{anti}-\tau-\mathrm{id}}(\mathrm{data})} - N^{\mathrm{SR}_{\mathrm{anti}-\tau-\mathrm{id}}(\mathrm{MC,true}-\tau)}\right) \times \mathrm{FF} . \tag{4}$$

The calculated FFs for the two CRs and $\tau+$jets and $\tau+$lepton SRs are shown in Fig. 1.

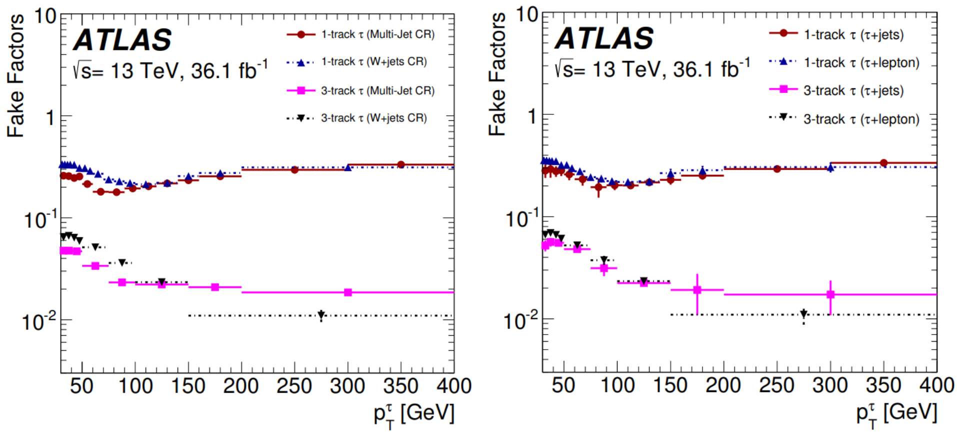

Fig. 1. The obtained fake factor parametrized as a function of p_{T}^{τ} and the number of the charged tracks associated with the reconstructed τ lepton in the (left) Multi-jets and W+jets CRs, as well as in (right) $\tau+$jets and $\tau+$lepton signal regions [6].

Validation of the background estimation is performed in the region enriched in misidentified τ leptons. As an example, the $\tau+$lepton b-veto region is used, which is defined by the same event selection as for the $\tau+$lepton SR, except that it has exactly zero b-tagged jets. The BDT score distributions for different $H^{\pm}$ mass ranges in this validation region are presented in Fig. 2, where a decent agreement between the data and the estimated background can be observed.

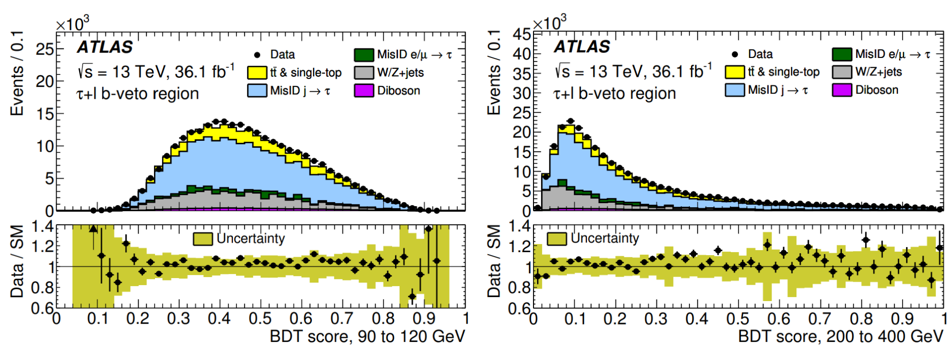

Fig. 2. Distribution of the BDT score for the predicted backgrounds and data in the $\tau+$lepton b-veto region. Two ranges of $H^{\pm}$ mass are presented: 90–120 GeV (left) and 200–400 GeV (right) [6].

Proper estimation of the fake jets $\to \tau$ background is a crucial part of the analysis, since this 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.

Systematic uncertainties for the following sources have been considered:

— the lower cut requirement on the τ identification BDT output score used in the definition of the anti-τ sample,

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

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

— the statistical uncertainty of the best-fit value of α_{MJ},

— for the Υ distribution (used in BDT-based τ-ID), the uncertainty of the inverse transform sampling method,

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

The biggest contribution comes from uncertainty about the contamination of the true τ candidates fulfilling the anti-τ selection and uncertainty of α_{MJ} fitting procedure.

4. Results

The search for $H^{\pm} \to \tau\nu$ based on 2015–2016 data collected by the ATLAS Collaboration at the LHC found an agreement between the data and the background-only hypothesis. Upper limits at the 95% confidence level for the mass range of $m_{H^{\pm}} = 90$–2000 GeV are set on the production cross section of the $H^{\pm}$ times the branching fraction $\mathcal{B}(H^{+} \to \tau\nu)$ in the range of 4.2–0.0025 pb. The data-driven Fake Factor method for the fake τ background estimation was presented with its uncertainties as a major source of systematic uncertainties. Currently, results from a new analysis using full Run 2 (2015–2018) dataset are expected to be published soon and, provided that no BSM signal is observed, the expected limits will improve.

REFERENCES

[1] ATLAS Collaboration (G. Aad *et al.*), *J. Instrum.* **3**, S08003 (2008).

[2] CMS Collaboration (S. Chatrchyan *et al.*), *J. Instrum.* **3**, S08004 (2008).

[3] ATLAS Collaboration (G. Aad *et al.*), *Phys. Lett. B* **716**, 1 (2012).

[4] CMS Collaboration (S. Chatrchyan *et al.*), *Phys. Lett. B* **716**, 30 (2012).

[5] A. Djouadi, *Phys. Rep.* **459**, 1 (2008).

[6] ATLAS Collaboration (M. Aaboud *et al.*), *J. High Energy Phys.* **2018**, 139 (2018).

[7] ATLAS Collaboration, *Eur. Phys. J. C* **78**, 903 (2018).