EUROPEAN ORGANISATION FOR NUCLEAR RESEARCH (CERN)

Submitted to: JHEP

CERN-EP-2022-279
10th February 2023

Searches for lepton-flavour-violating decays of the Higgs boson into $e\tau$ and $\mu\tau$ in $\sqrt{s} = 13$ TeV pp collisions with the ATLAS detector

The ATLAS Collaboration

This paper presents direct searches for lepton flavour violation in Higgs boson decays, $H \to e\tau$ and $H \to \mu\tau$, performed using data collected with the ATLAS detector at the LHC. The searches are based on a data sample of proton–proton collisions at a centre-of-mass energy $\sqrt{s} = 13$ TeV, corresponding to an integrated luminosity of 138 fb^{-1}. Leptonic ($\tau \to \ell\nu_\ell\nu_\tau$) and hadronic ($\tau \to$ hadrons ν_τ) decays of the τ-lepton are considered. Two background estimation techniques are employed: the MC-template method, based on data-corrected simulation samples, and the Symmetry method, based on exploiting the symmetry between electrons and muons in the Standard Model backgrounds. No significant excess of events is observed and the results are interpreted as upper limits on lepton-flavour-violating branching ratios of the Higgs boson. The observed (expected) upper limits set on the branching ratios at 95% confidence level, $\mathcal{B}(H \to e\tau) < 0.20\%$ (0.12%) and $\mathcal{B}(H \to \mu\tau) < 0.18\%$ (0.09%), are obtained with the MC-template method from a simultaneous measurement of potential $H \to e\tau$ and $H \to \mu\tau$ signals. The best-fit branching ratio difference, $\mathcal{B}(H \to \mu\tau) - \mathcal{B}(H \to e\tau)$, measured with the Symmetry method in the channel where the τ-lepton decays to leptons, is $(0.25 \pm 0.10)\%$, compatible with a value of zero within 2.5σ.

Contents

1 Introduction

One of the main goals of the Large Hadron Collider (LHC) physics programme at CERN is to discover physics beyond the Standard Model (SM). The discovery of a scalar Higgs boson at the LHC [1, 2] has provided important insight into the mechanism of electroweak symmetry breaking [3–8] and made it possible to search for physics phenomena beyond the SM (BSM physics phenomena) in the Higgs sector. A possible sign of new physics would be the observation of lepton flavour violation (LFV) in decays of the Higgs boson into a pair of leptons with different flavours.

The observation of neutrino oscillations indicates that LFV is realised in nature and that lepton flavour is not an exact symmetry, making it possible for BSM physics to participate in flavour-changing dynamics. LFV decays of the Higgs boson occur naturally in models with more than one Higgs doublet [9–12], composite Higgs models [13, 14], models with flavour symmetries [15] or warped extra dimensions [14, 16, 17] and other models [18, 19]. The flavour anomalies measured by BaBar, Belle and LHCb [20–24] could be linked to LFV decays of the Higgs boson or other massive particles [25–27].

The most stringent bounds on the LFV decays of the Higgs boson, $H \to e\tau$ and $H \to \mu\tau$, are derived from direct searches. These include a previous ATLAS search [28] which placed 95% confidence level (CL) upper limits on the branching ratios ($\mathcal{B}$) of $H \to e\tau$ and $H \to \mu\tau$ at 0.47% and 0.28%, respectively, using data collected at $\sqrt{s} = 13$ TeV, corresponding to an integrated luminosity of 36.1 fb^{-1}. Likewise, the CMS Collaboration set 95% CL upper limits restricting the branching ratios to $\mathcal{B}(H \to e\tau) < 0.22\%$ and $\mathcal{B}(H \to \mu\tau) < 0.15\%$ using data collected at $\sqrt{s} = 13$ TeV, with an integrated luminosity of 137 fb^{-1} [29]. The ATLAS Collaboration performed a direct search for $H \to e\mu$ decay and obtained a 95% CL upper limit on the branching ratio value of $\mathcal{B}(H \to e\mu) < 6.1 \times 10^{-5}$ [30], using data collected at $\sqrt{s} = 13$ TeV, with an integrated luminosity of 139 fb^{-1}. The most stringent indirect constraint on the $H \to e\mu$ decay is derived from the results of searches for $\mu \to e\gamma$ decays [31], and a bound of $\mathcal{B}(H \to e\mu) < O(10^{-8})$ is obtained [32, 33].

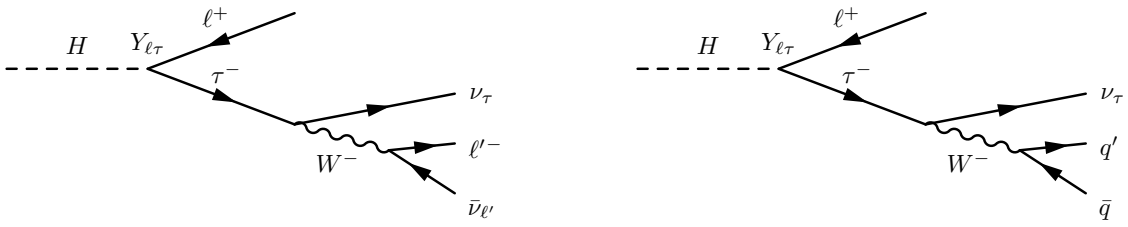
This document presents searches for two LFV decays of the Higgs boson, $H \to e\tau$ and $H \to \mu\tau$, with the ATLAS experiment at the LHC. The two decay modes $\ell\tau_{\ell'}$ and $\ell\tau_{\text{had}}$ illustrated in Figure 1 are considered for each search, where ℓ/ℓ' is used to denote electrons and muons,[1] also referred to as 'light leptons'. The former exploits the leptonic τ-decay mode $\tau \to \ell'\nu\bar{\nu}$, while the latter exploits the hadronic τ-decay mode $\tau \to$ hadrons + ν.

Due to the large background of same-flavour lepton pairs produced by Drell–Yan processes, the $\ell\tau_{\ell'}$ final state only considers pairs of different-flavour leptons.

Two independent methods are exploited to estimate the background in the $\ell\tau_{\ell'}$ final state: 'MC-template' and 'Symmetry'. The first method uses templates from Monte Carlo (MC) simulation, where the normalisation of the two main backgrounds is obtained from data, and a data-driven estimate of the 'misidentified background'. The second method relies on the assumption that the prompt-lepton[2] backgrounds in the SM are symmetric under the exchange of electrons and muons to derive a data-driven background estimate for the main backgrounds [34]. A separate data-driven estimate is employed for the misidentified background, and simulation is used for the remaining minor background contributions. For the $\ell\tau_{\text{had}}$ final state, only the MC-template method is used. The MC-template method targets the measurement of the actual values of $\mathcal{B}(H \to e\tau)$ and $\mathcal{B}(H \to \mu\tau)$ individually, while the Symmetry method is sensitive to the difference of

[1] Throughout this document the inclusion of charge-conjugate decay modes is implied.

[2] Leptons from the decay of the Higgs boson, heavy vector bosons, and τ-leptons are considered prompt leptons.



(a) (b)

Figure 1: LFV decay schemes of the Higgs boson for the (a) $\ell\tau_{\ell'}$ and (b) $\ell\tau_{\text{had}}$ final states. The off-diagonal Yukawa coupling term is indicated by the $Y_{\ell\tau}$ symbol.

the branching ratios. A multivariate analysis (MVA) technique is developed for each final state in both methods to achieve maximum separation between signal and background.

Three statistical analyses are performed: one independent search for each of the $H \to e\tau$ and $H \to \mu\tau$ processes, and one simultaneous determination of the $H \to e\tau$ and $H \to \mu\tau$ signals. In the independent search for the $H \to e\tau$ process, the $H \to \mu\tau$ signal is assumed to be zero, and vice versa in the case of the $H \to \mu\tau$ search. For each event category used in the statistical analysis, the MC-template method in the $\ell\tau_{\text{had}}$ final state is combined with either the Symmetry method or the MC-template method in the $\ell\tau_{\ell'}$ final state. The method having the higher expected sensitivity is chosen. In the simultaneous determination of the $H \to e\tau$ and $H \to \mu\tau$ signals, the assumption about the absence of one of the two signals is removed. Consequently, the MC-template method is used for both the $\ell\tau_{\ell'}$ and $\ell\tau_{\text{had}}$ final states.

2 ATLAS detector

The ATLAS detector [35] at the LHC covers nearly the entire solid angle around the collision point.[3] It consists of an inner tracking detector surrounded by a thin superconducting solenoid, electromagnetic and hadron calorimeters, and a muon spectrometer incorporating three large superconducting air-core toroidal magnets.

The inner-detector system is immersed in a 2 T axial magnetic field and provides charged-particle tracking in the range $|\eta| < 2.5$. The high-granularity silicon pixel detector covers the vertex region and typically provides four measurements per track, the first hit normally being in the insertable B-layer installed before Run 2 [36, 37]. It is followed by the silicon microstrip tracker, which usually provides eight measurements per track. These silicon detectors are complemented by the transition radiation tracker (TRT), which enables radially extended track reconstruction up to $|\eta| = 2.0$. The TRT also provides electron identification information based on a likelihood method.

[3] ATLAS uses a right-handed coordinate system with its origin at the nominal interaction point (IP) in the centre of the detector and the z-axis along the beam pipe. The x-axis points from the IP to the centre of the LHC ring, and the y-axis points upwards. Cylindrical coordinates (r, ϕ) are used in the transverse plane, ϕ being the azimuthal angle around the z-axis. The pseudorapidity is defined in terms of the polar angle θ as $\eta = -\ln\tan(\theta/2)$. Angular distance is measured in units of $\Delta R \equiv \sqrt{(\Delta\eta)^2 + (\Delta\phi)^2}$.

The calorimeter system covers the pseudorapidity range $|\eta| < 4.9$. Within the region $|\eta| < 3.2$, electromagnetic calorimetry is provided by barrel and endcap high-granularity lead/liquid-argon (LAr) calorimeters, with an additional thin LAr presampler covering $|\eta| < 1.8$ to correct for energy loss in material upstream of the calorimeters. Hadron calorimetry is provided by the steel/scintillator-tile calorimeter, segmented into three barrel structures within $|\eta| < 1.7$, and two copper/LAr hadron endcap calorimeters. The solid angle coverage is completed with forward copper/LAr and tungsten/LAr calorimeter modules optimised for electromagnetic and hadronic energy measurements respectively.

The muon spectrometer comprises separate trigger and high-precision tracking chambers measuring the deflection of muons in a magnetic field generated by the superconducting air-core toroidal magnets. The field integral of the toroids ranges between 2.0 and 6.0 T m across most of the detector. The region $|\eta| < 2.7$ is covered with three layers of precision chambers composed of monitored drift tubes, complemented by cathode-strip chambers in the forward region, where the background is highest. The muon trigger system covers the range $|\eta| < 2.4$ with resistive-plate chambers in the barrel, and thin-gap chambers in the endcap regions.

Interesting events are selected by the first-level trigger system implemented in custom hardware, followed by selections made by algorithms implemented in software in the high-level trigger [38]. The first-level trigger accepts events from the 40 MHz bunch crossings at a rate below 100 kHz, which is reduced to about 1 kHz by the high-level trigger and these events are recorded to disk.

An extensive software suite [39] is used in data simulation, in the reconstruction and analysis of real and simulated data, in detector operations, and in the trigger and data acquisition systems of the experiment.

3 Collision data and simulation samples

The dataset used for the searches consists of the LHC proton–proton collision data recorded by the ATLAS experiment at $\sqrt{s} = 13$ TeV during the period from 2015 to 2018. Events are selected for analysis only if they are of good quality and if all the relevant detector components are known to have been in good operating condition [40]. The total integrated luminosity of the analysed data is 138 fb^{-1}. The events considered were accepted by single-lepton or dilepton triggers [41–44]. The p_{T} thresholds of the single-lepton triggers were $p_{\mathrm{T}}^{e} > 27\,(25)$ GeV and $p_{\mathrm{T}}^{\mu} > 27.3\,(21)$ GeV for the 2016–2018 (2015) data-taking period. The p_{T} thresholds of the dilepton triggers were $p_{\mathrm{T}}^{e} > 18$ GeV and $p_{\mathrm{T}}^{\mu} > 14.7$ GeV.

Simulated events are used to model $H \to \ell\tau$ signal processes, as well as most of the backgrounds from SM processes. A summary of all the generators used for the simulation of the signal and background processes is shown in Table 1, and more details can be found in Ref. [45]. The measured Higgs boson mass of 125.09 GeV [46] is assumed in the calculation of the expected cross-sections and branching fractions. The Higgs-boson production cross-sections are fixed to the SM predictions [47] throughout this measurement.

The main Higgs boson production modes at the LHC are, in descending order of predicted cross-section, gluon–gluon fusion (ggF), followed by vector-boson fusion (VBF), and associated VH and $t\bar{t}H$ production. For the LFV signal, $H \to e\tau$ and $H \to \mu\tau$, the ggF, VBF and VH production mechanisms are considered. The $t\bar{t}H$ production process is not considered for the LFV signal due to its negligible contribution. The background contribution originating from the SM $H \to \tau\tau$ and $H \to WW$ decays is small and the SM predictions are assumed for the branching ratios. These two processes were modelled using the same simulation strategy as for the LFV signal, but the $t\bar{t}H$ production mode was also included. Other

background processes involve electroweak production of W/Z bosons via VBF, Drell–Yan production of W/Z in association with jets, and diboson, single top-quark and top-quark pair ($t\bar{t}$) production.

All samples of simulated events were processed through the ATLAS detector simulation [48] based on GEANT4 [49]. The effects of multiple interactions in the same and neighbouring bunch crossings (pile-up) were modelled by overlaying minimum-bias events, simulated using the soft QCD processes of PYTHIA 8.186 [50] with the A3 [51] set of tuned parameters and the NNPDF2.3LO [52] parton distribution functions (PDF).

Table 1: Overview of the MC generators used for the main signal and background samples. The last column, labelled 'Order', specifies the order of the cross-section calculation used for the normalisation of the simulated samples. More details can be found in Ref. [45]. The '*' symbol denotes additional V + jets (QCD/EW) samples used by the Symmetry method.

Process	Generator		PDF set		Tune	Order
	ME	PS	ME	PS		
Higgs boson						
ggF	POWHEG BOX v2 [53–57]	PYTHIA 8 [58]	PDF4LHC15NNLO [59]	CTEQ6L1 [60]	AZNLO [61]	N^{3}LO QCD + NLO EW
VBF	POWHEG BOX v2	PYTHIA 8	PDF4LHC15NLO	CTEQ6L1	AZNLO	NNLO QCD + NLO EW
VH	POWHEG BOX v2	PYTHIA 8	PDF4LHC15NLO	CTEQ6L1	AZNLO	NNLO QCD + NLO EW
$t\bar{t}H$	POWHEG BOX v2	PYTHIA 8	NNPDF3.0NNLO [62]	NNPDF2.3LO [52]	A14 [63]	NLO QCD + NLO EW
Background						
V + jets (QCD/EW)	SHERPA 2.2.1 [64]		NNPDF3.0NNLO		SHERPA [65]	NNLO QCD + LO EW
V + jets (QCD/EW)*	POWHEG BOX v2	PYTHIA 8	CT10NLO [66]	CTEQ6L1	AZNLO	NNLO
Diboson	SHERPA 2.2.1		NNPDF3.0NNLO		SHERPA	NLO
$t\bar{t}$	POWHEG BOX v2	PYTHIA 8	NNPDF3.0NNLO	NNPDF2.3LO	A14	NNLO + NNLL
Single top	POWHEG BOX v2	PYTHIA 8	NNPDF3.0NNLO	NNPDF2.3LO	A14	NLO

4 Object reconstruction and event selection

The topology of the $H \to \mu\tau$ and $H \to e\tau$ events requires the reconstruction of electrons, muons, visible products of hadronically decaying τ leptons ($\tau_{\text{had-vis}}$), jets and missing transverse momentum.

4.1 Object reconstruction

The tracks measured in the inner detector are used to reconstruct the interaction vertices. The vertex with the highest sum of squared transverse momenta of associated tracks is taken as the primary vertex [67].

Electrons and photons are reconstructed from energy deposits in the electromagnetic calorimeters [68]. Electron candidates are matched to inner-detector tracks. They are required to satisfy the 'Loose' likelihood-based identification criterion defined in Ref. [68], which has an efficiency of about 93%, and to have $p_{\text{T}} > 15\,\text{GeV}$ and $|\eta| < 1.37$ or $1.52 < |\eta| < 2.47$, thus excluding the transition region between the barrel and endcap calorimeters. Selected electrons are also required to satisfy the 'Gradient' isolation, which has an efficiency of 90% at $p_{\text{T}} = 25\,\text{GeV}$ and 99% at $p_{\text{T}} = 60\,\text{GeV}$, and 'Medium' identification criteria, which have an average efficiency of 88% for typical electroweak processes [68].

Muons are reconstructed from track candidates in the muon spectrometer matched with tracks in the inner detector. They are required to satisfy the 'Loose' identification criterion [69], and $p_{\text{T}} > 10\,\text{GeV}$ and $|\eta| < 2.5$. These criteria have an efficiency of about 98% with good uniformity in p_{T}. Selected muons are

also required to satisfy 'Medium' identification and 'Tight' isolation criteria [69]. The latter are based on calorimetric and track information for the MC-template method and on track information only for the Symmetry method, where a slightly larger track-p_{T} contribution around the muon is allowed in order to increase the number of events available for that method's data-driven background estimation.

Jets are reconstructed with the anti-k_t algorithm [70, 71] using a radius parameter $R = 0.4$. The jet-clustering input objects are based on particle flow [72] in the inner detector and the calorimeter. Cleaning criteria are used to identify jets arising from non-collision backgrounds or noise in the calorimeters [73], and events containing such jets are removed. Only jets with $p_{\mathrm{T}} > 20\,\mathrm{GeV}$ and $|\eta| < 4.5$ are considered. To identify and reject jets that are not associated with the primary vertex of the hard interaction (pile-up jets), a 'jet-vertex tagger' (JVT) [74] algorithm is applied to jets with $p_{\mathrm{T}} < 60\,\mathrm{GeV}$ and $|\eta| < 2.5$. To suppress pile-up jets in the forward region, a 'forward JVT' [75] algorithm which exploits jet shapes and topological jet correlations in pile-up interactions is applied to all jets with $p_{\mathrm{T}} < 60\,\mathrm{GeV}$ and $|\eta| > 2.5$. Jets with $|\eta| < 2.5$ containing b-hadrons are identified using the DL1r b-tagging algorithm [76–78]. The fixed 85% efficiency (measured in simulated $t\bar{t}$ events) working point is used. The rejection factors for b-tagging a jet initiated by a c-quark or light parton are 2.6 and 29 respectively at the 85% efficiency working point.

Leptonic τ-decays are reconstructed as electrons or muons. The τ_{had} decays are composed of a neutrino and a set of visible decay products, most frequently one or three charged pions and up to two neutral pions. The reconstruction of the $\tau_{\text{had-vis}}$ is seeded by jets reconstructed by the anti-k_t algorithm [70], using calibrated topological clusters [79] as inputs, with a radius parameter of $R = 0.4$ [80, 81]. The jets form $\tau_{\text{had-vis}}$ candidates and are additionally required to have $p_{\mathrm{T}} > 10\,\mathrm{GeV}$ and $|\eta| < 2.5$. Reconstructed tracks are matched to $\tau_{\text{had-vis}}$ candidates. To separate the $\tau_{\text{had-vis}}$ candidates originating from hadronic τ decays from quark/gluon-initiated jets, a recurrent neural network (RNN) $\tau_{\text{had-vis}}$ identification algorithm [82] is used. The $\tau_{\text{had-vis}}$ objects are required to satisfy the 'Very Loose' $\tau_{\text{had-vis}}$ identification criterion, which has an efficiency of 95% for simulated τ_{had} decays. A separate multivariate discriminant (eBDT) [83] is employed to reject backgrounds arising from electrons that are misreconstructed as single-track $\tau_{\text{had-vis}}$.

The reconstructed electrons, muons, $\tau_{\text{had-vis}}$ and jets used in this analysis are not built from a set of mutually exclusive tracks or calorimeter clusters; it is therefore possible that two different objects share most of their constituents. To resolve this ambiguity, an overlap removal procedure is applied. The p_{T} threshold of muons considered in the overlap removal with $\tau_{\text{had-vis}}$ is lowered to 2 GeV. More details can be found in Ref. [45].

The missing transverse momentum vector, $\vec{E}_{\mathrm{T}}^{\mathrm{miss}}$, is reconstructed as the negative vector sum of the transverse momenta of light leptons, photons, $\tau_{\text{had-vis}}$, jets, and the 'soft-term'. The soft-term is calculated as the vectorial sum of the p_{T} of tracks matched to the primary vertex but not to a reconstructed light lepton, $\tau_{\text{had-vis}}$ or jet [84]. The magnitude of $\vec{E}_{\mathrm{T}}^{\mathrm{miss}}$ is referred to as the missing transverse momentum, $E_{\mathrm{T}}^{\mathrm{miss}}$.

4.2 Event selection and categorisation

The two analysis channels are defined according to the τ decay mode. Events in the $\ell\tau_{\ell'}$ channel contain exactly two light leptons of opposite-sign electric charges and different flavours, while events in the $\ell\tau_{\mathrm{had}}$ channel contain exactly one light lepton and a $\tau_{\text{had-vis}}$ with opposite-sign electric charges. To preserve event separation between the channels, events containing $\tau_{\text{had-vis}}$ are vetoed in the $\ell\tau_{\ell'}$ channel.

For each channel, a *Baseline* selection is applied, based on the properties of the light leptons, $\tau_{\text{had-vis}}$, missing transverse momentum, and on event properties as the absence of b-tagged jets. Events satisfying

Table 2: *Baseline* event selection in the $\ell\tau_{\ell'}$ and $\ell\tau_{\text{had}}$ channels and further splitting into the *VBF* and *non-VBF* categories. The opposite-sign charge of the final-state particles is labelled as OS.

Selection	$\ell\tau_{\ell'}$	$\ell\tau_{\text{had}}$
Baseline	exactly $1e$ and 1μ, OS	exactly 1ℓ and $1\tau_{\text{had-vis}}$, OS
	τ_{had}-veto	τ_{had}Tight ID
		Medium eBDT ($e\tau_{\text{had}}$)
	b-veto	b-veto
	$p_{\text{T}}^{\ell_1} > 45$ (35) GeV MC-template (Symmetry method)	$p_{\text{T}}^{\ell} > 27.3$ GeV
	$p_{\text{T}}^{\ell_2} > 15$ GeV	$p_{\text{T}}^{\tau_{\text{had-vis}}} > 25$ GeV, $\lvert\eta^{\tau_{\text{had-vis}}}\rvert < 2.4$
	$30\,\text{GeV} < m_{\ell_1\ell_2} < 150\,\text{GeV}$	$\sum_{i=\ell,\tau_{\text{had-vis}}} \cos\Delta\phi(i, E_{\text{T}}^{\text{miss}}) > -0.35$
	$0.2 < p_{\text{T}}^{\text{track}}(\ell_2 = e)/p_{\text{T}}^{\text{cluster}}(\ell_2 = e) < 1.25$ (MC-template)	$\lvert\Delta\eta(\ell, \tau_{\text{had-vis}})\rvert < 2$
	track d_0 significance requirement (see text)	
	$\lvert z_0 \sin\theta\rvert < 0.5$ mm	
VBF	*Baseline*	
	≥ 2 jets, $p_{\text{T}}^{\text{j}_1} > 40$ GeV, $p_{\text{T}}^{\text{j}_2} > 30$ GeV	
	$\lvert\Delta\eta_{\text{jj}}\rvert > 3$, $m_{\text{jj}} > 400$ GeV	
non-VBF	*Baseline* plus fail *VBF* categorisation	
	–	veto events if
	–	$90 < m_{\text{vis}}(e, \tau_{\text{had-vis}}) < 100$ GeV

the *Baseline* selection criteria are further classified into two statistically independent categories, *VBF* and *non-VBF*, based on the kinematic properties of the jets produced in association with the Higgs boson candidate. The event selection for all regions is summarised in Table 2. Background control regions (CRs) used in the analyses are described in Section 5 and summarised in Table 5.

In the $\ell\tau_{\ell'}$ channel, the definition of the leading (ℓ_1) and subleading (ℓ_2) light leptons is based on the p_{T}-ordering of the light leptons in the laboratory frame. This ordering is used in the context of the rejection and estimation of backgrounds, where the p_{T} in the laboratory frame is relevant. Additionally, an approach based on the *approximate Higgs boson rest frame* is used for the classification of the events. The four-momentum of the Higgs boson rest frame is built from the two leptons and $E_{\text{T}}^{\text{miss}}$ four-momenta. The pseudorapidity component of the missing-momentum vector is assumed to be the same as that of the system formed by the two charged light leptons, and the resulting invariant mass of the Higgs boson system is constrained to be 125 GeV. The light lepton having the higher p_{T} in the approximate Higgs boson rest frame is called ℓ_H, and is assumed to be the lepton originating from the Higgs boson decay. The other charged light lepton is called ℓ_τ, and is assumed to be the light lepton originating from the τ decay. Events are divided into the $\mu\tau_e$ and $e\tau_\mu$ final states depending on whether ℓ_H is the muon or the electron. Using this approach, the lepton misassignment for the *Baseline* event selection is reduced to about 5% (7%) for the ggF (*VBF*) production mode of the LFV signal, compared with a 11% (19%) misassignment obtained when the laboratory frame is used.

In the $\ell\tau_{\ell'}$ channel of the MC-template method, the leading light lepton (ℓ_1) is required to have $p_{\text{T}}^{\ell_1} > 45$ GeV, while events with lower $p_{\text{T}}^{\ell_1}$ values are used to obtain a region enriched with $Z \to \tau\tau$ events, as described in Section 5.1. The events must satisfy a requirement on the invariant mass of the two final-state leptons, $30\,\text{GeV} < m_{\ell_1\ell_2} < 150\,\text{GeV}$, to reduce the single-top-quark and $t\bar{t}$ (hereafter collectively labelled 'top-quark') background. Additionally, to reduce the top-quark background contribution, events with one or more

identified b-tagged jets are rejected. In order to suppress the contribution from backgrounds with non-prompt light leptons such as heavy-flavoured hadron decays and $Z \to \tau\tau$ processes, and to ensure compatibility with the primary vertex, additional requirements are imposed on the transverse impact parameter (d_0) significance and the longitudinal impact parameter (z_0), weighted by the normalized transverse momentum of the track ($\sin\theta$). The d_0 significance [77] of the ℓ_1 track is required to be smaller than 5, the combination $|z_0 \sin\theta|$ is required to be smaller than 0.5 mm. If ℓ_2 is an electron, the d_0 significance of its track is required to be smaller than 10. When $\ell_2 = e$, the requirement $0.2 < p_{\mathrm{T}}^{\mathrm{track}}/p_{\mathrm{T}}^{\mathrm{cluster}} < 1.25$ on the ratio of the p_{T} measured using only the inner detector, $p_{\mathrm{T}}^{\mathrm{track}}$, to the p_{T} measured in the calorimeter, $p_{\mathrm{T}}^{\mathrm{cluster}}$, aims to reduce the number of $Z \to \mu\mu$ background events, in which one of the muons deposits a significant fraction of its energy in the calorimeter.

In the $\ell\tau_{\ell'}$ channel of the Symmetry method, the requirement on $p_{\mathrm{T}}^{\ell_1}$ is reduced relative to the MC-template method to $p_{\mathrm{T}}^{\ell_1} > 35$ GeV to increase the number of events used in the Neural Network (NN) training, introduced in Section 6. To preserve a symmetric selection, electron $|\eta|$ requirements from the object reconstruction are also applied to muons. Therefore, selected muons are required to have $|\eta| < 1.37$ or $1.52 < |\eta| < 2.47$. The requirement $30\,\mathrm{GeV} < m_{\ell_1\ell_2} < 150\,\mathrm{GeV}$ is used to reduce the top-quark background. The d_0 significance of the tracks of both light leptons is required to be smaller than 10. No requirements on $p_{\mathrm{T}}^{\mathrm{track}}(\ell_2)/p_{\mathrm{T}}^{\mathrm{cluster}}(\ell_2)$ are applied when $\ell_2 = e$ because it would break the symmetry between electrons and muons.

In the $\ell\tau_{\mathrm{had}}$ channel, angular selection criteria $(\cos\Delta\phi(\ell, E_{\mathrm{T}}^{\mathrm{miss}}) + \cos\Delta\phi(\tau_{\text{had-vis}}, E_{\mathrm{T}}^{\mathrm{miss}})) > -0.35$ and $|\Delta\eta(\ell, \tau_{\text{had-vis}})| < 2$ are imposed to reject W+jets and multi-jet production processes. To reduce the top-quark background contribution, events with one or more identified b-tagged jets are rejected. Selected $\tau_{\text{had-vis}}$ are required to satisfy the 'Tight' $\tau_{\text{had-vis}}$ identification criterion. In the $e\tau_{\mathrm{had}}$ channel, the reconstructed $\tau_{\text{had-vis}}$ is required to pass the eBDT 'Medium' working point in order to suppress $Z(\to ee)$+jets processes where one of the electrons is reconstructed as a $\tau_{\text{had-vis}}$ object.

The *VBF* and *non-VBF* categories in each of the $\ell\tau_{\ell'}$ and $\ell\tau_{\mathrm{had}}$ final states give rise to a total of four signal regions (SRs) in each search. The *VBF* category is designed to enhance the sensitivity to the VBF Higgs boson production mode. Dedicated requirements are applied to the jet kinematics and topology of the two jets to separate VBF Higgs boson production from the other production modes. The leading and subleading p_{T}-ordered jets are denoted by j_1 and j_2, respectively. The m_{jj} and $\Delta\eta_{\mathrm{jj}}$ variables represent the invariant mass and η-separation of the two leading jets, respectively. The *non-VBF* category contains events failing the *VBF* selection, and in the $e\tau_{\mathrm{had}}$ channel, an additional requirement on $m_{\mathrm{vis}}(e, \tau_{\text{had-vis}})$ reduces the $Z(\to ee)$+jets background component.

5 Background estimation

The $Z \to \tau\tau$ process is one of the dominant backgrounds in all the categories and channels. Other relevant backgrounds originate from W+jets and multi-jet events with at least one jet misidentified as an electron, muon or τ_{had} (referred to as 'misidentified' hereafter). The misidentified background in the two methods (MC-template and Symmetry) and the two analysis channels ($\ell\tau_{\ell'}$ and $\ell\tau_{\mathrm{had}}$) is evaluated using data-driven techniques. Additionally, top-quark decays form a non-negligible contribution, particularly in the $\ell\tau_{\ell'}$ channel. Other background components such as diboson production (WW, WZ and ZZ), $H \to \tau\tau$ and $H \to WW$ are also considered.

In the following, the background estimates are discussed in more detail for the MC-template $\ell\tau_{\ell'}$ channel in Section 5.1, for the MC-template $\ell\tau_{\text{had}}$ channel in Section 5.2, and for the Symmetry-based $\ell\tau_{\ell'}$ channel in Section 5.3.

5.1 MC-template$\ell\tau_{\ell'}$ channel

In the MC-template $\ell\tau_{\ell'}$ channel, the main background contributions arise from top-quark, $Z \to \tau\tau$ and diboson processes, and events with misidentified leptons. Sources of smaller backgrounds are $Z \to \ell\ell$ events and SM Higgs boson decays. In addition to simulation, data control and validation regions are used to estimate the background contributions, where possible. Backgrounds from Higgs boson processes such as $H \to \tau\tau$ and $H \to WW$ are expected to be small and are normalised to their SM predictions.

The top-quark processes contribute 34%–54% of the total background, depending on the category. For each category, the top-quark simulation is validated with data in a top-quark CR, statistically independent of the SRs. These CRs are defined by applying all the baseline criteria for *non-VBF* or *VBF* categorisation except for the b-tagged jet veto, and 95% of their events are top-quark events. Good modelling by the top-quark simulation is observed, as shown in Figure 2 (a, b). To account for possible theoretical uncertainties in the production cross-section for the simulated samples, the top-quark CRs and two normalisation factors, one per category, are included in the statistical analysis (Section 8). Each normalisation factor is determined during the signal extraction.

The $Z \to \tau\tau$ events account for 23% (11%) of the total background in the *non-VBF* (*VBF*) SR. For each category, the $Z \to \tau\tau$ simulation is validated with data in a $Z \to \tau\tau$ CR, statistically independent of the SRs. These CRs are defined by requiring $35\,\text{GeV} < p_{\text{T}}^{\ell_1} < 45\,\text{GeV}$, and the $Z \to \tau\tau$ process contributes ~65% (~32%) of all events in the *non-VBF* (*VBF*) $Z \to \tau\tau$ CR. Good modelling in the $Z \to \tau\tau$ CRs is illustrated for the *non-VBF* case in Figure 2 (c, d). In the statistical analysis, the $Z \to \tau\tau$ CRs are used jointly with two independent normalisation factors to constrain the $Z \to \tau\tau$ normalisation in the SRs.

Diboson events form 19%–32% of the total background, depending on the category. The shape and normalisation of diboson process distributions are estimated from the simulation and validated with data in a dedicated validation region, where the diboson processes contribute ~67% of the total background. This region is defined by applying the *Baseline* selection criteria and requiring $p_{\text{T}}^{\ell_2} > 30\,\text{GeV}$ and the mass of the two leptons to satisfy $100\,\text{GeV} < m_{\ell_1\ell_2} < 150\,\text{GeV}$. The transverse mass, m_{T}, calculated from the transverse momentum of the subleading light lepton and the $E_{\text{T}}^{\text{miss}}$, is required to be greater than 30 GeV and jets with $p_{\text{T}} > 30$ GeV are rejected. Good modelling in the diboson validation region is observed, as illustrated in Figure 3.

The $Z \to \mu\mu$ process contributes sizeably only in the $\mu\tau_e$ channel, where it represents up to 2% of the total background. The modelling of the $Z \to \mu\mu$ background by the simulation is validated in a dedicated $Z \to \mu\mu$ CR, which is not included in the statistical analysis. This CR is obtained by applying the *Baseline* selection, but with $35\,\text{GeV} < p_{\text{T}}^{\ell_1} < 45\,\text{GeV}$ and with the dilepton pair mass near the Z boson mass peak, $75\,\text{GeV} < m_{\ell_1\ell_2} < 100\,\text{GeV}$. Additionally, the ϕ separation between the subleading light lepton and the $\vec{E}_{\text{T}}^{\text{miss}}$ is required to satisfy $|\Delta\phi(\ell_2, E_{\text{T}}^{\text{miss}})| < 1.5$. To ensure that the $Z \to \mu\mu$ CR is statistically independent of the SRs and the $Z \to \tau\tau$ CRs, the events are required to have $1.25 < p_{\text{T}}^{\text{track}}(\ell_2)/p_{\text{T}}^{\text{cluster}}(\ell_2) < 3$. While no mismodelling of the MC template shape is found in the $Z \to \mu\mu$ CR, a global normalisation offset of 25% is observed. Therefore, the $Z \to \mu\mu$ event yields in the SRs are decreased by 25% and a normalisation uncertainty of that size is assigned to the $Z \to \mu\mu$ contribution in the statistical analysis.

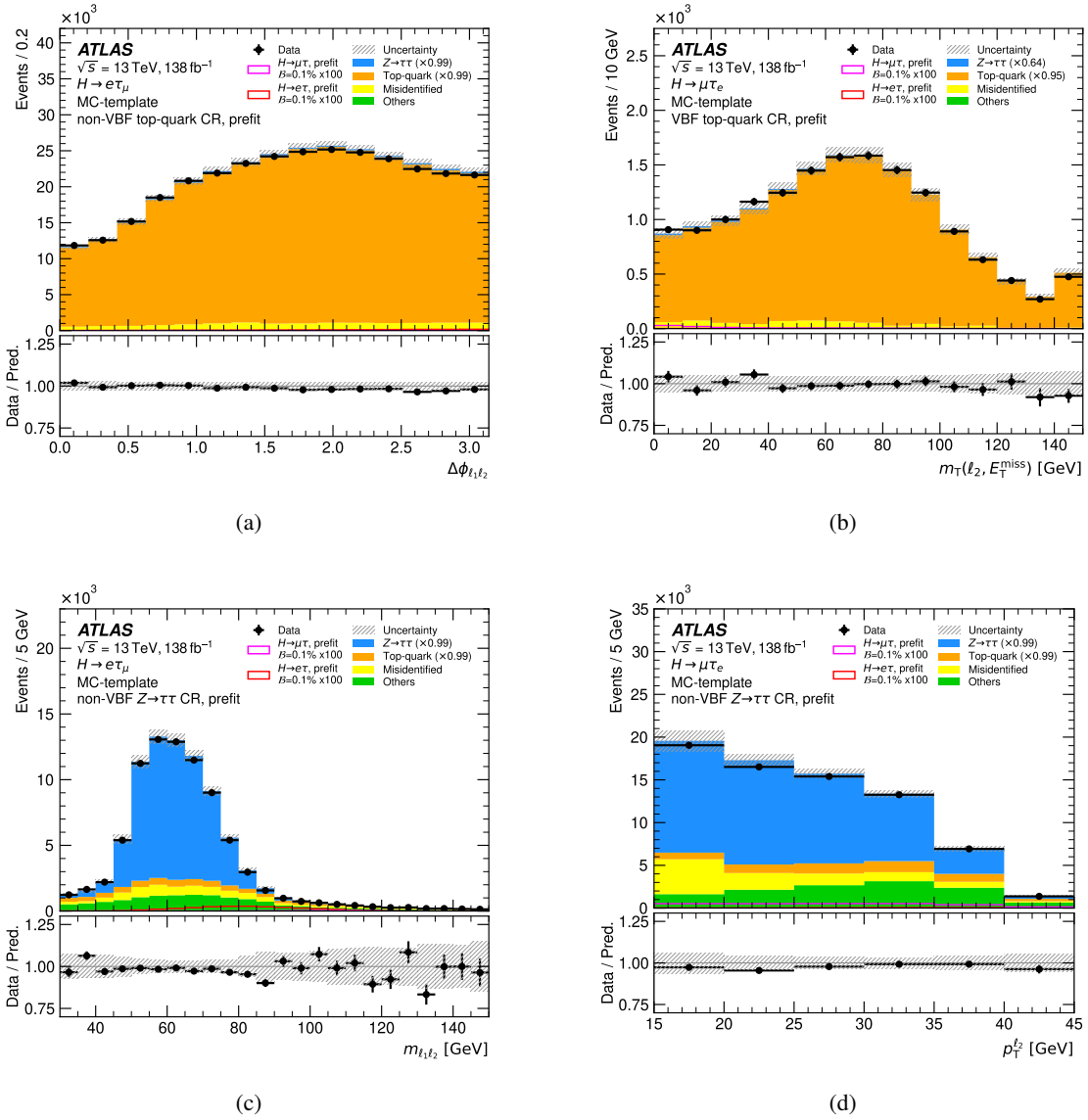
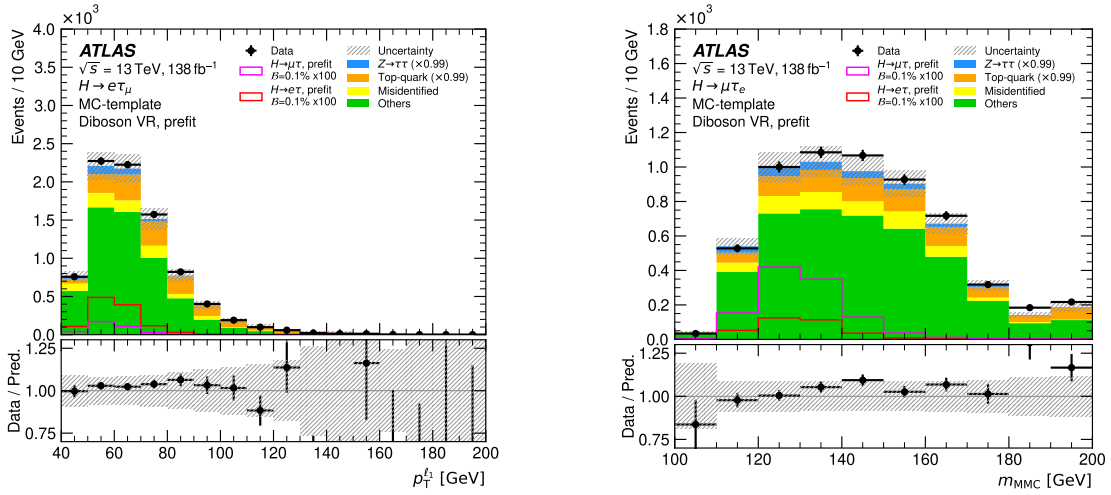


Figure 2: Distributions of representative quantities for the top-quark and $Z \to \tau\tau$ CRs before the statistical analysis (prefit) in the $e\tau_\mu$ (left) and $\mu\tau_e$ (right) final states: (a) $\Delta\phi_{\ell_1,\ell_2}$ and (b) $m_{\mathrm{T}}(\ell_2, E_{\mathrm{T}}^{\mathrm{miss}})$ for the top-quark CRs, (c) $m_{\ell_1\ell_2}$ and (d) $p_{\mathrm{T}}^{\ell_2}$ for the $Z \to \tau\tau$ CRs. Entries with values that would exceed the horizontal axis range are included in the last bin of each distribution. The hashed band indicates the prefit statistical, experimental and theoretical uncertainties. The $Z \to \tau\tau$ and top-quark contributions are scaled by a normalisation factor (indicated in the legend) obtained from the likelihood fit performed independently in *non-VBF* and *VBF* categories of the MC-template $\ell\tau_{\ell'}$ channel. Overlaid prefit signal shapes assume $\mathcal{B}(H \to \ell\tau) = 0.1\%$ and are enhanced by a factor 100 for visibility.

A data-driven method is used to estimate the misidentified background contribution from events having at least one light lepton originating from heavy-flavour decays, photon conversion, a jet or a τ_{had} misidentified as a light lepton. These events originate mostly from W+jets, multi-jet production and top-quark processes. The method used is based on the assumption that the ratio of opposite-sign to same-sign light-lepton events is approximately constant when inverting the isolation quality requirement for the subleading



(a) (b)

Figure 3: Distributions of representative quantities for the diboson validation region before the statistical analysis (prefit): (a) distribution of $p_{\mathrm{T}}^{\ell_1}$ in the $e\tau_\mu$ final state, (b) distribution of the m_{MMC} [85] mass in the $\mu\tau_e$ final state. Entries with values that would exceed the horizontal axis range are included in the last bin of each distribution. The hashed band indicates the prefit statistical, experimental and theoretical uncertainties. The $Z \to \tau\tau$ and top-quark contributions are scaled by a normalisation factor (indicated in the legend) obtained from the likelihood fit performed independently in *non-VBF* and *VBF* categories of the MC-template $\ell\tau_{\ell'}$ channel. Overlaid prefit signal shapes assume $\mathcal{B}(H \to \ell\tau) = 0.1\%$ and are enhanced by a factor 100 for visibility.

light lepton [28]. The ratio, named a transfer factor, is measured in regions enriched in misidentified background. These regions use the SRs' selection criteria, but the lepton isolation requirement is inverted for the subleading light lepton, and the electron is allowed to fail the 'Medium' identification requirement while still passing the 'Loose' requirement. To obtain the misidentified background prediction in the SRs, the transfer factor is applied to same-sign regions. Events in the same-sign regions are required to satisfy the same selection criteria as in the SRs, but have two light leptons with same-sign electric charges. Events containing two prompt light leptons (mostly from diboson events) form a small component of this background and are subtracted using simulation. The transfer factors are calculated independently in the $e\tau_\mu$ and $\mu\tau_e$ final states. In each of the two cases, transfer factors are obtained separately for events passing the single-lepton or dilepton trigger requirement, and separately for events passing or failing the b-veto requirement.

The statistical uncertainty of the misidentified background is separated into four uncorrelated components, based on whether the event is triggered by a single-lepton or dilepton trigger and whether the event passed or failed the b-veto requirement. The systematic components account for several effects. The uncertainty related to the subtraction of the prompt-lepton processes in the same-sign region is determined to be between 6% and 8%. The systematic uncertainty due to possible differences between the jet flavour composition in the misidentified background enriched CRs and the SRs is taken into account. To estimate this uncertainty, the transfer factors are calculated in misidentified background enriched CRs using the simulated samples of the main sources of fake leptons (W+jets and $V\gamma$). The transfer factors are applied to the simulated events passing the *Baseline* selection, except that two same-sign leptons instead of opposite-sign leptons are required. The obtained prediction for the BDT distribution (see Section 6) is

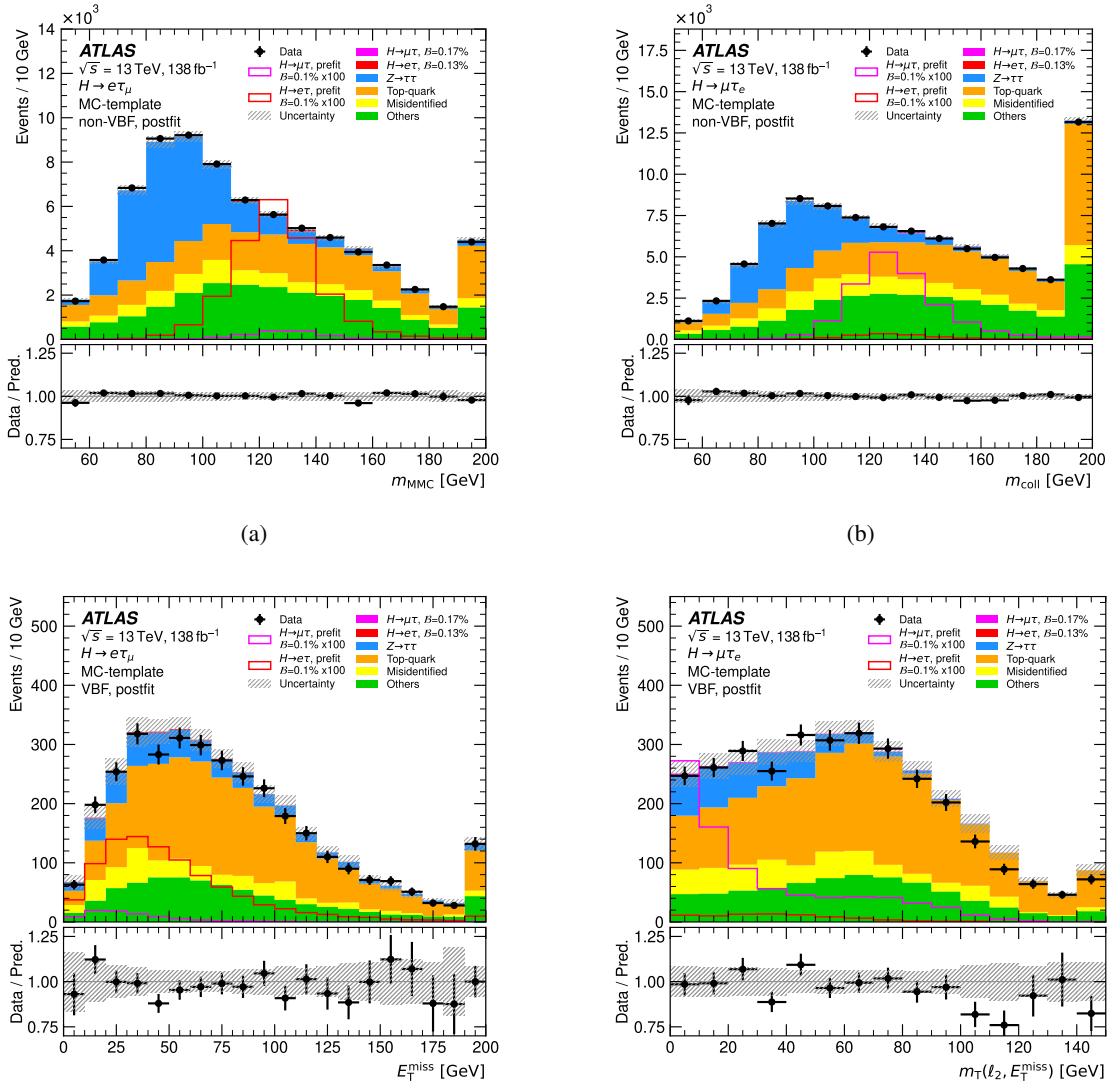


Figure 4: Distributions of representative kinematic quantities for $e\tau_\mu$ (left) and $\mu\tau_e$ (right) final states, after a simultaneous fit of the $H \to e\tau$ and $H \to \mu\tau$ signals, obtained by fitting the data of the MC-template $\ell\tau_{\ell'}$ channel: (a) the mass m_{MMC} and (b) m_{coll} [86] in the *non-VBF* category, (c) the missing transverse momentum $E_{\text{T}}^{\text{miss}}$ and (d) the transverse mass $m_{\text{T}}(\ell_2, E_{\text{T}}^{\text{miss}})$ in the *VBF* category. Entries with values that would exceed the horizontal axis range are included in the last bin of each distribution. The hashed band represents the prefit statistical uncertainty, and the experimental and theoretical uncertainties obtained from the likelihood fit. Overlaid prefit signal shapes assume $\mathcal{B}(H \to \ell\tau) = 0.1\%$ and are enhanced by a factor 100 for visibility. The postfit signal contributions are considered as part of the predictions.

compared with the simulated events passing the *Baseline* selection. The difference between the two BDT distribution shapes is taken as the magnitude of the systematic uncertainty, which is found to vary between 10% and 30%. To cover potential residual differences between the simulation and data, a non-closure uncertainty is obtained by defining additional opposite-sign misidentified background enriched CRs. These

regions are statistically independent of the SRs and use the same selection criteria as the *Baseline* selection but invert the requirement on d_0 significance. The transfer factors are recalculated using these additional regions and the background predictions are re-evaluated using the new transfer factors. The difference between data and the background prediction in the additional opposite-sign misidentified background enriched CRs is assigned as a systematic uncertainty. The magnitude of the uncertainty ranges from 2% to 12%. The dependence of the transfer factor on the d_0 significance requirement was investigated and found to be negligible.

Figure 4 shows the level of agreement between the data and simulation in the *non-VBF* and *VBF* categories for a subset of variables with the highest discrimination power used in the MVA as described in Section 6. The distributions show clear separation between signal and background, as well as good modelling of the data within statistical and systematic uncertainties. The collinear mass of the two final-state leptons (m_{coll}) [86], the Higgs boson mass obtained with the Missing Mass Calculator (MMC) technique (m_{MMC}) [85] and $m_{\ell_1\ell_2}$ show the highest discrimination power. The MMC algorithm is tuned specifically to reconstruct the mass of the LFV Higgs boson. The prediction for each sample is determined from the likelihood fit performed to measure $H \to e\tau$ and $H \to \mu\tau$ signals, based only on data in the $\ell\tau_{\ell'}$ final state. Corresponding SR and CR event yields are detailed in Tables 6 and 7 in Appendix B.1.

5.2 MC-template $\ell\tau_{\text{had}}$ channel

In the MC-template $\ell\tau_{\text{had}}$ channel, the main background contributions arise from the $Z \to \tau\tau$ events and are estimated using the simulated events. The $Z \to \tau\tau$ events form 48%–67% of the total background depending on the category. The second-largest background consists of events containing a jet misidentified as $\tau_{\text{had-vis}}$ and contributes 22%–30% of the total background. This misidentified background is evaluated using the Fake-Factor technique [87, 88]. Contributions with τ_{had} and a jet misidentified as an electron or a muon are estimated from simulation and found to be negligible. Backgrounds from other processes such as top-quark, $Z \to \ell\ell$, dibosons, $H \to \tau\tau$ and $H \to WW$ are much smaller and are estimated from simulation.

The shape of the $Z \to \tau\tau$ background distribution is modelled with simulations, and its normalisation in the *VBF* and *non-VBF* categories is constrained by data in the BDT score distributions (see Section 6) in the SRs. The shape of the top-quark background distribution is modelled with simulation. The corresponding normalisation factors in the *VBF* and *non-VBF* categories are constrained using the $\ell\tau_{\ell'}$ top-quark CRs in the simultaneous fit of the $\ell\tau_{\text{had}}$ and $\ell\tau_{\ell'}$ channels.

Diboson events form 2%–5% of the total background, depending on the category. In the $\mu\tau_{\text{had}}$ channel, the $Z \to \mu\mu$ process contributes about 5%–6% of the total background in the *non-VBF* category, and 3%–4% in the *VBF* category. Its modelling is validated in a dedicated validation region, where the $Z \to \mu\mu$ event fraction is ~65%. The $Z \to \mu\mu$ validation region is defined by applying the *Baseline* selection criteria and also requiring $|\eta(\tau)| < 0.1$ and the collinear mass [86] to be near the Z peak: $90\,\text{GeV} < m_{\text{coll}}(\mu,\tau) < 110\,\text{GeV}$. While no template mismodelling is observed in this region, a global normalisation offset of 13% is observed. Thus, an uncertainty of that size is assigned to the $Z \to \mu\mu$ contribution in the statistical analysis described in Section 8.

The distribution of the misidentified background component in the SR is obtained by multiplying the number of events that satisfy the SR selection criteria but fail the Tight $\tau_{\text{had-vis}}$ identification requirement by a Fake-Factor. The Fake-Factor is defined as the ratio of the number of jets misidentified as $\tau_{\text{had-vis}}$ that satisfy the Tight $\tau_{\text{had-vis}}$ identification criteria to the number that do not, but still satisfy the Very

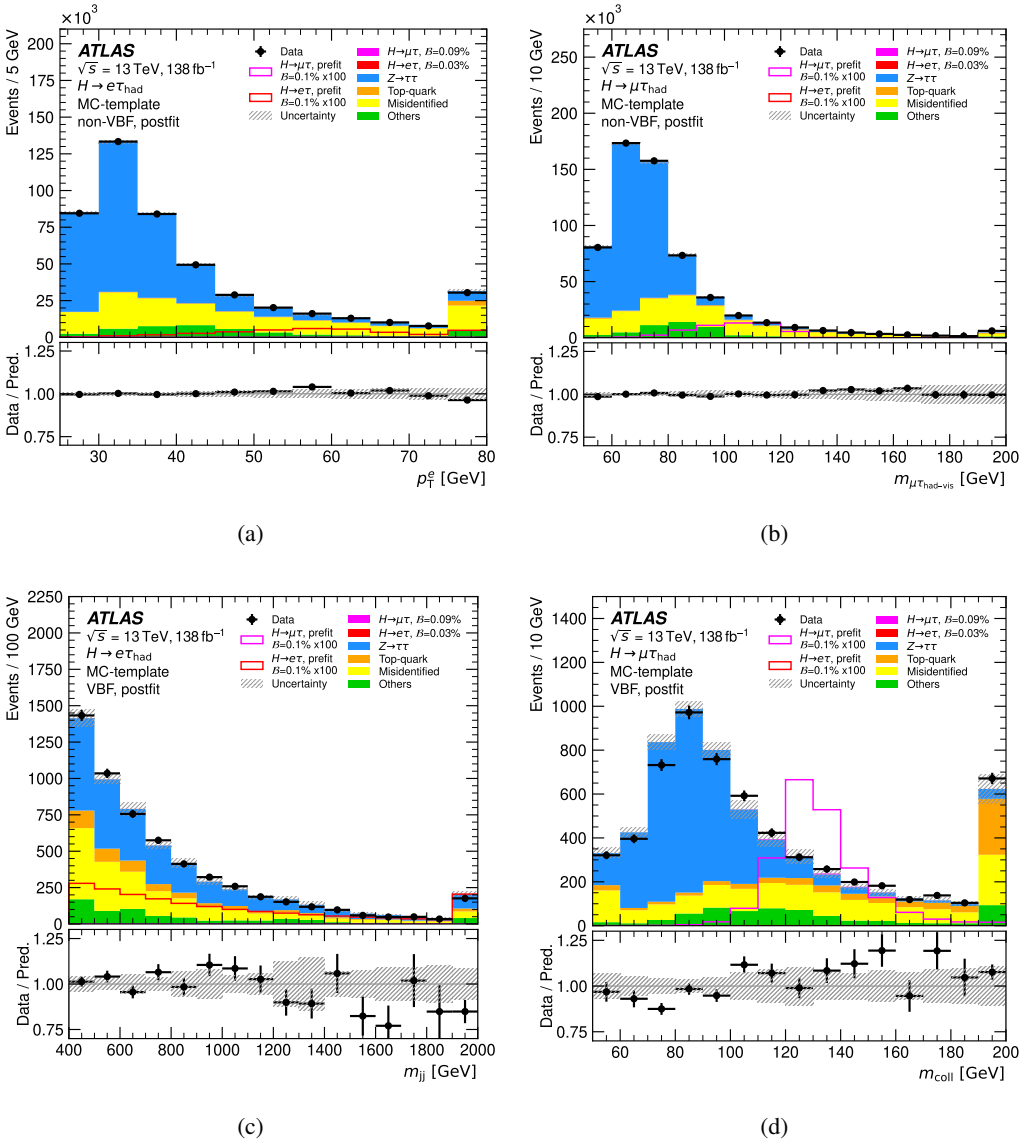


Figure 5: Distributions of representative kinematic quantities for $e\tau_{\text{had}}$ (left) and $\mu\tau_{\text{had}}$ (right) final states, after a simultaneous fit of the $H \to e\tau$ and $H \to \mu\tau$ signals, obtained by fitting the data of the MC-template $\ell\tau_{\text{had}}$ channel: (a) the electron transverse momentum, p_{T}^{e}, and (b) the visible mass m_{vis} in the *non-VBF* category, (c) the dijet invariant mass m_{jj} and (d) collinear mass m_{coll} in the *VBF* category. Entries with values that would exceed the horizontal axis range are included in the last bin of each distribution. The hashed band represents the prefit statistical uncertainty, and the experimental and theoretical uncertainties obtained from the likelihood fit. Overlaid prefit signal shapes assume $\mathcal{B}(H \to \ell\tau) = 0.1\%$ and are enhanced by a factor 100 for visibility. The postfit signal contributions are considered as part of the predictions.

Loose identification requirement. The Fake-Factor is parameterised as a function of the p_{T} and track multiplicity of the $\tau_{\text{had-vis}}$. Since the misidentified background originates from W+jets and multi-jet production processes, two independent Fake-Factors, F_W and F_{QCD}, are measured in dedicated W+jets and multi-jet production CRs, respectively. These regions are defined to be statistically independent of the SRs by inverting several selection criteria. In particular, $(\cos\Delta\phi(\ell, E_{\text{T}}^{\text{miss}}) + \cos\Delta\phi(\tau_{\text{had-vis}}, E_{\text{T}}^{\text{miss}})) < -0.35$,

$m_{\mathrm{T}}(\ell, E_{\mathrm{T}}^{\mathrm{miss}}) > 60$ GeV and $m_{\mathrm{T}}(\tau, E_{\mathrm{T}}^{\mathrm{miss}}) > 40$ GeV for the W+jets CR, and $|\Delta\eta(\ell, \tau_{\text{had-vis}})| \geq 2$ and $m_{\mathrm{T}}(\ell, E_{\mathrm{T}}^{\mathrm{miss}}) < 60$ GeV for the multi-jet CR. To obtain the final estimate of the misidentified background contribution in the SRs, the combined Fake-Factor $F = R_{\mathrm{QCD}}F_{\mathrm{QCD}} + (1 - R_{\mathrm{QCD}})F_W$ is used. The fraction of multi-jet events in the misidentified background, R_{QCD}, is obtained by scaling the number of events in the multi-jet CR by the ratio of the number of events where the light lepton passes the isolation requirements to the number where it does not [45]. This ratio is measured in another region enriched in multi-jet events where the ℓ and τ_{had} have the same electric charge.

The statistical uncertainties obtained from the Fake-Factor and R_{QCD} calculations are included in the statistical analysis. They are considered as independent sources of uncertainty and their magnitude varies between 2% and 3% in the *VBF* category, while it is $< 0.4\%$ in the *non-VBF* category. An additional uncertainty accounts for the residual difference between the misidentified background modelling and the data in the misidentified background enriched same-sign region defined by applying the SR selection criteria, but requiring the τ_{had} and ℓ to have the same charge. The magnitude of the difference between the two regions varies from 2% to 10%, depending on the category, and it is assigned to both the positive and negative component of the systematic uncertainty.

Figure 5 shows the level of agreement between the data and background modelling in the *non-VBF* and *VBF* categories for a subset of variables with the highest discrimination power used in the MVA as described in Section 6. Here the normalisation of signal and background yields is obtained by performing the simultaneous fit of the $H \to e\tau$ and $H \to \mu\tau$ signals (discussed in Section 8.2), based only on data in the $\ell\tau_{\mathrm{had}}$ final state. The distributions show different levels of separation between signal and background, and reasonable modelling of the data within statistical and systematic uncertainties. The corresponding SR event yields are detailed in Table 8 in Appendix B.2.

5.3 Symmetry-based $\ell\tau_{\ell'}$ channel

The data-driven e/μ-symmetry method was first employed in Ref. [89]. This method is sensitive to the difference between $\mathcal{B}(H \to e\tau)$ and $\mathcal{B}(H \to \mu\tau)$, and it is based on the following two assumptions:

1. SM processes are symmetric under the exchange of prompt electrons with prompt muons to a good approximation. As a consequence, the kinematic distributions of prompt electrons and prompt muons are approximately the same;
2. Flavour-violating decays of the Higgs boson break this symmetry.

According to the e/μ-symmetry assumption, the SM background is split equally between the $\mu\tau_e$ and $e\tau_\mu$ datasets. However, the $H \to \mu\tau$ signal is mostly present in the $\mu\tau_e$ dataset, since the p_{T} of the muon originating from the Higgs boson is typically larger than the p_{T} of the electron originating from the decay of the τ-lepton. Similarly, the $H \to e\tau$ signal is mostly present in the $e\tau_\mu$ dataset. Thus, in a search for $H \to \mu\tau$ decays, the SM background can be estimated using the $e\tau_\mu$ dataset, and vice versa.

Detector-related effects cause a distortion of the original e/μ-symmetry and need to be accounted for. The first effect is due to events containing misidentified and non-prompt light leptons, collectively referred to as misidentified background in the following. These particles originate from misidentified jets, misclassified light leptons, hadronic decays within a jet, and electrons from photon conversion. They contribute differently to the $\mu\tau_e$ and $e\tau_\mu$ datasets. The second effect is due to the muons and electrons having different trigger, reconstruction, identifications and isolation efficiencies which depend differently on kinematic properties such as p_{T}, $|\eta|$, and ϕ.

Events containing misidentified light leptons

The misidentified light leptons are subdivided into two contributions based on their estimation method: F_{MC} and F_{FF} for contributions estimated from the simulation and from the data using the Fake-Factor method, respectively.

Events contributing to F_{MC} include τ_{had} that are misidentified as light leptons, muons that are misidentified as electrons, and photons that are misidentified as electrons. Events containing one or more misidentified light leptons are estimated from simulation, taking into account the following processes: $Z \to \tau\tau$, top-quark, diboson, $Z \to \mu\mu$, $Z \to ee$ and $V\gamma$. Contributions from $H \to \tau\tau$ and $H \to WW$ are negligible. The most significant contribution to F_{MC} is from $Z \to \mu\mu$ where low-p_{T} muons are misidentified as electrons. The modelling of this process (total cross-section and shapes of various kinematic distributions) is validated in a dedicated validation region dominated by $Z \to \mu\mu$ events. This region's definition is similar to the one described in Section 5.1, replacing the requirement $\Delta\phi(\ell_\tau, E_{\text{T}}^{\text{miss}}) < 1.5$ with a requirement of $m_{\text{coll}} < 115$ GeV. The systematic uncertainty associated with the normalisation of this background is determined to be 16%. Conservatively, this uncertainty is applied to all the F_{MC} contributions.

Events contributing to F_{FF} include jets misidentified as light leptons and light leptons originating from hadronic decays within a jet. These mostly originate from W+jets processes. A smaller contribution originates from multi-jet production. This background is estimated using the Fake-Factor method where the Fake-Factors are measured in a dedicated Z+jets CR enriched in misidentified light leptons. This Z+jets CR is based on the selection of exactly three electrons or muons where two are required to be consistent with a Z-boson decay, while the third is assumed to originate from a jet which is misidentified either as an electron or muon and is used for the determination of the Fake-Factors. A difference in jet flavour composition between the Z+jets and *Baseline* regions could result in different Fake-Factors. This is taken into account by using a set of correction factors measured in simulation as the ratio of the Fake-Factors in the two regions.

The Fake-Factors and correction factors are binned as a function of lepton p_{T}, with electron Fake-Factors also being parameterised with respect to $\Delta\phi\left(e, E_{\text{T}}^{\text{miss}}\right)$. The product of the Fake-Factors and correction factors is applied as event weights, depending on the kinematic quantities of the non-prompt light leptons that fail to satisfy the lepton identification criteria. Contributions, where either the electron or muon is misidentified, are estimated separately and summed; the numbers of events where both light leptons are misidentified are treated specifically to avoid double counting.

The uncertainties associated with the Fake-Factor method consist of statistical and systematic components. The statistical component is estimated for each Fake-Factor bin separately and consists of the statistical uncertainties from the data and the subtracted prompt-lepton simulated yields within the Z+jets CR propagated to the Fake-Factors. The systematic component corresponds to the uncertainty from the subtraction of events containing three real light leptons, i.e. from WZ and ZZ processes where the subtracted yields are varied by the theory cross-section uncertainties of both processes separately. Uncertainties associated with the correction factors include a contribution from the limited number of simulated events in the Z+jets and *Baseline* regions, and from the choice of a specific MC generator. In the statistical model (described in Section 8), the statistical sources are treated as uncorrelated between Fake-Factor or correction factor bins, the systematic sources as correlated between bins, and the MC subtraction-uncertainty also as correlated between electron and muon misidentified background.

Efficiency correction

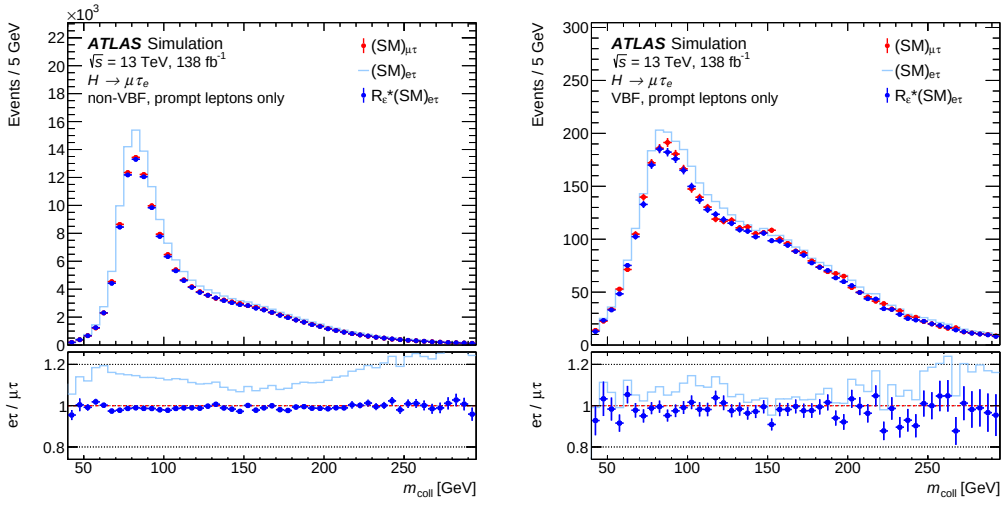
The efficiency to measure a $\mu\tau_e$ event ($\epsilon^{\mu\tau_e}$) or a $e\tau_\mu$ event ($\epsilon^{e\tau_\mu}$) is given by the product of the event trigger efficiency and the muon and electron efficiencies. The trigger efficiency calculation and the associated systematic uncertainties are described in Ref. [90]. Simulated events are scaled using dedicated scale factors to account for the modelling of trigger efficiency in the simulation.

The lepton efficiency is a product of the reconstruction, identification and isolation efficiencies. For the muons, these are measured in Ref. [91] along with their uncertainties and found to be independent of the event selection. For the electrons, regardless of the parameterisation used, the actual efficiencies are strongly dependent on the event selection, and need to be measured for the analysis-specific phase space. Simulated events are corrected to match the data efficiencies by using dedicated scale factors.

In this study, the electron efficiencies are estimated from the simulated events passing the *Baseline* event selection with the following two changes: the requirement $p_{\mathrm{T}}^{\ell_1} > 35\,\mathrm{GeV}$ is relaxed to $p_{\mathrm{T}}^{\ell_1} > 20\,\mathrm{GeV}$, in order to increase the number of events, and only events that contain either one electron and one muon candidate, or two electron candidates, are considered.

The number of events available after applying this selection determines the statistical uncertainty associated with the electron efficiency measurement. Systematic uncertainties associated with this measurement are estimated by comparing the efficiencies obtained when using either the electron or muon as the tag lepton, and by raising the p_{T} selection for ℓ_1 back to 35 GeV. The efficiencies are scaled to account for possible differences observed between data and simulation, using the set of scale factors and associated uncertainties described in Ref. [68] available for each efficiency separately.

The efficiency corrections are applied per event, depending on the kinematic properties of the electron and muon. Each event in the $e\tau_\mu$ sample is used to mimic an event in the $\mu\tau_e$ sample with the same kinematic properties by scaling it with the efficiency ratio $R_\epsilon = \epsilon^{\mu\tau_e}/\epsilon^{e\tau_\mu}$. The same procedure is applied with the inverted ratio when scaling $\mu\tau_e$ to $e\tau_\mu$ events. The procedure is tested in simulated events, as shown in Figure 6, by applying the efficiency correction to $e\tau_\mu$ events and comparing them with $\mu\tau_e$ events. By construction, the efficiency correction is only valid for events containing prompt leptons. In the simulation, events containing misidentified leptons are vetoed. In the data, contributions from both F_{MC} and F_{FF} are subtracted before the efficiency correction is applied. As shown in Figure 6, this procedure restores the symmetry between the two channels in simulation.



(a) (b)

Figure 6: The collinear mass [86] distributions, m_{coll}, in simulated events satisfying the (a) *non-VBF* and (b) *VBF* selection criteria. Events with a leading muon and a subleading electron from the τ-lepton decay ($(\text{SM})_{\mu\tau}$) are compared with events with a leading electron and a subleading muon without ($(\text{SM})_{e\tau}$) or with the efficiency correction applied ($R^*_\epsilon(\text{SM})_{e\tau}$). The bottom panel shows the ratio of the $e\tau$ and $\mu\tau$ predictions. The ratio for the $e\tau$ prediction with the efficiency correction applied ($R^*_\epsilon(\text{SM})_{e\tau}/(\text{SM})_{\mu\tau}$) is shown by dark blue points, while the $e\tau$ prediction without the efficiency correction applied ($(\text{SM})_{e\tau}/(\text{SM})_{\mu\tau}$) is shown by the light blue line. Uncertainties included are statistical only.

Background estimation

The number of background events in each dataset ($e\tau_\mu$ and $\mu\tau_e$) is estimated as a sum of three contributions: a symmetric component estimated by applying the efficiency correction procedure to the events in the other dataset, F_{MC} and F_{FF}. Since the efficiency correction is only valid for true leptons, the contribution from events containing misidentified leptons is subtracted from the other dataset before applying the efficiency correction.

Figure 7 compares the data and the background prediction for a selected set of kinematic variables with strong signal-vs-background discrimination power for the $e\tau_\mu$ and $\mu\tau_e$ datasets in the *non-VBF* and *VBF* categories. In the figures, the F_{FF} contribution is labelled as 'Misidentified', while F_{MC} contribution is labelled as 'Other'. The normalisation of signal and background yields is obtained by performing the independent fits of the $H \to e\tau$ and $H \to \mu\tau$ signals (discussed in Sections 8.1 and 8.3), based only on data in the $\ell\tau_{\ell'}$ final state. The distributions show either $H \to e\tau$ or $H \to \mu\tau$ signal, with clear shape differences to the backgrounds, particularly for mass-based variables, and good agreement of the prediction with data within statistical and systematic uncertainties. The corresponding SR event yields are shown in Table 9 in Appendix B.3.

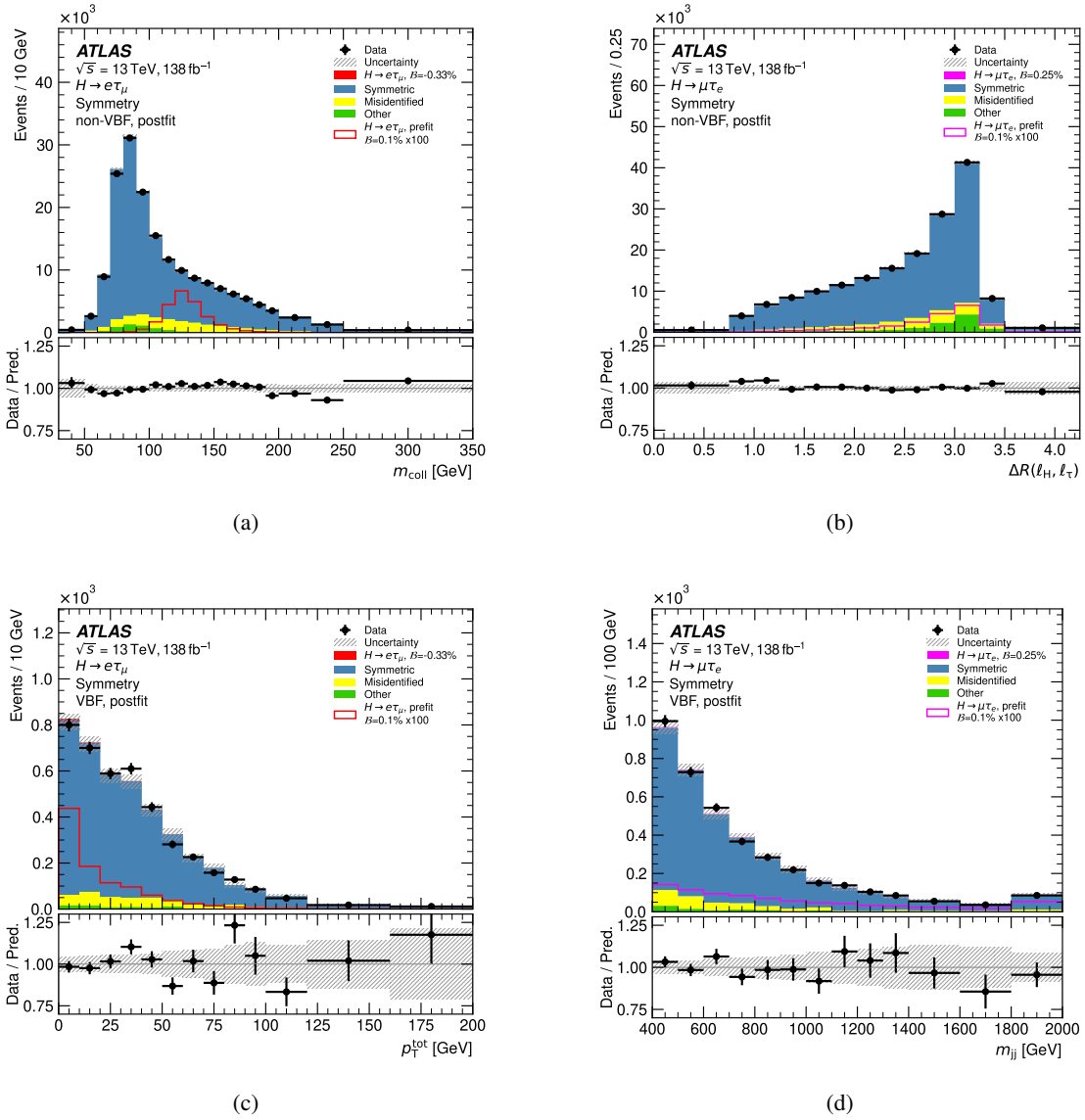


Figure 7: Distributions of representative kinematic quantities for $e\tau_\mu$ (left) and $\mu\tau_e$ (right) final states, after an independent fit of the $H \to e\tau$ and $H \to \mu\tau$ signals, respectively, obtained by fitting the data of the Symmetry $\ell\tau_{\ell'}$ channel: (a) m_{coll} and (b) the angular separation between ℓ_H and ℓ_τ, $\Delta R(\ell_H, \ell_\tau)$ in the *non-VBF* category, (c) $p_{\text{T}}^{\text{tot}}$, which is the absolute value of the vectorial sum of p_{T} of the two leptons, the two leading jets and the $E_{\text{T}}^{\text{miss}}$, and (d) the invariant mass of the dijet system, m_{jj}, in the *VBF* category. Entries with values that would exceed the horizontal axis range are included in the last bin of each distribution. The hashed band represents the prefit statistical uncertainty, and the experimental and theoretical uncertainties obtained from the likelihood fit. Overlaid prefit signal shapes assume $\mathcal{B}(H \to \ell\tau) = 0.1\%$ and are enhanced by a factor 100 for visibility. The postfit signal contributions are considered as part of the predictions.

6 Multivariate analysis strategy

In order to separate signal from various background contributions, two MVA techniques are used: boosted decision trees (BDTs), exploited in the MC-template method, and fully connected deep neural networks (NNs), used in the Symmetry method. Different MVA techniques are used for the *non-VBF* and *VBF* categories to exploit the VBF topology. The MVA technique with the best expected sensitivity was chosen for each background estimation method. The BDTs are trained using the Toolkit for Multivariate Data Analysis [92], while the NNs are trained using Keras (v2.2) [93], with Tensorflow (v1.12) [94] as the backend.

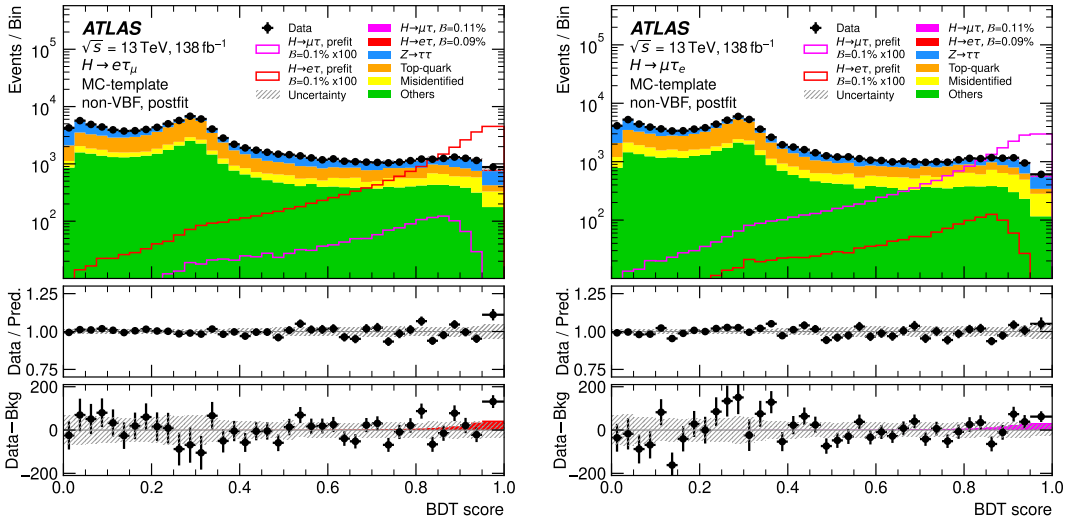
In the $\ell\tau_{\ell'}$ channel of the MC-template method, three different BDTs are utilised: $\text{BDT}_1^{\ell\tau_{\ell'}}$ is devoted to separating the signal from events with misidentified leptons; $\text{BDT}_2^{\ell\tau_{\ell'}}$ separates the signal from top-quark background processes, diboson production and $H \to WW$ decays; and $\text{BDT}_3^{\ell\tau_{\ell'}}$ separates the signal from $Z \to \tau\tau$, $Z \to \ell\ell$ and $H \to \tau\tau$ backgrounds. The three BDTs were utilised because they provided better expected signal significance than a single BDT. To increase the number of events in the BDT training, the $e\tau_\mu$ and the $\mu\tau_e$ datasets are combined. The resulting BDT scores are combined into a single score using a linear weighted sum, with the weights optimised using the expected signal significance as a figure-of-merit.

In the $\ell\tau_{\text{had}}$ channel of the MC-template method, two BDTs are used in all cases except for the *non-VBF* category in the $e\tau_{\text{had}}$ channel, where three BDTs are used. Two/three BDTs were chosen instead of a single BDT because they provide better expected signal significance. In the case where three BDTs are trained, $\text{BDT}_1^{\ell\tau_{\text{had}}}$ discriminates the signal from the $Z \to \tau\tau$ events; $\text{BDT}_2^{\ell\tau_{\text{had}}}$ separates the signal from misidentified background; and $\text{BDT}_3^{\ell\tau_{\text{had}}}$ separates the signal from the rest of the backgrounds. In the case where two BDTs are trained, $\text{BDT}_{1'}^{\ell\tau_{\text{had}}}$ separates the signal from $Z \to \tau\tau$ events and $\text{BDT}_{2'}^{\ell\tau_{\text{had}}}$ separates the signal from the rest of the backgrounds. The resulting BDT scores are combined into a single score using a linear weighted sum in the *non-VBF* categories, and a quadratic weighted sum in the *VBF* categories. The latter provided a 4% improvement in the expected signal significance, used as a figure-of-merit, for the *VBF* categories; in the *non-VBF* categories, the linear combination provided the better expected signal significance.

In the $\ell\tau_{\ell'}$ channel of the Symmetry method, the requirement on the jet invariant mass (m_{jj}) in the *VBF* category is lowered to 300 GeV to increase the size of the training set. The depth of the NN is chosen taking into account the number of events available for the training. For the *non-VBF* category, a multiclass classifier, with three output classes, is used. The three classes correspond to signal, flavour-symmetric SM background and misidentified background. The F_{MC} contribution is added to the flavour-symmetric SM background class since the distributions of its main processes, $Z \to \mu\mu$, $Z \to \tau\tau$ and $V\gamma$, are more similar to the distributions of the flavour-symmetric SM background processes than to the ones originating from jets misidentified as a lepton. For the *VBF* category, three single binary classification networks are trained and the resulting output distributions are combined via a linear combination to obtain a single discriminant. The three networks are trained to separate signal events from: (i) F_{MC}, $Z \to \tau\tau$ and $H \to \tau\tau$, (ii) top-quark, diboson and $H \to WW$ and (iii) F_{FF} events.

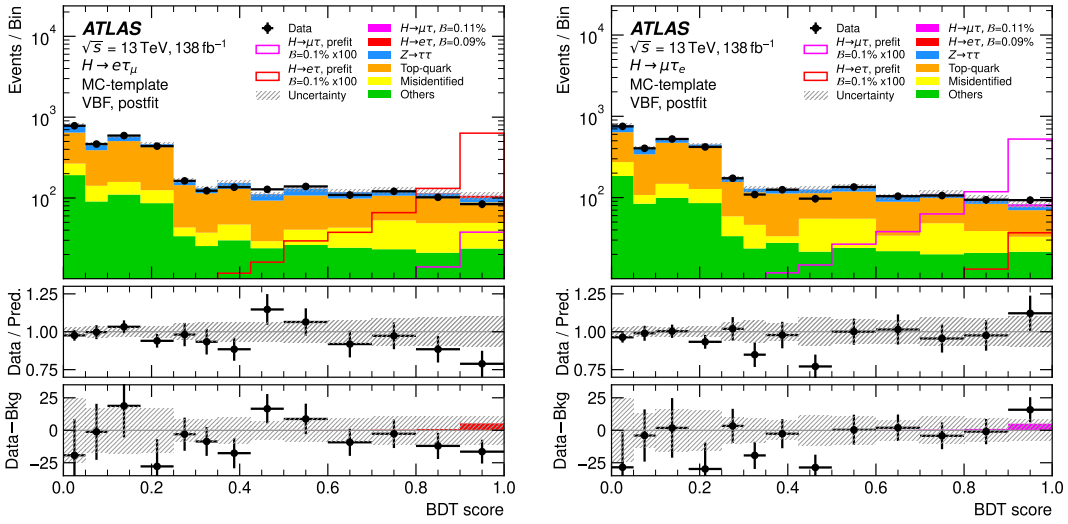
Details of the MVA optimisation procedure, as well as the input variables used in the training, are described in Appendix C. The distributions of the resulting MVA discriminants are shown in Figures 8, 9 and 10 for the MC-template $\ell\tau_{\ell'}$, MC-template $\ell\tau_{\text{had}}$ and Symmetry $\ell\tau_{\ell'}$ channels, respectively. The MC-template method distributions are obtained after performing the simultaneous fit of the $H \to e\tau$ and $H \to \mu\tau$ signals described in Section 8.2. The Symmetry method distributions use the set-up for independent searches for

$H \to e\tau$ and $H \to \mu\tau$ (see Section 8.1), setting the other signal to zero in each case, and are based on $\ell\tau_{\ell'}$ final-state data, which pass the Symmetry-method-specific selection. The conversion of the fit of the branching ratio difference into individual branching ratios results in positive contributions for $H \to \mu\tau$ and negative contributions for $H \to e\tau$, as discussed in Section 5.3. This is indicated with a positive or negative sign in Figure 10.



(a)

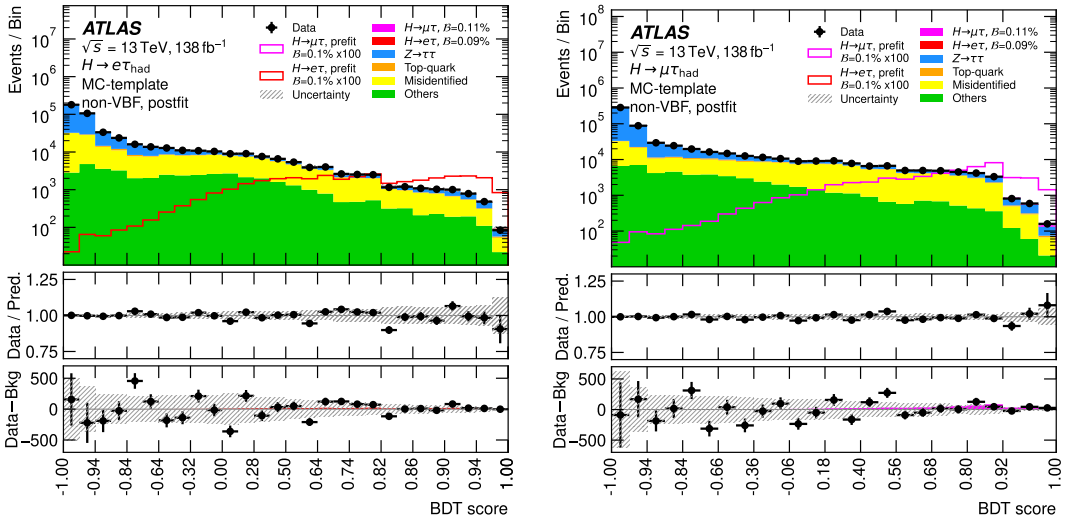
(b)



(c)

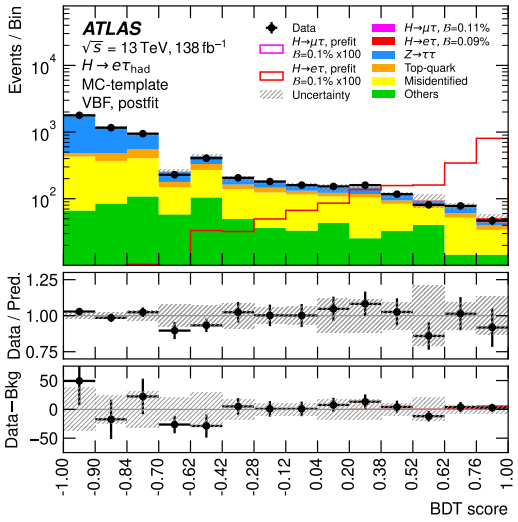
(d)

Figure 8: BDT score distributions for the $e\tau_\mu$ (left) and the $\mu\tau_e$ (right) final states in the *non-VBF* (top) and *VBF* (bottom) categories of the MC-template $\ell\tau_{\ell'}$ channel. The middle panel shows the ratio of data to prediction (background+signal). The bottom panel shows the residuals between data and the background after the fit. The hashed band indicates the total postfit uncertainty of the total predicted yields. The prediction for each sample is determined from the likelihood fit performed to measure $H \to e\tau$ and $H \to \mu\tau$ signals simultaneously, described in Section 8.2. The binning is shown as used in the likelihood fit. Overlaid prefit signal shapes assume $\mathcal{B}(H \to \ell\tau) = 0.1\%$ and are enhanced by a factor 100 for visibility. The postfit signal contributions are considered as part of the predictions.

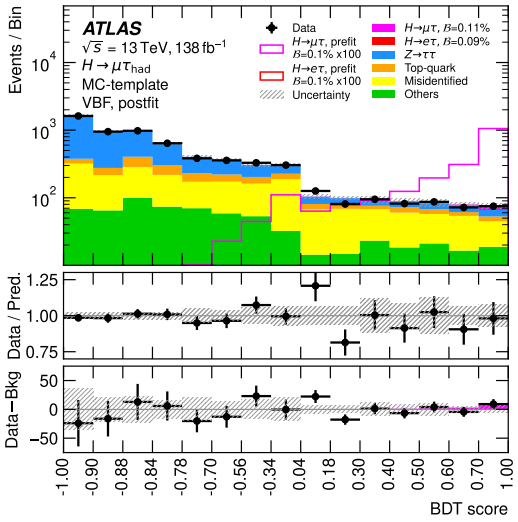


(a)

(b)



(c)



(d)

Figure 9: BDT score distributions for the $e\tau_{\text{had}}$ (left) and the $\mu\tau_{\text{had}}$ (right) final states in the *non-VBF* (top) and *VBF* (bottom) categories of the MC-template $\ell\tau_{\text{had}}$ channel. The middle panel shows the ratio of data to prediction (background+signal). The bottom panel shows the residuals between data and the background after the fit. The hashed band indicates the total postfit uncertainty of the total predicted yields. The prediction for each sample is determined from the likelihood fit performed to measure $H \to e\tau$ and $H \to \mu\tau$ signals simultaneously, described in Section 8.2. The binning is shown as used in the likelihood fit. Overlaid prefit signal shapes assume $\mathcal{B}(H \to \ell\tau) = 0.1\%$ and are enhanced by a factor 100 for visibility. The postfit signal contributions are considered as part of the predictions.

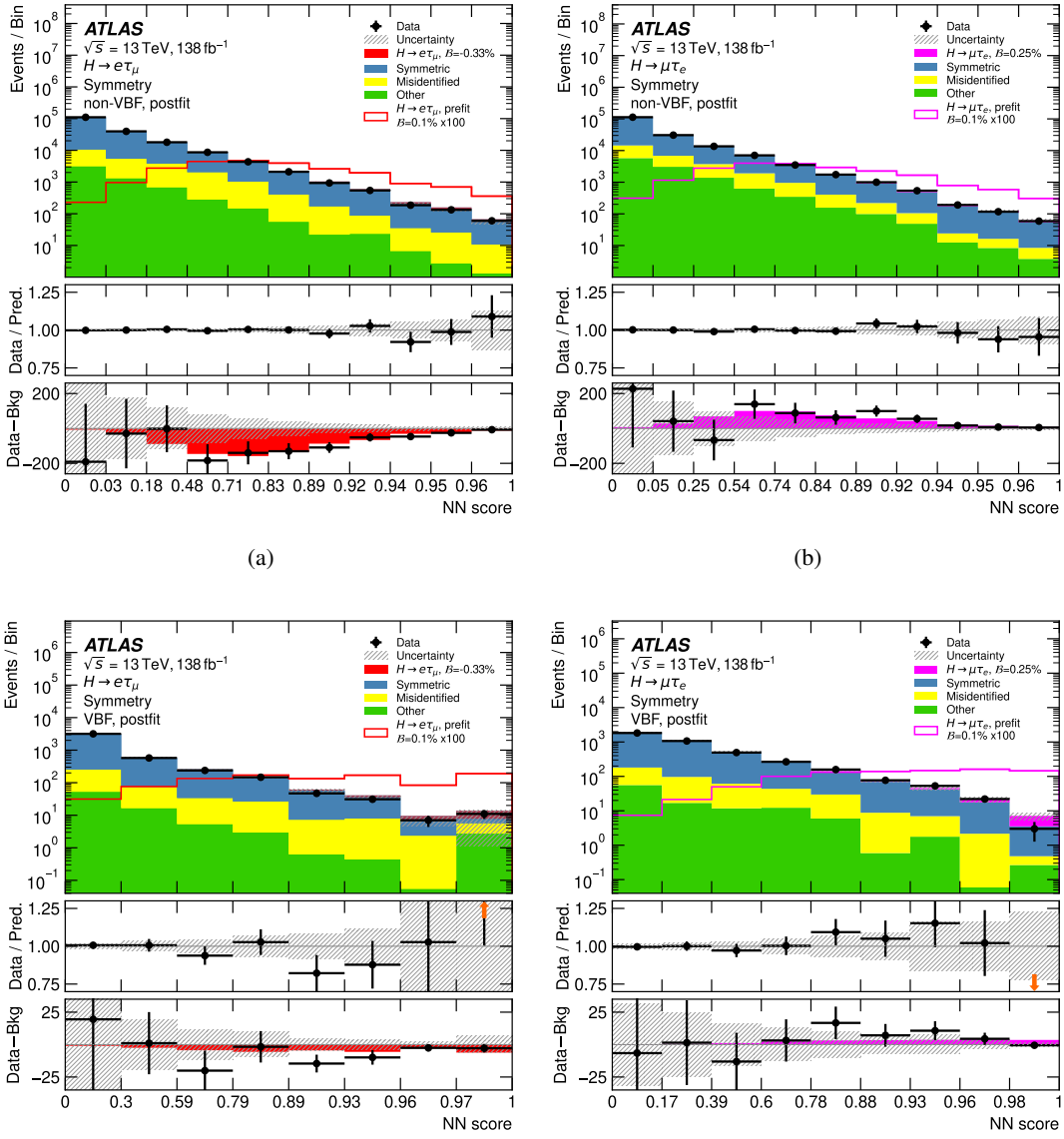


Figure 10: NN score distributions for the $e\tau_\mu$ (left) and the $\mu\tau_e$ (right) final states, in the *non-VBF* (top) and *VBF* (bottom) categories of the Symmetry $\ell\tau_{\ell'}$ channel. The middle panel shows the ratio of data to prediction (background+signal). The bottom panel shows the residuals between data and the background after the fit. The hashed band indicates the total postfit uncertainty on the total predicted yields. The prediction for each sample is determined from the likelihood fit performed to measure $H \to e\tau$ (left) and $H \to \mu\tau$ (right) signals, described in Section 8.1. The binning is shown as used in the likelihood fit and points outside the displayed y-axis range are indicated by arrows. Overlaid prefit signal shapes assume $\mathcal{B}(H \to \ell\tau) = 0.1\%$ and are enhanced by a factor 100 for visibility. The postfit signal contributions are considered as part of the predictions. As a result of the conversion of the fits of branching ratio differences to individual branching ratios, the upward deviation in $\mu\tau_e$ appears as the downward deviation in $e\tau_\mu$.

7 Systematic uncertainties

Systematic uncertainties affect the yields in the signal and control regions, as well as the shape of the MVA score distribution. They can be separated into three groups: experimental uncertainties, theoretical uncertainties for the backgrounds, and theoretical uncertainties for the signal. The systematic uncertainties related to the estimation of misidentified background are discussed in Section 5 for the MC-template $\ell\tau_{\ell'}$ (Section 5.1), the MC-template $\ell\tau_{\text{had}}$ (Section 5.2) and the Symmetry method (Section 5.3).

Experimental uncertainties include those originating from the trigger, reconstruction, identification and isolation efficiencies and affecting final-state objects such as electrons [90, 95], muons [91], $\tau_{\text{had-vis}}$ [81], jets [75, 96–98] and $E_{\text{T}}^{\text{miss}}$ [84]. Their energy scale and resolution uncertainties are taken into account as well. Experimental uncertainties affect the shape of the MVA score distribution, the background yields and the signal cross-section through their effects on the acceptance and the migration between different analysis categories. An additional uncertainty from the measurement of the luminosity [99, 100], amounting to 1.7%, is included.

In the MC-template channels, theoretical uncertainties are considered for the background processes estimated from simulation [45]. Their effect on the normalisation and shape of the MVA discriminant is considered in the statistical analysis. For Z + jets events, systematic uncertainties include those due to renormalisation (μ_{r}), factorisation (μ_{f}) and resummation scale (μ_{qsf}), the jet-to-parton matching scheme (CKKW) [101], and the choice of α_{s} value and the PDFs. For the top-quark background, uncertainties related to the choice of matrix element and parton shower generators [102, 103], the initial- and final-state radiation model [104], and the PDFs are considered [105]. For the diboson production processes, an uncertainty of 6% is assigned to the cross-section in the statistical analysis [106–108].

With the Symmetry method, these background contributions are estimated from data, and it is not necessary to consider the aforementioned theoretical uncertainties.

The Higgs boson production cross-section uncertainties are obtained from Ref. [109]. Effects on the signal expectations are treated as uncorrelated between production modes. Theoretical uncertainties affecting the ggF, VBF, WH, ZH and $t\bar{t}H$ production cross-sections are considered; more details can be found in Ref. [45]. The uncertainties include components for those estimated by varying the PDF or α_{s} value, or varying the choice of matrix element generator or parton shower and hadronisation model. For the matrix element variation, predictions by Powheg Box v2 are compared with those by MadGraph5_aMC@NLO [110]. The parton shower and hadronisation model variation replaces the nominal Pythia 8 simulation with Herwig 7 [102, 103]. For both the MC-template and Symmetry methods, the Higgs boson production cross-section uncertainties are considered for SM decays as well as for the $H \to e\tau$ and $H \to \mu\tau$ signals (ggF, VBF, WH, ZH production modes only).

Table 3 lists uncertainties for each measurement grouped by their respective sources. In the independent measurements of the $H \to e\tau$ and $H \to \mu\tau$ signal (discussed in Section 8.1), labelled as 1 POI, the results based on the Symmetry method are included. Hence, statistical uncertainties in the background estimate which result from the symmetric component of the background estimate play an important role. In the simultaneous measurement of the $H \to e\tau$ and $H \to \mu\tau$ signal (discussed in Section 8.2), labelled as 2 POI, the uncertainties are dominated by systematic sources, in particular the misidentified background.

Table 3: Summary of the different sources of uncertainty affecting the observed $\mathcal{B}(H \to \ell\tau)$ and their impact as computed by the independent fits (1 POI) described in Section 8.1 and the simultaneous fit (2 POI) described in Section 8.2. The values in the table are multiplied by a factor 100 to improve their readability. Experimental uncertainties for reconstructed objects combine efficiency and energy/momentum scale and resolution uncertainties. 'Background sample size' includes the bin-by-bin statistical uncertainties in the simulated backgrounds as well as statistical uncertainties in misidentified backgrounds, which are estimated using data.

1 POI	Impact on observed $[10^{-4}]$	
Source of uncertainty	$\hat{\mathcal{B}}(H \to e\tau)$	$\hat{\mathcal{B}}(H \to \mu\tau)$
Flavour tagging	0.6	0.4
Misidentified background ($\ell\tau_{\text{had}}$)	2.1	1.5
Misidentified background ($\ell\tau_{\ell'}$)	2.9	1.6
Jet and $E_{\text{T}}^{\text{miss}}$	1.1	1.1
Electrons and muons	0.2	0.5
Luminosity	0.6	0.5
Hadronic τ decays	0.9	1.0
Theory (signal)	0.9	0.7
Theory (Z + jets processes)	1.0	1.2
Theory (top-quark processes)	0.3	0.3
Theory (diboson processes)	0.4	0.7
$Z \to \ell\ell$ normalisation	0.2	0.7
Symmetric background estimate	0.2	0.1
Background sample size	4.2	2.4
Total systematic uncertainty	5.3	3.9
Data sample size	2.9	2.7
Total	6.1	4.7

2 POI	Impact on observed $[10^{-4}]$	
Source of uncertainty	$\hat{\mathcal{B}}(H \to e\tau)$	$\hat{\mathcal{B}}(H \to \mu\tau)$
Flavour tagging	0.7	0.2
Misidentified background ($e\tau_{\text{had}}$)	2.1	0.3
Misidentified background ($e\tau_{\mu}$)	5.8	0.3
Misidentified background ($\mu\tau_{\text{had}}$)	0.6	1.4
Misidentified background ($\mu\tau_{e}$)	0.9	1.1
Jet and $E_{\text{T}}^{\text{miss}}$	1.2	0.9
Electrons and muons	1.4	0.5
Luminosity	0.6	0.4
Hadronic τ decays	0.9	0.9
Theory (signal)	0.8	0.8
Theory (Z + jets processes)	0.8	1.0
$Z \to \ell\ell$ normalisation ($e\tau$)	<0.1	<0.1
$Z \to \ell\ell$ normalisation ($\mu\tau$)	0.2	0.9
Background sample size	3.7	2.3
Total systematic uncertainty	5.1	3.6
Data sample size	3.0	2.7
Total	5.9	4.5

8 Statistical analysis and results

The statistical analysis uses a likelihood function $\mathcal{L}(\mu, \theta)$, constructed as a product of Poisson probability terms over all bins considered in the search. These include the MVA score distributions of all the SRs and the event yields in the $\ell\tau_{\ell'}$ MC-template method CRs, when included, to constrain the normalisations of the major backgrounds estimated from simulation, in particular the top-quark and $Z \to \tau\tau$ background components. The likelihood function depends on the parameters of interest (POIs), $\mu_{e\tau}$ and $\mu_{\mu\tau}$, defined as the branching ratios $\mathcal{B}(H \to e\tau)$ and $\mathcal{B}(H \to \mu\tau)$, respectively, and a set of nuisance parameters θ that encode the effect of systematic uncertainties on the signal and background expectations. All nuisance parameters are implemented in the likelihood function as Gaussian or log-normal constraints. The latter are used for normalisation factors to ensure that they are always positive.

Table 4: Combinations of channels, regions and categories used in the statistical analysis. In the independent searches, 1 POI (see Section 8.1), $H \to e\tau$ and $H \to \mu\tau$ signals are fit separately. When performing the simultaneous measurement, 2 POI (see Section 8.2), $H \to e\tau$ and $H \to \mu\tau$ signals are fit simultaneously.

Method	Channel	Category	Region	1 POI fit	2 POI fit
MC-template	$\ell\tau_{\ell'}$	*non-VBF*	SR	✓	✓
			$Z \to \tau\tau$ CR	✓	✓
			Top-quark CR	✓	✓
		VBF	SR		✓
			$Z \to \tau\tau$ CR		✓
			Top-quark CR		✓
MC-template	$\ell\tau_{\text{had}}$	*non-VBF*	SR	✓	✓
		VBF	SR	✓	✓
Symmetry	$\ell\tau_{\ell'}$	*non-VBF*	SR		
		VBF	SR	✓	

The likelihood function is fitted to the data to test for the presence of a signal. Estimates of the POIs are calculated with the profile-likelihood-ratio test statistic $\tilde{q}_\mu$ [111], and if no signal is found, the upper limits on the branching ratios are derived by using $\tilde{q}_\mu$ and the CL_s method [112].

Three different statistical analyses are performed, differing in the POI definitions and relying on different inputs from the two background estimation methods:

- An independent search for the $H \to e\tau$ signal: a single POI, $\mu_{e\tau}$, is estimated in the fit combining the $e\tau_\mu$ and $e\tau_{\text{had}}$ final states (discussed in Section 8.1). $\mathcal{B}(H \to \mu\tau)$ is set to zero.
- An independent search for the $H \to \mu\tau$ signal: a single POI, $\mu_{\mu\tau}$, is estimated in the fit combining the $\mu\tau_e$ and $\mu\tau_{\text{had}}$ final states (discussed in Section 8.1). $\mathcal{B}(H \to e\tau)$ is set to zero.
- A simultaneous measurement of the $H \to e\tau$ and $H \to \mu\tau$ signals: two POIs ($\mu_{e\tau}$ and $\mu_{\mu\tau}$) are estimated in the simultaneous fit of the $e\tau_\mu$, $e\tau_{\text{had}}$, $\mu\tau_e$ and $\mu\tau_{\text{had}}$ final states (discussed in Section 8.2).

The independent searches combine the *non-VBF* and *VBF* categories of the $\ell\tau_{\text{had}}$ channel and the *non-VBF* category of the $\ell\tau_{\ell'}$ channel from the MC-template method with the *VBF* category of the $\ell\tau_{\ell'}$ channel

from the Symmetry method. For the simultaneous measurements of $H \to e\tau$ and $H \to \mu\tau$ signals, the combined fit is performed with the regions from the MC-template method for both $\ell\tau_{\ell'}$ and $\ell\tau_{\text{had}}$ final states. The SRs and CRs exploited in the different combined fits are specified in Table 4. In the fit combinations where a SR of the $\ell\tau_{\ell'}$ channel from the MC-template method is considered, the corresponding top-quark and $Z \to \tau\tau$ CRs are also included.

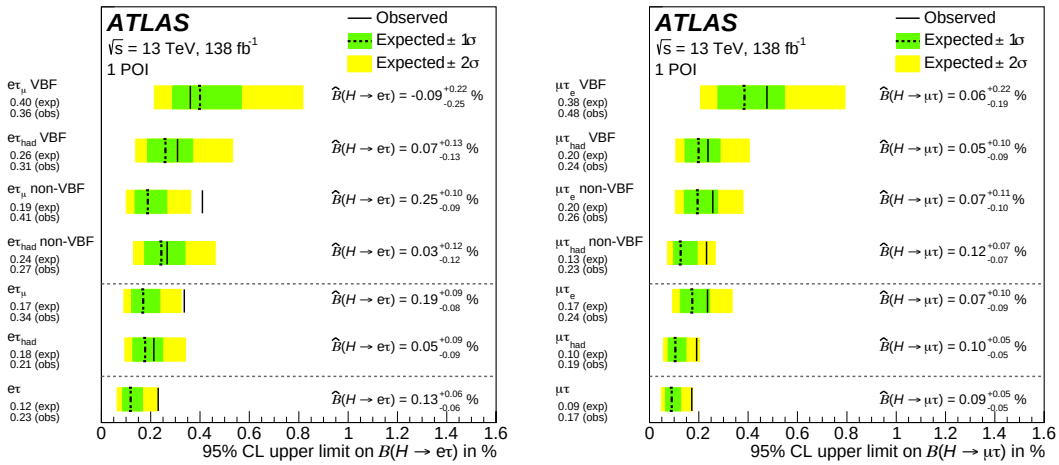
Additionally, stand-alone fits of individual channels and categories are performed employing either background estimation method. Since the Symmetry method directly measures the branching ratio difference, the results of the Symmetry method can be converted into individual branching ratios with the assumption that the other signal is zero (employed in Section 8.1). Alternatively, the results from the simultaneous fits with the MC-template method can be utilised to calculate the branching ratio difference to allow a direct comparison with the Symmetry method results. The branching ratio difference determined from the stand-alone fits with either background estimation method for the $\ell\tau_{\ell'}$ final state, together with their compatibility, is discussed in Section 8.3.

8.1 Independent searches for $H \to e\tau$ and $H \to \mu\tau$

A fit with a single POI is performed for each search separately by combining the $\ell\tau_{\ell'}$ and $\ell\tau_{\text{had}}$ final states. The best-fit branching ratios and upper limits are evaluated assuming $\mathcal{B}(H \to e\tau) = 0$ for the $H \to \mu\tau$ search and $\mathcal{B}(H \to \mu\tau) = 0$ for the $H \to e\tau$ search. Each search combines the $\ell\tau_{\text{had}}$ channel and the *non-VBF* category of the $\ell\tau_{\ell'}$ channel from the MC-template method with the *VBF* category of the $\ell\tau_{\ell'}$ channel from the Symmetry method. The choice of combining the MC-template and Symmetry methods is based on the expected sensitivity [111] of each individual category and selecting the method having higher significance. For the *VBF* category, the Symmetry method performs better than the MC-template method by 13% and 7% for the $e\tau$ and $\mu\tau$ final states, respectively. For the *non-VBF* category, the MC-template method performs better than the Symmetry method by 28% and 18% for the $e\tau$ and $\mu\tau$ final states, respectively. For the $\ell\tau_{\text{had}}$ channel, only the MC-template method is available.

Three normalisation factors are used to normalise the $Z \to \tau\tau$ production cross-section in the $\ell\tau_{\ell'}$ and $\ell\tau_{\text{had}}$ SRs estimated with the MC-template method. A common normalisation factor is used for the *non-VBF* category SR of the $\ell\tau_{\ell'}$ channel and the event yield of the dedicated $Z \to \tau\tau$ CR. In the $\ell\tau_{\text{had}}$ channel, two independent normalisation factors are used for the *non-VBF* and *VBF* categories. They are constrained by the data in the SRs. In general, top-quark processes form a very small background contribution in the $\ell\tau_{\text{had}}$ channel and no top-quark CR is defined in the $\ell\tau_{\text{had}}$ channel. Therefore, in the *non-VBF* category, where the MC-template method is used for both final states, a single normalisation factor is used to scale the top-quark processes' production cross-section jointly in the $\ell\tau_{\ell'}$ and $\ell\tau_{\text{had}}$ final state, making use of the $\ell\tau_{\ell'}$ top-quark CR. Since the Symmetry method is exploited in the $\ell\tau_{\ell'}$ channel in the *VBF* category, it is not possible to constrain the normalisation of the top-quark background in the $\ell\tau_{\text{had}}$ final state by a CR and hence uncertainties are applied to the top-quark cross-section.

Figure 11 shows the resulting 95% CL upper limits for the $H \to e\tau$ and $H \to \mu\tau$ searches, with the breakdown of the stand-alone fits from each channel. For the $H \to \mu\tau$ ($H \to e\tau$) signal, a 1.9σ (2.2σ) excess is observed. In the case of the $H \to e\tau$ signal, this is driven mainly by the *non-VBF* category of the $\ell\tau_{\ell'}$ final state based on the MC-template method. In the case of the $H \to \mu\tau$ signal, it is driven by the *non-VBF* category of the $\ell\tau_{\text{had}}$ final state.



(a) (b)

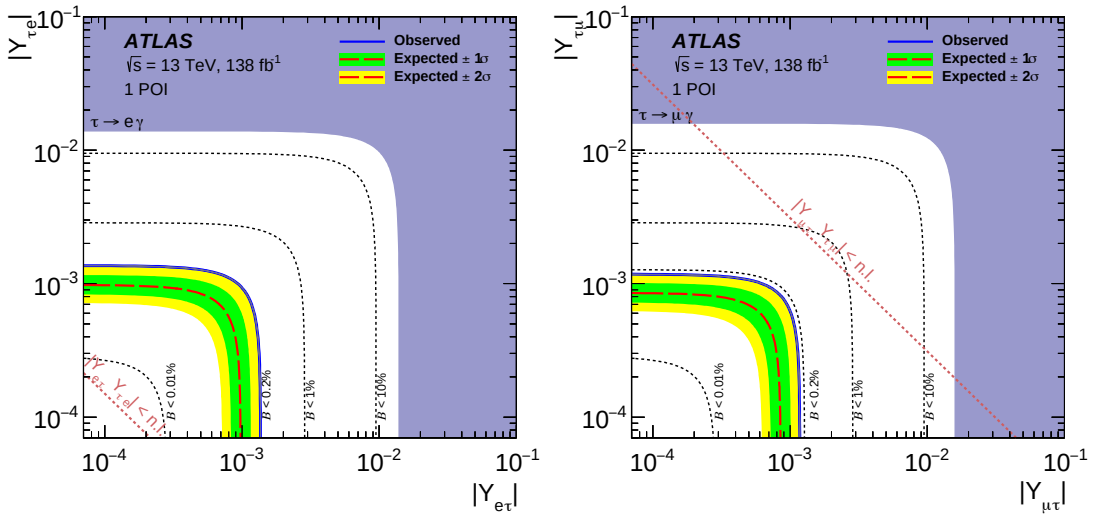
Figure 11: Fit results of the independent searches (1 POI) showing 95% CL upper limits on the LFV branching ratios of the Higgs boson, (a) $H \to e\tau$ and (b) $H \to \mu\tau$, indicated by solid lines (observed results) or dashed lines (expected results). Best-fit values of the branching ratios ($\hat{\mathcal{B}}$) are also provided, in %. The limits are computed while assuming that either (a) $\mathcal{B}(H \to \mu\tau) = 0$ or (b) $\mathcal{B}(H \to e\tau) = 0$. The channels and categories included in each likelihood fit are shown on the y-axis, and the signal and control regions from all other channels/categories are removed from the fit. The results from stand-alone channels/categories fits shown at the top are compared with the results of the combined fit displayed in the last row.

The branching ratio of the LFV Higgs boson decay is related to the non-diagonal Yukawa coupling matrix elements [32] by the formula

$$|Y_{\ell\tau}|^2 + |Y_{\tau\ell}|^2 = \frac{8\pi}{m_H} \frac{\mathcal{B}(H \to \ell\tau)}{1 - \mathcal{B}(H \to \ell\tau)} \Gamma_H(\mathrm{SM}),$$

where $\Gamma_H(\mathrm{SM}) = 4.07$ MeV [113] is the Higgs boson's width as predicted by the SM. The observed 95% CL upper limits on the branching ratio correspond to the following limits on the coupling matrix elements: $\sqrt{|Y_{\tau e}|^2 + |Y_{e\tau}|^2} < 0.0014$ and $\sqrt{|Y_{\tau\mu}|^2 + |Y_{\mu\tau}|^2} < 0.0012$.

Figure 12 shows the limits on the individual coupling matrix elements $Y_{\tau\ell}$ and $Y_{\ell\tau}$ obtained from the independent fits in the two searches. The same figure also shows the limits from the $\tau \to \ell\gamma$ searches [32, 114]. Compared with the indirect limits from the $\tau \to \ell\gamma$ searches, the direct limits obtained in this search are around (slightly less than) one order of magnitude tighter in $Y_{\tau\mu}$ and $Y_{\mu\tau}$ ($Y_{\tau e}$ and $Y_{e\tau}$), clearly indicating the strength of direct searches at the LHC. In the case of $H \to \mu\tau$, the constraints are tighter than the naturalness limit preventing a non-hierarchical mass spectrum from large off-diagonal terms in the Yukawa coupling matrix: $|Y_{\tau\ell}Y_{\ell\tau}| \lesssim m_\tau m_\ell / v^2$, where v is the vacuum expectation value of the Higgs field [32]; in the case of $H \to e\tau$, the naturalness limit has not been reached yet. This is in line with the previous ATLAS search [28] based on a partial dataset at $\sqrt{s} = 13$ TeV, but with tighter limits.



(a) (b)

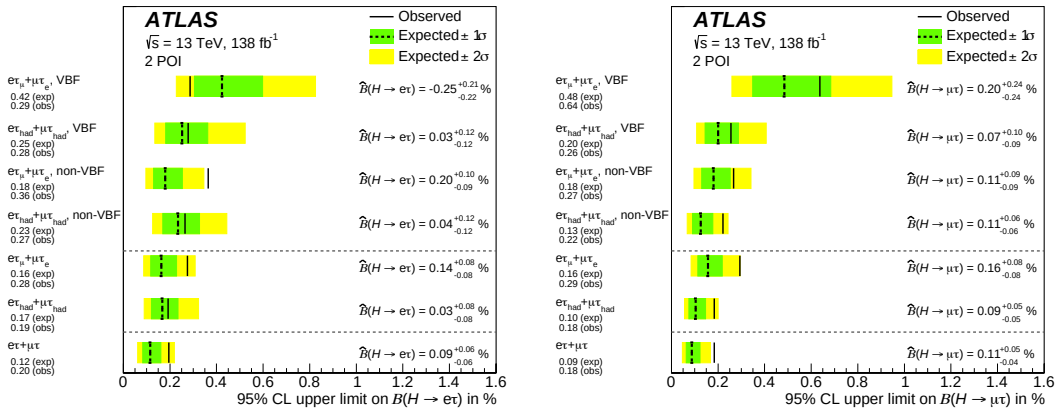
Figure 12: Expected (red long-dashed line) and observed (solid blue line) 95% CL upper limits from the independent fits (1 POI) on the absolute value of the couplings $Y_{\tau\ell}$ and $Y_{\ell\tau}$ together with the most stringent indirect limits from $\tau \to \ell\gamma$ searches (dark purple region) for (a) $\ell = e$ or (b) $\ell = \mu$. The short-dashed lines represent the limits corresponding to different branching ratios (0.01%, 0.2%, 1% and 10%), while the dotted line indicates the naturalness limit (denoted by n.l.).

8.2 Simultaneous measurement of $H \to e\tau$ and $H \to \mu\tau$ signal

A simultaneous measurement of the $H \to e\tau$ and $H \to \mu\tau$ signals is performed using the MC-template method, where it is possible to remove the assumption about the absence of a $H \to \mu\tau$ signal in the fit of $\mathcal{B}(H \to e\tau)$ and vice versa. The two POIs, corresponding to $\mathcal{B}(H \to e\tau)$ and $\mathcal{B}(H \to \mu\tau)$, are estimated in the simultaneous fit of the $e\tau_\mu$, $e\tau_{\text{had}}$, $\mu\tau_e$ and $\mu\tau_{\text{had}}$ final states. The analysis exploits eight SRs and eight CRs (one top-quark and one $Z \to \tau\tau$ CR for each $\ell\tau_{\ell'}$ SR). The $Z \to \tau\tau$ and top-quark normalisation factors are correlated between the SRs and CRs, but not between the *VBF* and *non-VBF* categories. The top-quark normalisation factors are common to the $e\tau_\mu$, $\mu\tau_e$, $e\tau_{\text{had}}$ and $\mu\tau_{\text{had}}$ channels, while $Z \to \tau\tau$ normalisation factors are decorrelated between $\ell\tau_{\ell'}$ and $\ell\tau_{\text{had}}$ final states. In total, two normalisation factors are used for the top-quark processes and four normalisation factors are used for the $Z \to \tau\tau$ processes.

Figure 13 shows the resulting 95% CL upper limits for the $H \to e\tau$ and $H \to \mu\tau$ signals, together with the contribution from each category. For the $H \to \mu\tau$ ($H \to e\tau$) signal, a 2.4σ (1.6σ) excess is observed. In the case of the $H \to e\tau$ signal, this is driven mainly by the *non-VBF* category of the $\ell\tau_{\ell'}$ final state based on the MC-template method. In the case of the $H \to \mu\tau$ signal, it is driven by the *non-VBF* category of the $\ell\tau_{\text{had}}$ final state. The result of the simultaneous fit of the $H \to e\tau$ and $H \to \mu\tau$ signals, along with the 68% and 95% CL contours, is shown in Figure 14. The result is found to be compatible with the SM prediction within 2.1σ.

Figure 15 shows the limits on the individual coupling matrix elements $Y_{\tau\ell}$ and $Y_{\ell\tau}$ obtained from the simultaneous fit of the two signals. The observed 95% CL upper limits on the branching ratios correspond to the following limits on the coupling matrix elements: $\sqrt{|Y_{\tau e}|^2 + |Y_{e\tau}|^2} < 0.0013$, and



(a) (b)

Figure 13: Fit results of the simultaneous measurement of the $H \to e\tau$ and $H \to \mu\tau$ signals (2 POI) showing 95% CL upper limits on the LFV branching ratios of the Higgs boson, (a) $H \to e\tau$ and (b) $H \to \mu\tau$, indicated by solid lines (observed results) or dashed lines (expected results). Best-fit values of the branching ratios ($\hat{\mathcal{B}}$) are also provided, in %. The channels and categories included in each likelihood fit are shown on the y-axis, and the signal and control regions from all other channels/categories are removed from the fit. The results from stand-alone channels/categories fits shown at the top are compared with the results of the combined fit displayed in the last row.

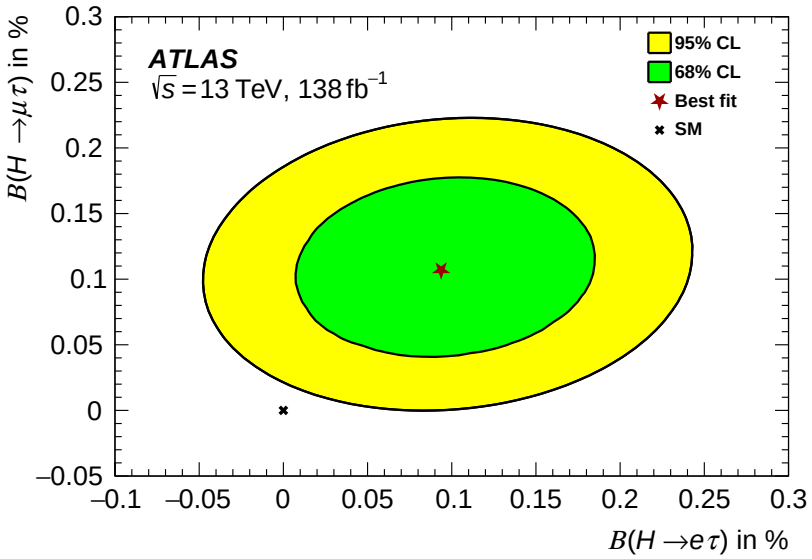
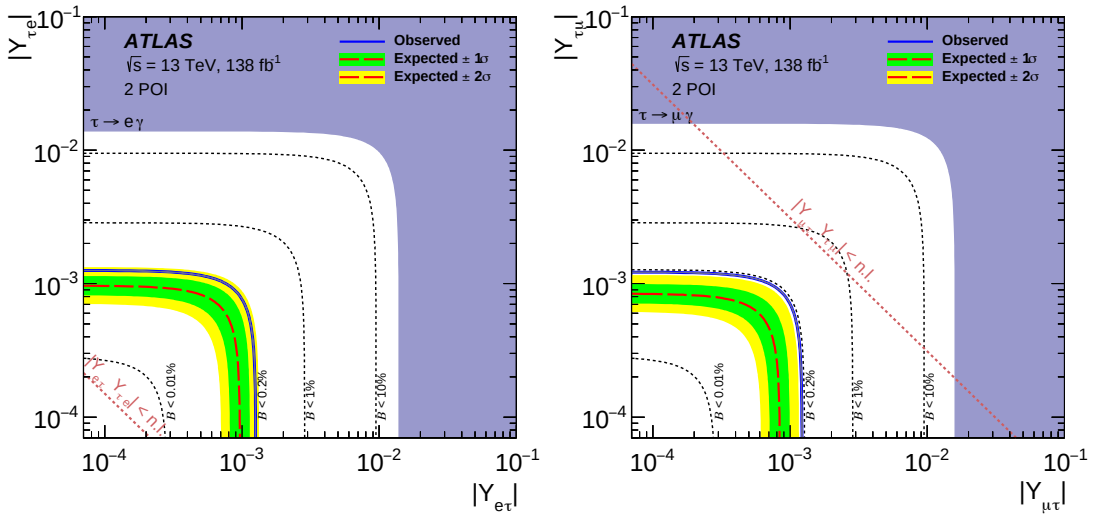


Figure 14: Best-fit value (red star) of the branching ratios $\mathcal{B}(H \to e\tau)$ and $\mathcal{B}(H \to \mu\tau)$, given in %, and likelihood contours at 68% and 95% CL obtained from the simultaneous fit of $H \to e\tau$ and $H \to \mu\tau$ signals based on the MC-template method, compared with the SM expectation (black cross).



(a) (b)

Figure 15: Expected (red long-dashed line) and observed (solid blue line) 95% CL upper limits from the simultaneous fit (2 POI) of the two searches on the absolute value of the couplings $Y_{\tau\ell}$ and $Y_{\ell\tau}$ together with the most stringent indirect limits from $\tau \to \ell\gamma$ searches (dark purple region) for (a) $\ell = e$ or (b) $\ell = \mu$. The short-dashed lines represent limits corresponding to different branching ratios (0.01%, 0.2%, 1% and 10%), while the dotted line indicates the naturalness limit (denoted by n.l.).

$\sqrt{|Y_{\tau\mu}|^2 + |Y_{\mu\tau}|^2} < 0.0012$. Both the indirect limits from $\tau \to \ell\gamma$ searches [32, 114] and the naturalness limit $|Y_{\tau\ell}Y_{\ell\tau}| \lesssim m_\tau m_\ell / v^2$ [32] are shown as well.

8.3 Measurement of the branching ratio difference in the $\ell\tau_{\ell'}$ final state

The Symmetry method uses the $e\tau_\mu$ dataset to estimate the background in the $H \to \mu\tau$ search, and the $\mu\tau_e$ dataset to estimate the background in the $H \to e\tau$ search. Therefore, this method measures the difference between the two branching ratios, $\mathcal{B}(H \to e\tau)$ and $\mathcal{B}(H \to \mu\tau)$. In the search for $H \to \mu\tau$, the difference $\mathcal{B}(H \to \mu\tau) - \mathcal{B}(H \to e\tau)$ is measured and in the search for $H \to e\tau$, the difference $\mathcal{B}(H \to e\tau) - \mathcal{B}(H \to \mu\tau)$ is measured. The two measurements of the branching ratio difference with the Symmetry method are found to be compatible, taking into account the anti-correlation between the two measurements. A more detailed description can be found in Appendix D. The $\mathcal{B}(H \to \mu\tau) - \mathcal{B}(H \to e\tau)$ measurement has the smaller expected uncertainty and is chosen as main result for the branching ratio difference from the Symmetry method. For the combination of the *non-VBF* and *VBF* categories, $\mathcal{B}(H \to \mu\tau) - \mathcal{B}(H \to e\tau)$ measured by the Symmetry method is (0.25 ± 0.10)%, compatible with zero within 2.5σ.

The branching ratio difference for the MC-template method is determined from the simultaneous fit of $\mathcal{B}(H \to e\tau)$ and $\mathcal{B}(H \to \mu\tau)$ following the approach discussed in Section 8.2, but based only on data in the $\ell\tau_{\ell'}$ final state. The obtained difference $\mathcal{B}(H \to \mu\tau) - \mathcal{B}(H \to e\tau)$ takes into account the correlation between the two branching ratio parameters from the fit. The difference $\mathcal{B}(H \to \mu\tau) - \mathcal{B}(H \to e\tau)$ is measured to be (0.02 ± 0.12)% for the combination of the *non-VBF* and *VBF* categories, compatible with zero well within 1σ.

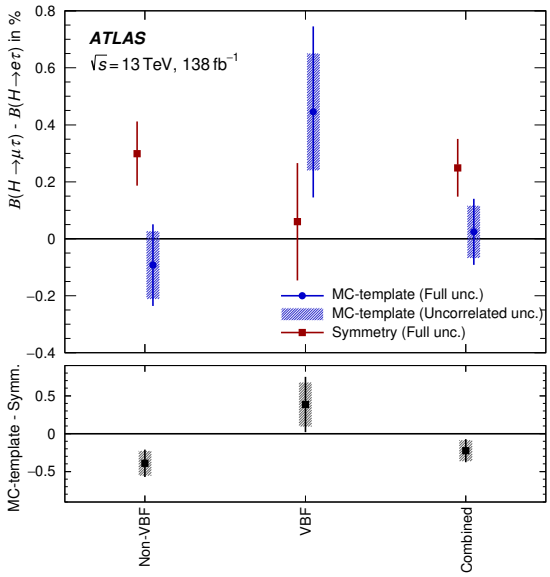


Figure 16: Best-fit $\mathcal{B}(H \to \mu\tau) - \mathcal{B}(H \to e\tau)$ values, given in %, obtained from the stand-alone fits of the $\ell\tau_{\ell'}$ final state with either background estimation method. The uncertainty of the difference displayed in the lower panel is calculated taking into account either only the sources of uncertainty that are uncorrelated between the two analyses (grey bars) or all uncertainties (black lines).

The results of the branching ratio difference measurement are shown in Figure 16, displaying the results for *non-VBF* and *VBF* categories separately as well as combined. The Symmetry method favours a larger branching ratio for the $H \to \mu\tau$ signal than for the $H \to e\tau$ signal, with a significance of 2.5σ for the combination of the *non-VBF* and *VBF* categories, driven mainly by the *non-VBF* category. The MC-template method observes a branching ratio difference compatible with zero for the combination of the *non-VBF* and *VBF* categories, driven also by the *non-VBF* category.

The measurement of the branching ratio difference with two independent background estimation methods in the $\ell\tau_{\ell'}$ final state allows the results of the two methods to be checked for compatibility. The compatibility of the two background estimation methods is tested, taking into account the correlations between the Symmetry and MC-template methods. Due to the large overlap between the data that are used by the two methods, the data statistical uncertainties are correlated. Likewise, the two methods use the same simulated samples for signal, so the signal uncertainties are correlated. All other uncertainties are considered to be uncorrelated. For the MC-template method, the correlated uncertainties are fixed to their best-fit values and the uncertainties are re-evaluated considering only the uncorrelated sources. For the Symmetry method, the full uncertainty is considered. The results obtained with the MC-template and Symmetry methods are shown in Figure 16 and are found to be compatible within 2σ.

9 Conclusion

Two direct searches for lepton-flavour-violating Higgs boson decays, $H \to e\tau$ and $H \to \mu\tau$, are presented. They are based on $\sqrt{s}$ = 13 TeV proton–proton collision data collected by the ATLAS experiment at the LHC, corresponding to an integrated luminosity of 138 fb^{-1}. Two complementary background estimation techniques are exploited. The first relies on simulation and is used in both the $\ell\tau_{\ell'}$ and $\ell\tau_{\text{had}}$ channels. The second one relies on the Symmetry method and is applied in the $\ell\tau_{\ell'}$ channel.

A simultaneous fit of possible $H \to e\tau$ and $H \to \mu\tau$ signals was performed combining the $\ell\tau_{\ell'}$ and $\ell\tau_{\text{had}}$ channels, relying on inputs from the MC-template method only. In line with the independent fits, small excesses with respect to the SM background are observed, but below the threshold for evidence of a new signal. The observed (expected) 95% CL upper limits on the branching ratios of $H \to e\tau$ and $H \to \mu\tau$ are 0.20% (0.12%) and 0.18% (0.09%), respectively. The result of the simultaneous fit of $H \to e\tau$ and $H \to \mu\tau$ signals is found to be compatible with a branching ratio value of zero for both processes within 2.1σ.

The measurement of the branching ratio difference with the Symmetry method, based only on data in the $\ell\tau_{\ell'}$ final state, favours a larger branching ratio for the $H \to \mu\tau$ signal than for the $H \to e\tau$ signal, though the difference is not statistically significant. The best-fit value for the difference $\mathcal{B}(H \to \mu\tau) - \mathcal{B}(H \to e\tau)$ obtained with the Symmetry method is (0.25 ± 0.10)%, which is compatible with a value of zero within 2.5σ. It is also compatible within 2σ with the best-fit value of (0.02 ± 0.12)%, obtained with the MC-template method in the $\ell\tau_{\ell'}$ final state.

The observed limits based on the full dataset recorded by ATLAS at $\sqrt{s}$ = 13 TeV are more stringent by factors of up to 2.5 (1.6) than the corresponding previous limits for the $H \to e\tau$ ($H \to \mu\tau$) decay based on a partial dataset at $\sqrt{s}$ = 13 TeV, while the expected sensitivity for the $H \to e\tau$ ($H \to \mu\tau$) signal improves by a factor of about 3.1 (4.1). In addition to the approximately four times larger dataset, the main sources of improvement are significantly more sophisticated features of the analysis method. These include lepton assignment in the approximate Higgs boson rest frame for the $\ell\tau_{\ell'}$ channel, use of the symmetry method for the $\ell\tau_{\ell'}$ *VBF* category in the fit with one parameter of interest, introduction of a simultaneous fit of $H \to e\tau$ and $H \to \mu\tau$ signals (two parameters of interest), more advanced multivariate classifiers for signal extraction, and improved object reconstruction, in particular the τ identification algorithm.

Appendix

A Control and validation regions summary

Table 5 summarizes the control and validation regions (CR and VR respectively) used in the analysis to characterize the various background sources.

Table 5: Control and validation regions selection in the $\ell\tau_{\ell'}$ and $\ell\tau_{\text{had}}$ channels. The detailed description of the misidentified background CR can be found in Sections 5.1 and 5.2.

Selection	$\ell\tau_{\ell'}$	$\ell\tau_{\text{had}}$
misidentified background CR	*non-VBF* (or *VBF*) category with statistically independent lepton (ℓ or $\tau_{\text{had-vis}}$) selection, see text	
$Z \to \mu\mu$ CR/VR ($\ell\tau_{\ell'}/\ell\tau_{\text{had}}$)	*Baseline* with $35\,\text{GeV} < p_{\text{T}}^{\ell_1} < 45\,\text{GeV}$ $75\,\text{GeV} < m_{\ell_1\ell_2} < 100\,\text{GeV}$ $\lvert\Delta\phi(\ell_2, E_{\text{T}}^{\text{miss}})\rvert < 1.5$ $1.25 < p_{\text{T}}^{\text{track}}(\ell_2)/p_{\text{T}}^{\text{cluster}}(\ell_2) < 3$	*Baseline* $\lvert\eta(\tau)\rvert < 0.1$ $90\,\text{GeV} < m_{\text{coll}}(\mu,\tau) < 110\,\text{GeV}$
top-quark CR	*non-VBF* (or *VBF*) selection with inverted b-veto requirement	–
$Z \to \tau\tau$ CR	*non-VBF* (or *VBF*) selection with $35\,\text{GeV} < p_{\text{T}}^{\ell_1} < 45\,\text{GeV}$	–
Diboson VR	*Baseline* $p_{\text{T}}^{\ell_2} > 30\,\text{GeV}$ $100\,\text{GeV} < m_{\ell_1\ell_2} < 150\,\text{GeV}$ $m_{\text{T}} > 30\,\text{GeV}$ veto events with jets with $p_{\text{T}} > 30$ GeV	–

B Background and signal yields

B.1 MC-template$\ell\tau_{\ell'}$ channel

The signal and background event yields in all SRs and CRs obtained after performing the simultaneous fit of the $H \to e\tau$ and $H \to \mu\tau$ signals (Section 8.2) using only data from the $\ell\tau_{\ell'}$ final state are detailed in Tables 6 and 7.

Table 6: Observed event yields and predictions as computed by the fit in the signal region (SR) and control regions (CRs) of the $e\tau_\mu$ channel. The prediction for each sample is determined from the likelihood fit performed to measure the $H \to e\tau$ and $H \to \mu\tau$ signals using the $\ell\tau_{\ell'}$ data only (Section 8.2). The signal event yields are given for $\mathcal{B}(H \to e\tau) = 0.13\%$ and $\mathcal{B}(H \to \mu\tau) = 0.17\%$, which correspond to the best-fit values obtained from the fit mentioned above. Uncertainties include statistical and systematic components. Uncertainties in the total background prediction include the effect of correlations between individual uncertainty sources as determined by the fit.

	Channel	$e\tau_\mu$					
	Region	SR		$Z \to \tau\tau$ CR		Top-quark CR	
	Category	*non-VBF*	*VBF*	*non-VBF*	*VBF*	*non-VBF*	*VBF*
Signal, $H \to e\tau$	ggF	270 ± 160	4.3 ± 2.6	37 ± 22	0.55 ± 0.33	14 ± 8	0.57 ± 0.34
	VBF	15 ± 9	8 ± 5	2.0 ± 1.2	0.8 ± 0.5	1.6 ± 1.0	0.8 ± 0.5
	VH	12 ± 7	0.22 ± 0.13	1.4 ± 0.9	0.027 ± 0.023	2.9 ± 1.7	0.053 ± 0.033
Signal, $H \to \mu\tau$	ggF	19 ± 9	0.57 ± 0.30	3.6 ± 1.7	0.045 ± 0.023	1.4 ± 0.7	0.07 ± 0.05
	VBF	1.7 ± 0.8	0.7 ± 0.4	0.18 ± 0.09	0.045 ± 0.024	0.17 ± 0.09	0.09 ± 0.05
	VH	2.4 ± 1.2	0.06 ± 0.04	0.32 ± 0.16	0.004 ± 0.002	0.51 ± 0.26	0.011 ± 0.013
Background	$Z \to \tau\tau$	23000 ± 400	450 ± 80	57500 ± 900	250 ± 40	2620 ± 180	133 ± 15
	$Z \to \ell\ell$	510 ± 150	9 ± 4	340 ± 120	5 ± 5	53 ± 19	4.3 ± 2.5
	Top-quark	34600 ± 1300	1850 ± 70	6270 ± 220	319 ± 22	298800 ± 1700	15470 ± 120
	SM ggF	1200 ± 100	37 ± 7	1090 ± 60	12.1 ± 2.3	92 ± 13	5.6 ± 2.1
	SM VBF	109 ± 4	74.9 ± 3.2	46.6 ± 2.4	19.5 ± 1.1	13.1 ± 0.8	8.0 ± 0.6
	SM VH	34 ± 5	0.6 ± 0.4	12.0 ± 1.9	0.19 ± 0.12	7.6 ± 1.2	0.19 ± 0.07
	SM ttH	8.5 ± 1.4	0.45 ± 0.08	1.28 ± 0.21	0.080 ± 0.018	214 ± 34	8.8 ± 1.5
	VV	31800 ± 1500	640 ± 70	11800 ± 600	100 ± 12	2020 ± 180	106 ± 12
	Misidentified	9200 ± 900	370 ± 50	6000 ± 1000	98 ± 25	9200 ± 1600	590 ± 100
Total Background		100800 ± 400	3450 ± 60	82700 ± 400	810 ± 31	313100 ± 600	16340 ± 120
Data		100769	3383	82902	878	312902	16382

Table 7: Observed event yields and predictions as computed by the fit in the signal region (SR) and control regions (CRs) of the $\mu\tau_e$ channel. The prediction for each sample is determined from the likelihood fit performed to measure the $H \to e\tau$ and $H \to \mu\tau$ signals using the $\ell\tau_{\ell'}$ data only (Section 8.2). The signal event yields are given for $\mathcal{B}(H \to e\tau) = 0.13\%$ and $\mathcal{B}(H \to \mu\tau) = 0.17\%$, which correspond to the best-fit values obtained from the fit mentioned above. Uncertainties include statistical and systematic components. Uncertainties in the total background prediction include the effect of correlations between individual uncertainty sources as determined by the fit.

	Channel	$\mu\tau_e$					
	Region	SR		$Z \to \tau\tau$ CR		Top-quark CR	
	Category	*non-VBF*	*VBF*	*non-VBF*	*VBF*	*non-VBF*	*VBF*
Signal, $H \to e\tau$	ggF	15 ± 9	0.44 ± 0.27	2.7 ± 1.6	0.041 ± 0.026	1.1 ± 0.6	0.08 ± 0.06
	VBF	1.2 ± 0.8	0.6 ± 0.4	0.14 ± 0.09	0.042 ± 0.026	0.15 ± 0.09	0.07 ± 0.04
	VH	1.9 ± 1.1	0.06 ± 0.04	0.26 ± 0.15	0.005 ± 0.004	0.41 ± 0.25	0.012 ± 0.008
Signal, $H \to \mu\tau$	ggF	270 ± 130	4.5 ± 2.3	38 ± 18	0.55 ± 0.28	14 ± 7	0.64 ± 0.33
	VBF	15 ± 7	8 ± 4	2.1 ± 1.1	0.9 ± 0.5	1.6 ± 0.8	0.8 ± 0.4
	VH	12 ± 6	0.2 ± 0.1	1.5 ± 0.8	0.022 ± 0.011	3.0 ± 1.5	0.059 ± 0.031
Background	$Z \to \tau\tau$	18200 ± 400	360 ± 70	46400 ± 600	213 ± 32	2000 ± 160	113 ± 13
	$Z \to \ell\ell$	610 ± 190	5.1 ± 2.9	640 ± 180	2.1 ± 1.4	48 ± 22	1.5 ± 1.1
	Top-quark	30900 ± 1200	1750 ± 60	5770 ± 200	306 ± 22	274500 ± 900	14500 ± 100
	SM ggF	942 ± 70	32 ± 6	910 ± 50	10.5 ± 1.9	79 ± 11	4.4 ± 1.6
	SM VBF	92 ± 4	64.7 ± 2.8	40.9 ± 2.1	17.4 ± 1.0	11.1 ± 0.7	7.0 ± 0.5
	SM VH	27 ± 4	0.6 ± 0.4	9.7 ± 1.4	0.17 ± 0.09	6.7 ± 1.3	0.07 ± 0.10
	SM ttH	7.7 ± 1.3	0.43 ± 0.08	1.25 ± 0.21	0.061 ± 0.023	200 ± 110	8.0 ± 1.3
	VV	28000 ± 1300	610 ± 60	10600 ± 500	100 ± 12	1820 ± 160	100 ± 10
	Misidentified	11400 ± 700	390 ± 40	8200 ± 600	81 ± 18	8900 ± 600	460 ± 50
Total Background		90500 ± 400	3230 ± 60	72680 ± 330	733 ± 28	287600 ± 500	15210 ± 110
Data		90531	3138	72511	712	287734	15228

B.2 MC-template $\ell\tau_{\text{had}}$ channel

The signal and background event yields in all SRs obtained after performing the simultaneous fit of the $H \to e\tau$ and $H \to \mu\tau$ signals (Section 8.2) using only data of the $\ell\tau_{\text{had}}$ final state are detailed in Table 8.

Table 8: Observed event yields and predictions as computed by the fit in the signal regions (SRs) of the $e\tau_{\text{had}}$ and $\mu\tau_{\text{had}}$ channels. The prediction for each sample is determined from the likelihood fit performed to measure the $H \to e\tau$ and $H \to \mu\tau$ signals using the $\ell\tau_{\text{had}}$ data only (Section 8.2). The signal event yields are given for $\mathcal{B}(H \to e\tau) = 0.03\%$ and $\mathcal{B}(H \to \mu\tau) = 0.09\%$, which correspond to the best-fit values obtained from the fit mentioned above. Uncertainties include statistical and systematic components. Uncertainties in the total background prediction include the effect of correlations between individual uncertainty sources as determined by the fit.

	Channel	$e\tau_{\text{had}}$		$\mu\tau_{\text{had}}$	
	Region	SR		SR	
	Category	*non-VBF*	*VBF*	*non-VBF*	*VBF*
Signal, $H \to e\tau$	ggF	110 ± 270	2 ± 5	<0.001	<0.001
	VBF	5 ± 14	4 ± 10	<0.001	<0.001
	VH	4 ± 9	0.06 ± 0.16	0.07 ± 0.17	0.003 ± 0.006
Signal, $H \to \mu\tau$	ggF	0.019 ± 0.013	0.0027 ± 0.0029	440 ± 250	5.7 ± 3.4
	VBF	0.0014 ± 0.0010	<0.001	22 ± 13	12 ± 7
	VH	0.08 ± 0.05	0.0024 ± 0.0019	14 ± 8	0.21 ± 0.13
Background	$Z \to \tau\tau$	286800 ± 1800	2700 ± 100	389100 ± 2600	3440 ± 130
	$Z \to \ell\ell$	32100 ± 2000	280 ± 80	33200 ± 2900	180 ± 40
	Top-quark	7400 ± 2000	520 ± 150	9500 ± 2500	590 ± 160
	SM ggF	2110 ± 130	49 ± 10	3220 ± 190	50 ± 10
	SM VBF	135 ± 13	103 ± 7	196 ± 16	121 ± 8
	SM VH	100 ± 13	2.1 ± 0.6	131 ± 17	2.5 ± 0.5
	SM ttH	4.3 ± 0.6	0.28 ± 0.05	4.9 ± 0.7	0.32 ± 0.06
	VV, W	9000 ± 500	250 ± 40	13100 ± 700	280 ± 30
	Misidentified	140200 ± 2100	1810 ± 130	140800 ± 1900	1520 ± 90
Total Background		478000 ± 1200	5700 ± 110	589700 ± 1200	6200 ± 100
Data		477927	5713	589198	6178

B.3 Symmetry-based $\ell\tau_{\ell'}$ channel

The signal and background event yields in the SRs obtained after performing the independent fit (Section 8.1) using only data passing the symmetry-method specific $\ell\tau_{\ell'}$ event selection are detailed in Table 9. The background contribution is divided into the symmetric, F_{MC} and F_{FF} components. Contributions from the other lepton channel are marked with the 'tilde' sign.

Table 9: Observed event yields in the *VBF* and *non-VBF* categories for the $H \to e\tau$ and $H \to \mu\tau$ searches following the Symmetry-method, after the statistical analysis based on events passing the $\ell\tau_{\ell'}$ Symmetry-method specific selection. The signal event yields are given for the best-fit branching ratio values based on the $\ell\tau_{\ell'}$ data, $\mathcal{B}(H \to e\tau) = -0.33\%$ and $\mathcal{B}(H \to \mu\tau) = 0.25\%$, where the fit of the difference is translated into individual branching ratios, by setting the other branching ratio explicitly to zero. This results in a negative shift for $\mathcal{B}(H \to e\tau)$, and a positive shift for $\mathcal{B}(H \to \mu\tau)$ (cf. Section 5.3). Contributions in the search for $H \to e\tau$ coming from the $\mu\tau_e$ selection are marked with the 'tilde' sign, those from the $e\tau_\mu$ selection are listed without. For the $H \to \mu\tau$ search, the notation is the same with interchanged regions. Uncertainties include statistical and systematic components. Uncertainties in the total background prediction include the effect of correlations between individual uncertainty sources as determined by the fit.

		$H \to e\tau$ (no $H \to \mu\tau$)		$H \to \mu\tau$ (no $H \to e\tau$)	
	Process	*non-VBF*	*VBF*	*non-VBF*	*VBF*
Signal	ggF	-720 ± 230	-11 ± 4	480 ± 190	8 ± 4
	VBF	-38 ± 12	-20 ± 7	26 ± 11	14 ± 6
	VH	-31 ± 11	-0.55 ± 0.21	21 ± 9	0.33 ± 0.16
	$\widetilde{ggF}$	48 ± 17	1.3 ± 0.5	-35 ± 15	-1.0 ± 0.5
	$\widetilde{VBF}$	3.9 ± 1.3	1.7 ± 0.6	-2.8 ± 1.2	-1.2 ± 0.5
	$\widetilde{VH}$	5.7 ± 2	0.14 ± 0.06	-4.2 ± 1.8	-0.10 ± 0.05
Symmetric background	$\widetilde{\text{Data}}$	193700 ± 1200	4300 ± 50	162900 ± 900	3890 ± 50
	$\widetilde{F_{\text{FF}}}$	-19000 ± 2000	-340 ± 40	-15100 ± 1500	-300 ± 34
	$\widetilde{F_{\text{MC}}}$	-12600 ± 1800	-112 ± 30	-5400 ± 700	-72 ± 17
Misidentified background	F_{FF}	17900 ± 1800	330 ± 40	17200 ± 1700	320 ± 40
	F_{MC}	5700 ± 800	81 ± 19	11500 ± 1600	106 ± 28
Total background		185700 ± 500	4260 ± 50	171000 ± 500	3950 ± 50
Data		184887	4230	171675	3967

C MVA optimisation

The MVAs are trained using the four-momenta of the selected reconstructed particles as well as other derived observables, such as the invariant masses, angular separations and the $E_{\mathrm{T}}^{\mathrm{miss}}$. In the *VBF* category, additional jet-related variables are used to exploit the characteristic VBF topology. The complete list of the variables used is shown in Table 10. For the MC-template method, the list of input variables is optimised by maximising the expected signal significance based on fits to Asimov datasets [111] considering statistical uncertainties only. For the Symmetry method, the hyperparameters and the input variables are optimised with respect to the expected signal significance by performing the statistical analysis taking into account all systematic uncertainties and using only $\ell\tau_{\ell'}$ data (as described in Section 8). For each category, the lowest-ranked variables making marginal contributions to the expected sensitivity are removed.

The collinear mass (m_{coll}) [86], the Higgs boson mass obtained with the Missing Mass Calculator (MMC) technique (m_{MMC}) [85] and m_{vis} show the highest separation power together with transverse mass calculated from the Higgs decay products and the $E_{\mathrm{T}}^{\mathrm{miss}}$ ($m_{\mathrm{T}}(\tau, E_{\mathrm{T}}^{\mathrm{miss}})$, $m_{\mathrm{T}}(\ell, E_{\mathrm{T}}^{\mathrm{miss}})$). The MMC algorithm is tuned specifically to reconstruct the mass of the LFV Higgs boson. In the collinear approximation, the azimuthal angle between the prompt lepton from the Higgs boson and the $E_{\mathrm{T}}^{\mathrm{miss}}$, $\Delta\phi(\ell_H, E_{\mathrm{T}}^{\mathrm{miss}})$, is expected to be large for signal events, while the azimuthal angle between the visible products of the τ-lepton decay and the $E_{\mathrm{T}}^{\mathrm{miss}}$, $\Delta\phi(\tau, E_{\mathrm{T}}^{\mathrm{miss}})$, is expected to be small for signal events. Other angular separations are included as well: $\Delta R(\ell_H, \tau)$, $\Delta\eta(\ell_H, \tau)$, $\Delta\phi(\ell_H, \tau)$, where τ refers to τ_ℓ or $\tau_{\text{had-vis}}$ in the corresponding channel. The $\Delta\phi$ angular separations involving ℓ_H are evaluated in the approximate Higgs boson rest frame.

In the $\ell\tau_{\ell'}$ channel, two vertex-related variables are also included because they provide additional discrimination power: the difference of the transverse impact parameters of the leptons ($\Delta d_0\,(\ell_1, \ell_2)$) and the d_0 significance of the lepton from the τ: $\sigma_{d_0}^{\ell_\tau}$. The $\Delta\alpha$ discriminant [115] exploited in the $\ell\tau_{\ell'}$ channels is expected to be close to zero if the decay products of τ-lepton are collinear and the transverse momentum of the Higgs boson can be neglected, while for the background events, this value deviates from zero. Additionally, in the Symmetry $\ell\tau_{\ell'}$ channel, the η-centrality, defined as

$$\eta\text{-centrality}(\ell) = \exp(-4/(\eta_{\mathrm{j}_1} - \eta_{\mathrm{j}_2})^2 \cdot (\eta_\ell - 0.5(\eta_{\mathrm{j}_1} + \eta_{\mathrm{j}_2}))^2),$$

is included in the *VBF* category.

Different MVA techniques are used for the *non-VBF* and *VBF* categories to exploit the VBF topology, taking into account for the different limited training sample sizes.

For the MC-template method, the parameters used to configure the BDT are summarised in Table 11. The BDTs are trained using the Toolkit for Multivariate Data Analysis [92]. A k-fold cross validation procedure [116] is implemented for the BDT training to avoid overtraining and to enable efficient usage of the available events in the training process. In both the $\ell\tau_{\ell'}$ and $\ell\tau_{\mathrm{had}}$ final states, 5-folds are used.

In the $\ell\tau_{\ell'}$ channel of the Symmetry method, the fully connected deep NN training with Keras [93] and Tensorflow [94] as the backend uses supervised learning with categorical cross-entropy as the cost function and L2 weight regularisation to prevent overtraining, as well as 10-fold cross validation. The hyperparameter optimisation is carried out using the *Optuna* framework [117], where the L2 weight regularisation parameter is included in the hyperparameter optimisation. The full set of hyperparameters used for the NNs in the Symmetry $\ell\tau_{\ell'}$ channel is summarised in Table 12.

Table 10: List of input variables used in the *non-VBF* and *VBF* categories. Those labelled with 'rest' are evaluated in the approximate Higgs boson rest frame as described in Section 4.2. The variable $p_{\mathrm{T}}^{\mathrm{tot}}$, used only in the *VBF* category of the MC-template $\ell\tau_{\ell'}$ channel, corresponds to the absolute value of the vectorial sum of the p_{T} of the two leptons, the two jets and the $E_{\mathrm{T}}^{\mathrm{miss}}$, and it is used to veto the third jet. When a variable is used in both the $\ell\tau_{\ell'}$ and $\ell\tau_{\mathrm{had}}$ channels, the τ symbol refers to either τ_{ℓ} or $\tau_{\text{had-vis}}$ in the corresponding channel.

Variable	$\ell\tau_{\mathrm{had}}$		$\ell\tau_{\ell'}$ MC-template		$\ell\tau_{\ell'}$ Symmetry	
	non-VBF	*VBF*	*non-VBF*	*VBF*	*non-VBF*	*VBF*
m_{coll}	✓	✓	✓	✓	✓	✓
m_{vis}	✓	✓	✓	✓	✓	✓
m_{MMC}			✓	✓	✓	✓
$m_{\mathrm{T}}(\tau, E_{\mathrm{T}}^{\mathrm{miss}})$	✓	✓			✓	✓
$m_{\mathrm{T}}(\ell_H, E_{\mathrm{T}}^{\mathrm{miss}})$	✓	✓			✓	✓
$m_{\mathrm{T}}(\ell_1, E_{\mathrm{T}}^{\mathrm{miss}})$			✓	✓		
$m_{\mathrm{T}}(\ell_2, E_{\mathrm{T}}^{\mathrm{miss}})$			✓	✓		
$E_{\mathrm{T}}^{\mathrm{miss}}$	✓	✓	✓	✓	✓	✓
$p_{\mathrm{T}}(\ell_H)$	✓	✓				
$p_{\mathrm{T}}(\tau_{\text{had-vis}})$	✓	✓				
$p_{\mathrm{T}}(\ell_2 + E_{\mathrm{T}}^{\mathrm{miss}})/p_{\mathrm{T}}(\ell_1)$			✓	✓		
$p_{\mathrm{T}}^{\mathrm{rest}}(\ell_H)$			✓	✓	✓	✓
$p_{\mathrm{T}}^{\mathrm{rest}}(\ell_\tau)$			✓	✓	✓	✓
$p_{\mathrm{T}}^{\mathrm{tot}}$				✓	✓	✓
$p_{\mathrm{T}}(\ell_H)/E_{\mathrm{T}}^{\mathrm{miss}}$					✓	✓
$p_{\mathrm{T}}(\ell_H)/p_{\mathrm{T}}(\ell_\tau)$					✓	✓
$p_{\mathrm{T}}(\ell_\tau + E_{\mathrm{T}}^{\mathrm{miss}})/p_{\mathrm{T}}(\ell_H)$					✓	✓
$\sum p_{\mathrm{T}}$					✓	✓
$\Delta R(\ell_H, \tau)$	✓	✓	✓	✓	✓	✓
$\Delta\eta(\ell_H, \tau)$	✓	✓			✓	✓
$\Delta\phi(\ell_H, \tau)$	✓				✓	
$\Delta\phi(\ell_\tau, E_{\mathrm{T}}^{\mathrm{miss}})$			✓	✓	✓	✓
$\lvert\Delta\phi(\ell_H, E_{\mathrm{T}}^{\mathrm{miss}})\rvert - \lvert\Delta\phi(\tau_{\text{had-vis}}, E_{\mathrm{T}}^{\mathrm{miss}})\rvert$	✓					
$\Delta\alpha$			✓	✓	✓	✓
$\Delta\Phi\left(\ell_H, E_{\mathrm{T}}^{\mathrm{miss}}\right)$		✓	✓	✓	✓	✓
$\Delta d_0\,(\ell_1, \ell_2)$			✓	✓	✓	
$\sigma_{d_0}^{\ell_\tau}$			✓	✓		
$\eta(\tau_{\text{had-vis}})$	✓	✓				
m_{jj}		✓		✓		✓
$N_{\mathrm{jets}(p_{\mathrm{T}}>30\mathrm{GeV})}$		✓				
$\lvert\Delta\eta_{\mathrm{jj}}\rvert$		✓				✓
$\Delta R(\mathrm{j}, \mathrm{j})$				✓		✓
$\lvert\Delta\eta_{\mathrm{jj}}\rvert \cdot \eta_{\mathrm{j}_1} \cdot \eta_{\mathrm{j}_2}$				✓		
$p_{\mathrm{T}}(\mathrm{j}_1)$						✓
$p_{\mathrm{T}}(\mathrm{j}_2)$						✓
$\Delta\phi(\mathrm{j}_1, E_{\mathrm{T}}^{\mathrm{miss}})$				✓		✓
$\Delta\phi(\mathrm{j}_2, E_{\mathrm{T}}^{\mathrm{miss}})$				✓		✓
η-centrality(ℓ_H)				✓		✓
η-centrality(ℓ_τ)				✓		✓

Table 11: Optimised BDT parameters, for all the regions and channels used in the MC-template method. The Boost type and Number-of-cuts are set to Gradient and 20 respectively.

Region	Channel	NTrees	MaxDepth	MinNodeSize	Shrinkage
non-VBF	$e\tau_{\text{had}}$,$\mu\tau_{\text{had}}$	500	7	1	0.05
VBF	$e\tau_{\text{had}}$	300	10	1	0.01
	$\mu\tau_{\text{had}}$ BDT_1	300	8	1	0.009
	$\mu\tau_{\text{had}}$ BDT_2	300	6	1	0.0095
non-VBF,*VBF*	$\ell\tau_{\ell'}$	750	8	2.5	0.1

Table 12: Hyperparameter configuration for the NNs used in the Symmetry $\ell\tau_{\ell'}$ analysis.

Hyperparameter	Value			
	non-VBF NN	$VBF_{Z\to\tau\tau}$ NN	$VBF_{\text{Top-quark}}$ NN	VBF_{misID} NN
# nodes in 1st layer	512	128	128	128
# hidden layers	2	4	3	4
# output layers	3	2	2	2
L2 weight reg. param.	0.000048	0.000292	0.000094	0.000356
Leaky ReLU slope below 0	0.080537	0.019614	0.062515	0.084219
Optimiser	SGD	Adam	Adam	Adam
Learning rate	0.025810	0.000142	0.000215	0.003507
Batch size	128	128	512	1024
Epochs	100	100	100	100

D Compatibility of branching ratio differences

The compatibility of the two measured branching ratio differences, $\Delta_{e\tau} = \mathcal{B}(H \to e\tau) - \mathcal{B}(H \to \mu\tau)$ and $\Delta_{\mu\tau} = \mathcal{B}(H \to \mu\tau) - \mathcal{B}(H \to e\tau)$, in the Symmetry method is tested. Measuring $\Delta_{e\tau}$ ($\Delta_{\mu\tau}$) is equivalent to measuring $\mu_{e\tau}$ ($\mu_{\mu\tau}$) when assuming $\mu_{\mu\tau}$ ($\mu_{e\tau}$) = 0. This implies that the sum of the two fitted values for $\mu_{e\tau}$ and $\mu_{\mu\tau}$ should be consistent with zero. This was quantified using a χ^2 test, accounting for the error of the mean value of the upward and downward systematic variations of $\mu_{e\tau}$ and $\mu_{\mu\tau}$, and the correlation coefficient ρ between the $e\tau_\mu$ and $\mu\tau_e$ datasets. The latter is found to be $\rho = -0.80$ using the bootstrap methodology [118] when considering the full systematic uncertainties in the fit. The results obtained in the *non-VBF* and *VBF* categories, as well as in the combined fit, indicate that the $\Delta_{e\tau}$ and $\Delta_{\mu\tau}$ measurements are compatible within 1.3σ.

The branching ratio difference, $\Delta_{\mu\tau}$, from the Symmetry method can be compared with the value from the MC-template method, as discussed in Section 8.3. The result is displayed in Table 13.

Table 13: Best-fit $\mathcal{B}(H \to \mu\tau) - \mathcal{B}(H \to e\tau)$ values, given in %, obtained from the stand-alone fits of the $\ell\tau_{\ell'}$ final state with either background estimation method. The results for *non-VBF* and *VBF* category are provided, as well as their combination. The last row shows the difference between the MC-template method and the Symmetry method. The results for the MC-template method are quoted once with the full uncertainty and once considering only the sources of uncertainty that are uncorrelated between the two analyses. The uncertainty in the difference between the two methods is calculated taking into account only the uncorrelated uncertainty sources for the MC-template method and the full uncertainty for Symmetry method.

	$\hat{\mathcal{B}}(H \to \mu\tau) - \hat{\mathcal{B}}(H \to e\tau)$ [%]		
Method	*non-VBF*	*VBF*	Combined
MC-template (full uncertainties)	-0.10 ± 0.14	0.45 ± 0.30	0.02 ± 0.12
MC-template (uncorrelated uncertainties)	-0.10 ± 0.12	0.45 ± 0.21	0.02 ± 0.09
Symmetry (full uncertainties)	0.30 ± 0.11	0.06 ± 0.21	0.25 ± 0.10
MC-template (uncorr. uncertainties) − Symmetry	-0.39 ± 0.16	0.39 ± 0.29	-0.22 ± 0.14

References

[1] ATLAS Collaboration, *Observation of a new particle in the search for the Standard Model Higgs boson with the ATLAS detector at the LHC*, Phys. Lett. **B 716** (2012) 1, arXiv: 1207.7214 [hep-ex].

[2] CMS Collaboration, *Observation of a new boson at a mass of 125 GeV with the CMS experiment at the LHC*, Phys. Lett. **B 716** (2012) 30, arXiv: 1207.7235 [hep-ex].

[3] F. Englert and R. Brout, *Broken Symmetry and the Mass of Gauge Vector Mesons*, Phys. Rev. Lett. **13** (1964) 321.

[4] P. W. Higgs, *Broken symmetries, massless particles and gauge fields*, Phys. Lett. **12** (1964) 132.

[5] P. W. Higgs, *Broken Symmetries and the Masses of Gauge Bosons*, Phys. Rev. Lett. **13** (1964) 508.

[6] P. W. Higgs, *Spontaneous Symmetry Breakdown without Massless Bosons*, Phys. Rev. **145** (1966) 1156.

[7] G. Guralnik, C. Hagen and T. Kibble, *Global Conservation Laws and Massless Particles*, Phys. Rev. Lett. **13** (1964) 585.

[8] T. Kibble, *Symmetry Breaking in Non-Abelian Gauge Theories*, Phys. Rev. **155** (1967) 1554.

[9] J. D. Bjorken and S. Weinberg, *Mechanism for Nonconservation of Muon Number*, Phys. Rev. Lett. **38** (1977) 622.

[10] J. L. Diaz-Cruz and J. J. Toscano, *Lepton flavor violating decays of Higgs bosons beyond the standard model*, Phys. Rev. D **62** (2000) 116005.

[11] M. Arana-Catania, E. Arganda and M. J. Herrero, *Non-decoupling SUSY in LFV Higgs decays: a window to new physics at the LHC*, JHEP **09** (2013) 160, (Erratum: JHEP **10** (2015) 192), arXiv: 1304.3371 [hep-ph].

[12] A. Arhrib, Y. Cheng and O. C. W. Kong, *Comprehensive analysis on lepton flavor violating Higgs boson to $\mu^{\mp}\tau^{\pm}$ decay in supersymmetry without R parity*, Phys. Rev. D **87** (2013) 015025, arXiv: 1210.8241 [hep-ph].

[13] K. Agashe and R. Contino, *Composite Higgs-mediated flavor-changing neutral current*, Phys. Rev. D **80** (2009) 075016, arXiv: 0906.1542 [hep-ph].

[14] A. Azatov, M. Toharia and L. Zhu, *Higgs mediated flavor changing neutral currents in warped extra dimensions*, Phys. Rev. D **80** (2009) 035016, arXiv: 0906.1990 [hep-ph].

[15] H. Ishimori et al., *Non-Abelian Discrete Symmetries in Particle Physics*, Prog. Theor. Phys. Suppl. **183** (2010) 1, arXiv: 1003.3552 [hep-th].

[16] G. Perez and L. Randall, *Natural neutrino masses and mixings from warped geometry*, JHEP **01** (2009) 077, arXiv: 0805.4652 [hep-ph].

[17] M. E. Albrecht, M. Blanke, A. J. Buras, B. Duling and K. Gemmler, *Electroweak and flavour structure of a warped extra dimension with custodial protection*, JHEP **09** (2009) 064, arXiv: 0903.2415 [hep-ph].

[18] A. Goudelis, O. Lebedev and J.-h. Park, *Higgs-induced lepton flavor violation*, Phys. Lett. B **707** (2012) 369, arXiv: 1111.1715 [hep-ph].

[19] D. McKeen, M. Pospelov and A. Ritz,
Modified Higgs branching ratios versus CP and lepton flavor violation,
Phys. Rev. D **86** (2012) 113004, arXiv: 1208.4597 [hep-ph].

[20] BaBar Collaboration, *Evidence for an Excess of* $\bar{B} \to D^{(*)}\tau^-\bar{\nu}_\tau$ *Decays*,
Phys. Rev. Lett. **109** (2012) 101802, arXiv: 1205.5442 [hep-ex].

[21] LHCb Collaboration,
Measurement of the Ratio of Branching Fractions $\mathcal{B}(\overline{B}^0 \to D^{*+}\tau^-\overline{\nu}_\tau)/\mathcal{B}(\overline{B}^0 \to D^{*+}\mu^-\overline{\nu}_\mu)$,
Phys. Rev. Lett. **115** (2015) 111803, arXiv: 1506.08614 [hep-ex].

[22] LHCb Collaboration, *Measurement of the Ratio of the* $B^0 \to D^{*-}\tau^+\nu_\tau$ *and* $B^0 \to D^{*-}\mu^+\nu_\mu$ *Branching Fractions Using Three-Prong* τ*-Lepton Decays*, Phys. Rev. Lett. **120** (2018) 171802,
arXiv: 1708.08856 [hep-ex].

[23] LHCb Collaboration,
Differential branching fraction and angular analysis of $\Lambda_b^0 \to \Lambda\mu^+\mu^-$ *decays*,
JHEP **06** (2015) 115, (Erratum: JHEP **09** (2018) 145), arXiv: 1503.07138 [hep-ex].

[24] Belle Collaboration, *Measurement of* $\mathcal{R}(D)$ *and* $\mathcal{R}(D^*)$ *with a Semileptonic Tagging Method*,
Phys. Rev. Lett. **124** (2020) 161803, arXiv: 1910.05864 [hep-ex].

[25] P. Ko, Y. Omura, Y. Shigekami and C. Yu,
LHCb anomaly and B physics in flavored Z' models with flavored Higgs doublets,
Phys. Rev. D **95** (2017) 115040, arXiv: 1702.08666 [hep-ph].

[26] A. Carmona and F. Goertz,
Lepton Flavor and Nonuniversality from Minimal Composite Higgs Setups,
Phys. Rev. Lett. **116** (2016) 251801, arXiv: 1510.07658 [hep-ph].

[27] F. J. Botella, G. C. Branco, M. Nebot and M. N. Rebelo,
Flavour-changing Higgs couplings in a class of two Higgs doublet models,
Eur. Phys. J. C **76** (2016) 161, arXiv: 1508.05101 [hep-ph].

[28] ATLAS Collaboration, *Searches for lepton-flavour-violating decays of the Higgs boson in* $\sqrt{s} = 13$ *TeV* pp *collisions with the ATLAS detector*, Phys. Lett. B **800** (2020) 135069,
arXiv: 1907.06131 [hep-ex].

[29] CMS Collaboration, *Search for lepton-flavor violating decays of the Higgs boson in the* $\mu\tau$ *and* $e\tau$ *final states in proton–proton collisions at* $\sqrt{s} = 13$ *TeV*, Phys. Rev. D **104** (2021) 032013,
arXiv: 2105.03007 [hep-ex].

[30] ATLAS Collaboration, *Search for the Higgs boson decays* $H \to ee$ *and* $H \to e\mu$ *in* pp *collisions at* $\sqrt{s} = 13$ *TeV with the ATLAS detector*, Phys. Lett. B **801** (2020) 135148,
arXiv: 1909.10235 [hep-ex].

[31] J. Adam et al., *New Constraint on the Existence of the* $\mu^+ \to e^+\gamma$ *Decay*,
Phys. Rev. Lett. **110** (2013) 201801, arXiv: 1303.0754 [hep-ex].

[32] R. Harnik, J. Kopp and J. Zupan, *Flavor violating Higgs decays*, JHEP **03** (2013) 026,
arXiv: 1209.1397 [hep-ph].

[33] G. Blankenburg, J. Ellis and G. Isidori, *Flavour-changing decays of a 125 GeV Higgs-like particle*,
Phys. Lett. B **712** (2012) 386, arXiv: 1202.5704 [hep-ph].

[34] S. Bressler, A. Dery and A. Efrati, *Asymmetric lepton-flavor violating Higgs boson decays*,
Phys. Rev. D **90** (2014) 015025, arXiv: 1405.4545 [hep-ph].

[35] ATLAS Collaboration, *The ATLAS Experiment at the CERN Large Hadron Collider*, JINST **3** (2008) S08003.

[36] ATLAS Collaboration, *ATLAS Insertable B-Layer: Technical Design Report*, ATLAS-TDR-19; CERN-LHCC-2010-013, 2010, URL: https://cds.cern.ch/record/1291633, Addendum: ATLAS-TDR-19-ADD-1; CERN-LHCC-2012-009, 2012, URL: https://cds.cern.ch/record/1451888.

[37] B. Abbott et al., *Production and integration of the ATLAS Insertable B-Layer*, JINST **13** (2018) T05008, arXiv: 1803.00844 [physics.ins-det].

[38] ATLAS Collaboration, *Performance of the ATLAS trigger system in 2015*, Eur. Phys. J. C **77** (2017) 317, arXiv: 1611.09661 [hep-ex].

[39] ATLAS Collaboration, *The ATLAS Collaboration Software and Firmware*, ATL-SOFT-PUB-2021-001, 2021, URL: https://cds.cern.ch/record/2767187.

[40] ATLAS Collaboration, *ATLAS data quality operations and performance for 2015–2018 data-taking*, JINST **15** (2020) P04003, arXiv: 1911.04632 [physics.ins-det].

[41] ATLAS Collaboration, *Performance of electron and photon triggers in ATLAS during LHC Run 2*, Eur. Phys. J. C **80** (2020) 47, arXiv: 1909.00761 [hep-ex].

[42] ATLAS Collaboration, *Performance of the ATLAS muon triggers in Run 2*, JINST **15** (2020) P09015, arXiv: 2004.13447 [hep-ex].

[43] ATLAS Collaboration, *The ATLAS Inner Detector Trigger performance in pp collisions at* 13 *TeV during LHC Run 2*, Eur. Phys. J. C **82** (2021) 206, arXiv: 2107.02485 [hep-ex].

[44] ATLAS Collaboration, *Performance of the ATLAS Level-1 topological trigger in Run 2*, Eur. Phys. J. C **82** (2021) 7, arXiv: 2105.01416 [hep-ex].

[45] ATLAS Collaboration, *Measurements of Higgs boson production cross-sections in the $H \to \tau^+\tau^-$ decay channel in pp collisions at $\sqrt{s} = 13$ TeV with the ATLAS detector*, JHEP **08** (2022) 175, arXiv: 2201.08269 [hep-ex].

[46] ATLAS and CMS Collaborations, *Combined Measurement of the Higgs Boson Mass in pp Collisions at $\sqrt{s} = 7$ and 8 TeV with the ATLAS and CMS Experiments*, Phys. Rev. Lett. **114** (2015) 191803, arXiv: 1503.07589 [hep-ex].

[47] D. de Florian et al., LHC Higgs Cross Section Working Group, *Handbook of LHC Higgs cross sections: 4. Deciphering the nature of the Higgs sector*, (2017), arXiv: 1610.07922 [hep-ph].

[48] ATLAS Collaboration, *The ATLAS Simulation Infrastructure*, Eur. Phys. J. C **70** (2010) 823, arXiv: 1005.4568 [physics.ins-det].

[49] S. Agostinelli et al., GEANT4 Collaboration, *GEANT4—a simulation toolkit*, Nucl. Instrum. Meth. A **506** (2003) 250.

[50] T. Sjöstrand, S. Mrenna and P. Skands, *A Brief Introduction to PYTHIA 8.1*, Comput. Phys. Commun. **178** (2008) 852, arXiv: 0710.3820 [hep-ph].

[51] ATLAS Collaboration, *The Pythia 8 A3 tune description of ATLAS minimum bias and inelastic measurements incorporating the Donnachie–Landshoff diffractive model*, ATL-PHYS-PUB-2016-017, 2016, URL: https://cds.cern.ch/record/2206965.

[52] R. D. Ball et al., *Parton distributions with LHC data*, Nucl. Phys. B **867** (2013) 244, arXiv: 1207.1303 [hep-ph].

[53] S. Frixione, G. Ridolfi and P. Nason, *A positive-weight next-to-leading-order Monte Carlo for heavy flavour hadroproduction*, JHEP **09** (2007) 126, arXiv: 0707.3088 [hep-ph].

[54] P. Nason, *A new method for combining NLO QCD with shower Monte Carlo algorithms*, JHEP **11** (2004) 040, arXiv: hep-ph/0409146.

[55] S. Frixione, P. Nason and C. Oleari, *Matching NLO QCD computations with parton shower simulations: the POWHEG method*, JHEP **11** (2007) 070, arXiv: 0709.2092 [hep-ph].

[56] S. Alioli, P. Nason, C. Oleari and E. Re, *A general framework for implementing NLO calculations in shower Monte Carlo programs: the POWHEG BOX*, JHEP **06** (2010) 043, arXiv: 1002.2581 [hep-ph].

[57] H. B. Hartanto, B. Jäger, L. Reina and D. Wackeroth, *Higgs boson production in association with top quarks in the POWHEG BOX*, Phys. Rev. D **91** (2015) 094003, arXiv: 1501.04498 [hep-ph].

[58] T. Sjöstrand et al., *An introduction to PYTHIA 8.2*, Comput. Phys. Commun. **191** (2015) 159, arXiv: 1410.3012 [hep-ph].

[59] J. Butterworth et al., *PDF4LHC recommendations for LHC Run II*, J. Phys. G **43** (2016) 023001, arXiv: 1510.03865 [hep-ph].

[60] J. Pumplin et al., *New Generation of Parton Distributions with Uncertainties from Global QCD Analysis*, JHEP **07** (2002) 012, arXiv: hep-ph/0201195.

[61] ATLAS Collaboration, *Measurement of the Z/γ^* boson transverse momentum distribution in pp collisions at $\sqrt{s} = 7$ TeV with the ATLAS detector*, JHEP **09** (2014) 145, arXiv: 1406.3660 [hep-ex].

[62] R. D. Ball et al., *Parton distributions for the LHC run II*, JHEP **04** (2015) 040, arXiv: 1410.8849 [hep-ph].

[63] ATLAS Collaboration, *ATLAS Pythia 8 tunes to 7 TeV data*, ATL-PHYS-PUB-2014-021, 2014, URL: https://cds.cern.ch/record/1966419.

[64] E. Bothmann et al., *Event generation with Sherpa 2.2*, SciPost Phys. **7** (2019) 034, arXiv: 1905.09127 [hep-ph].

[65] S. Schumann and F. Krauss, *A parton shower algorithm based on Catani–Seymour dipole factorisation*, JHEP **03** (2008) 038, arXiv: 0709.1027 [hep-ph].

[66] H.-L. Lai et al., *New parton distributions for collider physics*, Phys. Rev. D **82** (2010) 074024, arXiv: 1007.2241 [hep-ph].

[67] ATLAS Collaboration, *Reconstruction of primary vertices at the ATLAS experiment in Run 1 proton–proton collisions at the LHC*, Eur. Phys. J. C **77** (2017) 332, arXiv: 1611.10235 [hep-ex].

[68] ATLAS Collaboration, *Electron and photon performance measurements with the ATLAS detector using the 2015–2017 LHC proton–proton collision data*, JINST **14** (2019) P12006, arXiv: 1908.00005 [hep-ex].

[69] ATLAS Collaboration, *Muon reconstruction and identification efficiency in ATLAS using the full Run 2 pp collision data set at* $\sqrt{s} = 13$ *TeV*, Eur. Phys. J. C **81** (2021) 578, arXiv: 2012.00578 [hep-ex].

[70] M. Cacciari, G. P. Salam and G. Soyez, *The anti-k_t jet clustering algorithm*, JHEP **04** (2008) 063, arXiv: 0802.1189 [hep-ph].

[71] M. Cacciari, G. P. Salam and G. Soyez, *FastJet user manual*, Eur. Phys. J. C **72** (2012) 1896, arXiv: 1111.6097 [hep-ph].

[72] ATLAS Collaboration, *Jet reconstruction and performance using particle flow with the ATLAS Detector*, Eur. Phys. J. C **77** (2017) 466, arXiv: 1703.10485 [hep-ex].

[73] ATLAS Collaboration, *Selection of jets produced in* 13 *TeV proton–proton collisions with the ATLAS detector*, ATLAS-CONF-2015-029, 2015, URL: https://cds.cern.ch/record/2037702.

[74] ATLAS Collaboration, *Performance of pile-up mitigation techniques for jets in pp collisions at* $\sqrt{s} = 8$ *TeV using the ATLAS detector*, Eur. Phys. J. C **76** (2016) 581, arXiv: 1510.03823 [hep-ex].

[75] ATLAS Collaboration, *Identification and rejection of pile-up jets at high pseudorapidity with the ATLAS detector*, Eur. Phys. J. C **77** (2017) 580, arXiv: 1705.02211 [hep-ex], Erratum: Eur. Phys. J. C **77** (2017) 712.

[76] ATLAS Collaboration, *ATLAS b-jet identification performance and efficiency measurement with $t\bar{t}$ events in pp collisions at* $\sqrt{s} = 13$ *TeV*, Eur. Phys. J. C **79** (2019) 970, arXiv: 1907.05120 [hep-ex].

[77] ATLAS Collaboration, *Optimisation and performance studies of the ATLAS b-tagging algorithms for the 2017-18 LHC run*, ATL-PHYS-PUB-2017-013, 2017, URL: https://cds.cern.ch/record/2273281.

[78] ATLAS Collaboration, *ATLAS flavour-tagging algorithms for the LHC Run 2 pp collision dataset*, 2022, arXiv: 2211.16345 [hep-ex].

[79] ATLAS Collaboration, *Topological cell clustering in the ATLAS calorimeters and its performance in LHC Run 1*, Eur. Phys. J. C **77** (2017) 490, arXiv: 1603.02934 [hep-ex].

[80] ATLAS Collaboration, *Identification and energy calibration of hadronically decaying tau leptons with the ATLAS experiment in pp collisions at* $\sqrt{s} = 8$ *TeV*, Eur. Phys. J. C **75** (2015) 303, arXiv: 1412.7086 [hep-ex].

[81] ATLAS Collaboration, *Measurement of the tau lepton reconstruction and identification performance in the ATLAS experiment using pp collisions at* $\sqrt{s} = 13$ *TeV*, ATLAS-CONF-2017-029, 2017, URL: https://cds.cern.ch/record/2261772.

[82] ATLAS Collaboration, *Identification of hadronic tau lepton decays using neural networks in the ATLAS experiment*, ATL-PHYS-PUB-2019-033, 2019, URL: https://cds.cern.ch/record/2688062.

[83] ATLAS Collaboration, *Reconstruction, Energy Calibration, and Identification of Hadronically Decaying Tau Leptons in the ATLAS Experiment for Run-2 of the LHC*, ATL-PHYS-PUB-2015-045, 2015, URL: https://cds.cern.ch/record/2064383.

[84] ATLAS Collaboration, *Performance of missing transverse momentum reconstruction with the ATLAS detector using proton–proton collisions at* $\sqrt{s} = 13$ *TeV*, Eur. Phys. J. C **78** (2018) 903, arXiv: 1802.08168 [hep-ex].

[85] A. Elagin, P. Murat, A. Pranko and A. Safonov, *A new mass reconstruction technique for resonances decaying to* $\tau\tau$, Nucl. Instrum. Meth. A **654** (2011) 481, arXiv: 1012.4686 [hep-ex].

[86] R. K. Ellis, I. Hinchliffe, M. Soldate and J. J. van der Bij, *Higgs decay to* $\tau^+ \tau^-$ *A possible signature of intermediate mass Higgs bosons at high energy hadron colliders*, Nucl. Phys. B **297** (1988) 221.

[87] ATLAS Collaboration, *Evidence for the Higgs-boson Yukawa coupling to tau leptons with the ATLAS detector*, JHEP **04** (2015) 117, arXiv: 1501.04943 [hep-ex].

[88] ATLAS Collaboration, *Measurements of Higgs boson production and couplings in the four-lepton channel in* pp *collisions at center-of-mass energies of* 7 *and* 8 *TeV with the ATLAS detector*, Phys. Rev. D **91** (2015) 012006, arXiv: 1408.5191 [hep-ex].

[89] ATLAS Collaboration, *Search for lepton-flavour-violating decays of the Higgs and Z bosons with the ATLAS detector*, Eur. Phys. J. C **77** (2017) 70, arXiv: 1604.07730 [hep-ex].

[90] ATLAS Collaboration, *Electron reconstruction and identification in the ATLAS experiment using the 2015 and 2016 LHC proton–proton collision data at* $\sqrt{s} = 13$ *TeV*, Eur. Phys. J. C **79** (2019) 639, arXiv: 1902.04655 [hep-ex].

[91] ATLAS Collaboration, *Muon reconstruction performance of the ATLAS detector in proton–proton collision data at* $\sqrt{s} = 13$ *TeV*, Eur. Phys. J. C **76** (2016) 292, arXiv: 1603.05598 [hep-ex].

[92] A. Hoecker et al., *TMVA - Toolkit for Multivariate Data Analysis*, 2007, arXiv: physics/0703039 [physics.data-an].

[93] F. Chollet et al., *Keras*, 2015, URL: https://keras.io.

[94] M. Abadi et al., 'Tensorflow: A system for large-scale machine learning', *12th USENIX Symposium on Operating Systems Design and Implementation (OSDI'16)*, 2016 265.

[95] ATLAS Collaboration, *Electron and photon energy calibration with the ATLAS detector using 2015–2016 LHC proton–proton collision data*, JINST **14** (2019) P03017, arXiv: 1812.03848 [hep-ex].

[96] ATLAS Collaboration, *Jet energy scale measurements and their systematic uncertainties in proton–proton collisions at* $\sqrt{s} = 13$ *TeV with the ATLAS detector*, Phys. Rev. D **96** (2017) 072002, arXiv: 1703.09665 [hep-ex].

[97] ATLAS Collaboration, *Tagging and suppression of pileup jets with the ATLAS detector*, ATLAS-CONF-2014-018, 2014, URL: https://cds.cern.ch/record/1700870.

[98] ATLAS Collaboration, *Jet energy scale and resolution measured in proton–proton collisions at* $\sqrt{s} = 13$ *TeV with the ATLAS detector*, Eur. Phys. J. C **81** (2020) 689, arXiv: 2007.02645 [hep-ex].

[99] ATLAS Collaboration,
Luminosity determination in pp collisions at $\sqrt{s}$ = 13 TeV using the ATLAS detector at the LHC,
ATLAS-CONF-2019-021, 2019, URL: https://cds.cern.ch/record/2677054.

[100] G. Avoni et al., *The new LUCID-2 detector for luminosity measurement and monitoring in ATLAS*,
JINST **13** (2018) P07017.

[101] L. Lönnblad and S. Prestel, *Matching tree-level matrix elements with interleaved showers*,
JHEP **03** (2012) 019, arXiv: 1109.4829 [hep-ph].

[102] M. Bähr et al., *Herwig++ physics and manual*, Eur. Phys. J. C **58** (2008) 639,
arXiv: 0803.0883 [hep-ph].

[103] J. Bellm et al., *Herwig 7.0/Herwig++ 3.0 release note*, Eur. Phys. J. C **76** (2016) 196,
arXiv: 1512.01178 [hep-ph].

[104] ATLAS Collaboration,
Studies on top-quark Monte Carlo modelling with Sherpa and MG5_aMC@NLO,
ATL-PHYS-PUB-2017-007, 2017, URL: https://cds.cern.ch/record/2261938.

[105] L. A. Harland-Lang, A. D. Martin, P. Motylinski and R. S. Thorne,
Parton distributions in the LHC era: MMHT 2014 PDFs, Eur. Phys. J. C **75** (2015) 204,
arXiv: 1412.3989 [hep-ph].

[106] S. Hoche, F. Krauss, M. Schonherr and F. Siegert, *NLO matrix elements and truncated showers*,
JHEP **08** (2011) 123, arXiv: 1009.1127 [hep-ph].

[107] T. Gleisberg, F. Krauss, A. Schalicke, S. Schumann and J.-C. Winter,
Studying W^+W^- production at the Fermilab Tevatron with SHERPA,
Phys. Rev. D **72** (2005) 034028, arXiv: hep-ph/0504032.

[108] T. Binoth, T. Gleisberg, S. Karg, N. Kauer and G. Sanguinetti,
NLO QCD corrections to ZZ+ jet production at hadron colliders, Phys. Lett. B **683** (2010) 154,
arXiv: 0911.3181 [hep-ph].

[109] S. Dittmaier et al., LHC Higgs Cross Section Working Group,
Handbook of LHC Higgs Cross Sections: 1. Inclusive Observables, (2011),
arXiv: 1101.0593 [hep-ph].

[110] J. Alwall et al., *The automated computation of tree-level and next-to-leading order differential cross sections, and their matching to parton shower simulations*, JHEP **07** (2014) 079,
arXiv: 1405.0301 [hep-ph].

[111] G. Cowan, K. Cranmer, E. Gross and O. Vitells,
Asymptotic formulae for likelihood-based tests of new physics, Eur. Phys. J. C **71** (2011) 1554,
arXiv: 1007.1727 [physics.data-an], Erratum: Eur. Phys. J. C **73** (2013) 2501.

[112] A. L. Read, *Presentation of search results: the CL_S technique*, J. Phys. G **28** (2002) 2693.

[113] LHC Higgs Cross Section Working Group,
Handbook of LHC Higgs Cross Sections: 3. Higgs Properties, (2013),
arXiv: 1307.1347 [hep-ph].

[114] BaBar Collaboration,
Searches for Lepton Flavor Violation in the Decays $\tau^\pm \to e^\pm\gamma$ and $\tau^\pm \to \mu^\pm\gamma$,
Phys. Rev. Lett. **104** (2010) 021802, arXiv: 0908.2381 [hep-ex].

[115] S. Davidson, S. Lacroix and P. Verdier, *LHC sensitivity to lepton flavour violating Z boson decays*, JHEP **09** (2012) 092, arXiv: 1207.4894 [hep-ph].

[116] A. Blum, A. Kalai and J. Langford,
'Beating the Hold-out: Bounds for K-Fold and Progressive Cross-Validation',
Proceedings of the Twelfth Annual Conference on Computational Learning Theory, 1999 203,
URL: https://doi.org/10.1145/307400.307439.

[117] T. Akiba, S. Sano, T. Yanase, T. Ohta and M. Koyama,
'Optuna: A Next-generation Hyperparameter Optimization Framework', *Proceedings of the 25rd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, 2019.

[118] ATLAS Collaboration,
Evaluating statistical uncertainties and correlations using the bootstrap method,
ATL-PHYS-PUB-2021-011, 2021, URL: https://cds.cern.ch/record/2759945.