Title
Searches for lepton-flavour-violating decays of the Higgs boson in $s = 13$ TeV pp collisions with the ATLAS detector

Permalink
https://escholarship.org/uc/item/7bx35096

Authors
Aad, G
Abbott, B
Abbott, DC
et al.

Publication Date
2020

DOI
10.1016/j.physletb.2019.135069

Peer reviewed
Searches for lepton-flavour-violating decays of the Higgs boson in \(\sqrt{s} = 13\) TeV \(pp\) collisions with the ATLAS detector

The ATLAS Collaboration*

**Abstract**

This Letter presents direct searches for lepton flavour violation in Higgs boson decays, \(H \to e\tau\) and \(H \to \mu\tau\), performed 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 36.1 fb\(^{-1}\). No significant excess is observed above the expected background from Standard Model processes. The observed (median expected) 95% confidence-level upper limits on the lepton-flavour-violating branching ratios are 0.47% (0.34\%\(^{+0.15\%}_{-0.20\%}\)) and 0.28% (0.37\%\(^{+0.14\%}_{-0.16\%}\)) for \(H \to e\tau\) and \(H \to \mu\tau\), respectively.

© 2019 The Author(s). Published by Elsevier B.V. This is an open access article under the CC BY license (http://creativecommons.org/licenses/by/4.0/). Funded by SCOAP3.

1. Introduction

The search for processes at the Standard Model (SM) is one of the main goals of the Large Hadron Collider (LHC) programme at CERN. A possible sign of such processes is lepton flavour violation (LFV) in decays of the Higgs boson \([1,2]\). Many beyond-SM theories predict LFV decays of the Higgs boson, such as supersymmetry \([3,4]\), other models with more than one Higgs doublet \([5,6]\), composite Higgs models \([7]\), models with flavour symmetries \([8]\) or warped extra dimensions \([9–11]\) models and others \([12,13]\).

In this Letter, searches for LFV decays of the Higgs boson, \(H \to e\tau\) and \(H \to \mu\tau\), at the LHC with the ATLAS experiment are presented. Studies are based on proton–proton \((pp)\) collision data recorded in 2015–2016 at a centre-of-mass energy \(\sqrt{s} = 13\) TeV. The dataset corresponds to an integrated luminosity of 36.1 fb\(^{-1}\).

Previous ATLAS searches \([14,15]\) placed an upper limit of 1.04% (1.43%) on the \(H \to e\tau\) (\(H \to \mu\tau\)) branching ratio \((B)\) with a 95% confidence level (CL) using Run 1 data collected at \(\sqrt{s} = 8\) TeV, corresponding to an integrated luminosity of 20.3 fb\(^{-1}\). The CMS Collaboration recently provided 95% CL upper limits on these branching ratios of 0.61% and 0.25%, respectively, using data collected at \(\sqrt{s} = 13\) TeV, with an integrated luminosity of 35.9 fb\(^{-1}\) \([16]\).

The searches presented here involve both leptonic \((\tau \to \ell'\nu\bar{\nu})\) and hadronic \((\tau \to \text{hadrons} + \nu)\) decays of \(\tau\)-leptons, denoted \(\tau_{\ell'}\) and \(\tau_{\text{had}}\) respectively. The dilepton final state \(\ell\ell'\) only considers pairs of different-flavour leptons. Same-flavour lepton pairs are rejected due to the large lepton pair-production Drell-Yan background. Two channels are considered for each of the two searches: \(e\tau_{\mu}\) and \(e\tau_{\text{had}}\) for the \(H \to e\tau\) search, \(\mu\tau_{\mu}\) and \(\mu\tau_{\text{had}}\) for the \(H \to \mu\tau\) search. The analysis is designed such that any potential LFV signal overlap between the \(H \to e\tau\) and \(H \to \mu\tau\) searches is negligible. Many methods are reused from the measurement of the Higgs boson cross-section in the \(H \to \tau\tau\) final state \([17]\).

The ATLAS detector\(^2\) is described in Refs. \([18–20]\). It consists of an inner tracking detector covering the range \(|\eta| < 2.5\), surrounded by a superconducting solenoid providing a 2 T axial magnetic field, high-granularity electromagnetic \((|\eta| < 3.2)\) and hadronic calorimeters \((|\eta| < 4.9)\), and a muon spectrometer (MS) which covers the range \(|\eta| < 2.7\) and includes fast trigger chambers \((|\eta| < 2.4)\) and superconducting toroidal magnets.

2. Simulation samples

Samples of Monte Carlo (MC) simulated events are used to optimize the event selection, and to model the signal and several of the background processes. The samples were produced with the ATLAS simulation infrastructure \([21]\) using the full detector simulation performed by the GEANT4 \([22]\) toolkit. The Higgs boson mass was set to \(m_H = 125\) GeV \([23]\). The four leading Higgs boson production mechanisms are considered: the gluon–gluon fusion (ggF), vector-boson fusion (VBF) and two associated production modes

---

* E-mail address: atlas.publications@cern.ch.

1 Unless explicitly mentioned otherwise, leptons (denoted by \(\ell\) or \(\ell'\)) refer to electrons or muons.

2 ATLAS uses a right-handed coordinate system with its origin at the nominal interaction point in the centre of the detector and the \(z\)-axis along the beam pipe. The azimuthal angle \(\phi\) runs around the beam pipe, the pseudorapidity is defined in terms of the polar angle \(\theta\) as \(\eta = \ln(\tan(\theta/2))\). Angular distance in the \(\eta\)–\(\phi\) space is defined as \(\Delta R = \sqrt{(\Delta\eta)^2 + (\Delta\phi)^2}\).
(WH, ZH), while the others give negligible contributions and are ignored. The cross-sections of all Higgs boson production processes were normalized to the SM predictions [24]. The LFV Higgs boson decays as well as the $H \rightarrow \tau\tau$ and $H \rightarrow WW$ background decays were modelled with Pythia 8 [25]. Other background processes involve electroweak production of $W/Z$ bosons via VBF, Drell–Yan production of $W/Z$ in association with jet(s) as well as diboson, single top-quark and top–quark pair ($tt\ell$) production. The MC generators used for each production cross-section measurement [17] were also employed here for all background components. The generators and parton shower models used to simulate different processes are summarized in Table 1.

### 3. Object reconstruction

The correct identification of $H \rightarrow \ell\tau$ events requires reconstruction of several different objects (electrons, muons, and jets, including those initiated by hadronic decays of $\tau$–leptons) and the missing transverse momentum $p_T^{miss}$, whose magnitude is called $E_T^{miss}$.

Electrons are reconstructed by matching tracks in the inner detector to clustered energy deposits in the electromagnetic calorimeter [57]. Loose likelihood-based identification [58], $p_T > 15$ GeV and fiducial volume requirements ($|\eta| < 1.4$) are applied. Medium identification, corresponding to an efficiency of 87% at $p_T = 20$ GeV, is imposed for the baseline electron selection.

Muons are identified by tracks reconstructed in the inner detector and matched to tracks in the MS. Loose identification [59], $p_T > 10$ GeV and $|\eta| < 2.5$ requirements are applied. Medium identification (efficiency of 96.1% for muons with $p_T > 20$ GeV) is imposed for the baseline muon selection.

Isolation criteria exploiting calorimeter and track-based information are applied to both electrons and muons. The gradient working point is used, featuring an efficiency of 90% (99%) obtained for leptons with $p_T > 25$ GeV (60 GeV) originating from the $Z \rightarrow \ell\ell$ process [58,59].

Jets are reconstructed using the anti-$k_t$ algorithm [60] as implemented by the FastJet [61] package. The algorithm is applied to topological clusters of calorimeter cells [62] with a radius parameter $R = 0.4$. Only jets with $p_T > 20$ GeV and $|\eta| < 4.5$ are considered. Jets from other $pp$ interactions in the same and neighbouring bunch crossings (pile-up) are suppressed using jet vertex tagger [JVT] algorithms [63,64]. Jets containing b-hadrons (b-jets) are identified by the MV2c20 algorithm [65,66] in the central region ($|\eta| < 2.4$). A working point corresponding to 85% average efficiency determined for b-jets in $\tau$ simulated events is chosen, rejection factors are 2.8 and 28 against c-jets and light-flavour jets respectively.

The reconstruction of the object formed by the visible products of the $t\bar{t}$ hadron decay ($t\bar{t}_{\text{had-vis}}$) begins from jets reconstructed by the anti-$k_t$ jet algorithm with a radius parameter $R = 0.4$. Information from the inner detector tracks associated with the energy deposits in the calorimeter is incorporated in the reconstruction. Only $t\bar{t}_{\text{had-vis}}$ candidates with $p_T > 20$ GeV and $|\eta| < 2.5$ are considered. One or three associated tracks with an absolute total charge $q = 1$ are required. An identification algorithm [67,68] based on boosted decision trees (BDT) [69–71] is used to reject $t\bar{t}_{\text{had-vis}}$ candidates arising from misidentification of jets or from decays of hadrons with $b$- or $c$-quark content. Unless otherwise indicated, a tight identification (ID) working point is used for the $t\bar{t}_{\text{had-vis}}$, corresponding to an efficiency of 60% (45%) for 1-prong (3-prong) candidates. Jets corresponding to identified $t\bar{t}_{\text{had-vis}}$ candidates are removed from the jet collection. The $t\bar{t}_{\text{had-vis}}$ candidates with one track overlapping with an electron candidate with high ID score, as determined by a multivariate (MVA) approach, are rejected. Leptonic $\tau$-decays are reconstructed as electrons or muons.

Events considered in the analysis are triggered with single-electron or single-muon triggers. The $p_T$ thresholds depend on the isolation requirement and data-taking period [72,73]. The lowest trigger thresholds correspond to $25-27$ GeV (electrons) and $21-27$ GeV (muons).

### 4. Event selection and categorization

Events selected in the $\ell\tau$ channel contain exactly one electron and one muon of opposite-sign (OS) charges. Similarly in the $\ell\ell$ channel, a lepton and a $t\bar{t}_{\text{had-vis}}$ of OS charges are required, and events with more than one baseline lepton are rejected. The selection criteria are summarized in Table 2 for the analysis categories as well as the control regions (CRs), which are described in Section 5.

In the $\ell\tau$ channel, $\ell_1$ and $\ell_2$ denote the leading and subleading lepton in $p_T$, respectively. Events where the leading lepton is an electron (muon) are used in the search for $H \rightarrow \ell\tau_\ell$ ($H \rightarrow \mu\tau_\mu$). A requirement on the dilepton invariant mass, equal to the invariant mass of the lepton and the visible $\tau$–decay products, $M_{\text{vis}}$, reduces backgrounds with top quarks, and the criterion applied to the track-to-cluster $p_T$ ratio of the electron reduces the $Z \rightarrow \mu\mu$ background where a muon deposits a large amount of energy in the electromagnetic calorimeter and is misidentified as an electron in the $\mu\tau$ channel. The contribution from the $H \rightarrow \tau\tau$ decay is reduced by the asymmetric $p_T$ selection of the two leptons.

In the $\ell\ell$ channel, the criterion based on the azimuthal separations of leptons–$E_T^{miss}$ and $t\bar{t}_{\text{had-vis}}$–$E_T^{miss}$, $\sum_{i=1}^{2}\delta_{\text{had-vis}} \sum_{\ell=1}^{2} \cos \Delta \phi_{i,\ell}$, reduces the $W + \text{jets}$ background whereas the requirement on $|\Delta\eta(\ell, t\bar{t}_{\text{had-vis}})|$ reduces backgrounds with misidentified $t\bar{t}_{\text{had-vis}}$ candidates.

For both channels of each search, a $b$- veto requirement reduces the single-top-quark and $t\bar{t}$ backgrounds. Events are further cate-

---

1. The transition region in $\eta$ is excluded, similarly to electrons.
Table 2
Baseline event selection and further categorization for the $\ell_1\tau$ and $\ell_1T_{\text{had}}$ channels. The same criteria are also used for the control region (CR) definitions in the $\ell_1\tau$ channel (Section 5), but one requirement of the baseline selection is inverted to achieve orthogonal event selection. There is no CR in the $\ell_1T_{\text{had}}$ channel.

<table>
<thead>
<tr>
<th>Selection</th>
<th>$\ell_1\tau$</th>
<th>$\ell_1T_{\text{had}}$</th>
</tr>
</thead>
<tbody>
<tr>
<td>Baseline</td>
<td>exactly 1e and 1$\mu$, OS</td>
<td>exactly 1e and 1$\tau_{\text{had-viss.}}$, OS</td>
</tr>
<tr>
<td></td>
<td>$p_T^{1e} &gt; 45$ GeV</td>
<td>$p_T^{1e} &gt; 27.3$ GeV</td>
</tr>
<tr>
<td></td>
<td>$p_T^{1\mu} &gt; 15$ GeV</td>
<td>$p_T^{1\tau_{\text{had-viss.}}} &gt; 25$ GeV, $</td>
</tr>
<tr>
<td></td>
<td>$30$ GeV $&lt; m_{4\ell} &lt; 150$ GeV</td>
<td>$\sum_{i} \cos \Delta\phi(E_{T_{\text{miss}}}) - 0.35$</td>
</tr>
<tr>
<td></td>
<td>$p_T^{(\text{track})}/p_T^{(\text{cluster})} &lt; 1.2$ (for jets with $p_T &gt; 25$ GeV and $</td>
<td>\eta</td>
</tr>
<tr>
<td></td>
<td>$b$-veto (for jets with $p_T &gt; 25$ GeV and $</td>
<td>\eta</td>
</tr>
<tr>
<td>VBF</td>
<td>Baseline plus fail VBF categorization</td>
<td></td>
</tr>
<tr>
<td>Non-VBF</td>
<td>$m_T(t_1, E_{T_{\text{miss}}}) &gt; 50$ GeV</td>
<td></td>
</tr>
<tr>
<td></td>
<td>$m_T(t_1, E_{T_{\text{miss}}}) &lt; 40$ GeV</td>
<td></td>
</tr>
<tr>
<td></td>
<td>$</td>
<td>\Delta\phi(t_2, E_{T_{\text{miss}}})</td>
</tr>
<tr>
<td></td>
<td>$p_T^{(1)}/p_T^{(2)} &gt; 0.5$</td>
<td></td>
</tr>
<tr>
<td>Non-VBF</td>
<td>$p_T^{(1)} &gt; 40$ GeV, $p_T^{(2)} &gt; 30$ GeV</td>
<td></td>
</tr>
<tr>
<td></td>
<td>$</td>
<td>\Delta\eta(j_1, j_2)</td>
</tr>
<tr>
<td></td>
<td>$p_T^{(\text{had-viss.})} &gt; 45$ GeV</td>
<td></td>
</tr>
<tr>
<td>VBF</td>
<td>$\geq 2$ jets, $p_T^{(1)} &gt; 40$ GeV, $p_T^{(2)} &gt; 30$ GeV</td>
<td></td>
</tr>
<tr>
<td></td>
<td>$</td>
<td>\Delta\eta(j_1, j_2)</td>
</tr>
<tr>
<td></td>
<td>$p_T^{(\text{had-viss.})} &gt; 45$ GeV</td>
<td></td>
</tr>
<tr>
<td>Top-quark CR</td>
<td>inverted $b$-veto:</td>
<td></td>
</tr>
<tr>
<td>VBF and non-VBF</td>
<td>$\geq 1$ $b$-tagged jet ($p_T &gt; 25$ GeV and $</td>
<td>\eta</td>
</tr>
<tr>
<td>Top-quark CR</td>
<td>$Z \rightarrow \tau\tau$ CR</td>
<td></td>
</tr>
<tr>
<td>VBF and non-VBF</td>
<td>inverted $p_T^{(1)}$ requirement:</td>
<td></td>
</tr>
<tr>
<td>VBF and non-VBF</td>
<td>$35$ GeV $&lt; p_T^{(1)} &lt; 45$ GeV</td>
<td></td>
</tr>
</tbody>
</table>

The analysis exploits BDT algorithms to enhance the signal separation from the background in the individual searches, channels and categories. The components of the four-momenta of the analysis objects as well as derived event variables (e.g., invariant masses and angular separations) are the input variables of the BDT discriminant. Correlations between these input variables have been carefully checked, highly correlated variables have been removed and the remaining ones are ranked according to their discrimination power [74,75]. The list of variables is then optimized, removing the lowest-ranked variables with marginal contribution to the sensitivity. The final list of variables is presented in Table 3 for each channel and category. The invariant mass of the Higgs boson reconstructed under the $H \rightarrow \tau\tau$ decay hypothesis exhibits the highest signal-to-background separation power and it helps to distinguish LFV signal from $H \rightarrow \tau\tau$ and $H \rightarrow WW$ backgrounds. For the $\ell_1\tau$ channel the invariant mass is reconstructed with the MMC algorithm [76] and is denoted by $m_{\text{MMC}}$; for the $\ell_1T_{\text{had}}$ channel it is reconstructed with the collinear approximation [76] and is denoted by $m_{\text{coll}}$. The analysis CRs are used to validate the level of agreement between data and simulated distributions of the BDT score and input variables, as well as their correlations.

5. Background modelling

The most significant background in the search are from events with $Z \rightarrow \tau\tau$ decays or with (single or pair-produced) top quarks, especially in the $\ell_1\tau$ channel, as well as from events with misidentified objects, which are estimated using data-driven (d.d.) techniques. The relative contribution from misidentified objects to the total background yield is 5–25% in the $\ell_1\tau$ channel and 25–45% in the $\ell_1T_{\text{had}}$ channel, depending on the search and the analysis category. The shapes of distributions from the $Z \rightarrow \tau\tau$ and top-quark (single-top-quark and $t\bar{t}$) processes are modelled by simulation in both the $\ell_1\tau$ and $\ell_1T_{\text{had}}$ decay channels. In the $\ell_1\tau$ channel, the relative contributions of $Z \rightarrow \tau\tau$ and top-quark production processes are 20–35% and 20–55%, respectively; the top-quark background dominates in the VBF category. In the $\ell_1T_{\text{had}}$ channel, the top-quark background fraction is 1–10%, while the $Z \rightarrow \tau\tau$ process contributes 45–55% of the total background. The individual contributions are listed in Tables 4 and 5. Smaller background components are also modelled by simulation and are grouped together: $Z \rightarrow \mu\mu$, diboson production, $H \rightarrow \tau\tau$ and $H \rightarrow WW$.

Good modelling of the background is demonstrated in Fig. 1 for a selection of important BDT input variables. Details of the background estimation techniques are given below.

5.1. $\ell_1\tau$ Channel

Two sets of CRs, as defined in Table 2, are used to constrain the normalization of $Z \rightarrow \tau\tau$ and top-quark background components. These CRs inherit their definitions from the corresponding analysis category but invert one requirement to ensure orthogonality with the nominal selection. The normalization factors are determined during the statistical analysis by fitting the event yields in all signal and control regions simultaneously. For each search, separate $Z \rightarrow \tau\tau$ normalization factors are used for the VBF and non-VBF categories. In the case of the top-quark background, in...
Table 3

BDT input variables used in the analysis. For each channel and category, used input variables are marked with HR (indicating the five variables with the highest rank) or a bullet. Analogous variables between the two channels are listed on the same line.

<table>
<thead>
<tr>
<th>Variable</th>
<th>VBF</th>
<th>non-VBF</th>
</tr>
</thead>
<tbody>
<tr>
<td>( m_{\text{dNN}} )</td>
<td>HR</td>
<td>HR</td>
</tr>
<tr>
<td>( p_T^{\ell} )</td>
<td>●</td>
<td>●</td>
</tr>
<tr>
<td>( \Delta R(\ell, j) )</td>
<td>HR</td>
<td>HR</td>
</tr>
<tr>
<td>( m_{\ell \tau j} )</td>
<td>●</td>
<td>●</td>
</tr>
<tr>
<td>( \Delta \Phi(\ell, E_{\text{T}}^{\text{miss}}) )</td>
<td>HR</td>
<td>HR</td>
</tr>
<tr>
<td>( \Delta \Phi(j_1, j_2) )</td>
<td>HR</td>
<td>HR</td>
</tr>
<tr>
<td>( p_T^{\ell}/p_T^{j} )</td>
<td>HR</td>
<td>HR</td>
</tr>
</tbody>
</table>

Table 4

Event yields and predictions as determined by the background-only fit in different signal regions of the \( H \to \tau \tau \) analysis. Uncertainties include both the statistical and systematic contributions. "Other" contains diboson, \( Z \to \ell \ell \), \( H \to \tau \tau \) and \( H \to WW \) backgrounds. For the \( \ell \tau \text{had} \) channel the "Z \to \ell \ell \,(d.d.)" component corresponds to electrons misidentified as \( \tau \text{had} \). This contribution is summed with "Other" since there are few events in the VBF category. The uncertainty of the total background includes all correlations between channels. The normalizations of top-quark \( (\ell \tau \text{had} \text{channel only}) \) and \( Z \to \tau \tau \) background components are determined by the fit, while the expected signal event yields are given by \( B(H \to \tau \tau) = 1\% \).

<table>
<thead>
<tr>
<th>( \ell \tau \text{had} )</th>
<th>VBF</th>
<th>non-VBF</th>
</tr>
</thead>
<tbody>
<tr>
<td>( \epsilon_{\ell \tau} )</td>
<td>379 ± 37</td>
<td>19.8 ± 2.7</td>
</tr>
<tr>
<td>( Z \to \tau \tau )</td>
<td>2470 ± 230</td>
<td>221 ± 34</td>
</tr>
<tr>
<td>Top-quark</td>
<td>1640 ± 140</td>
<td>490 ± 40</td>
</tr>
<tr>
<td>Mis-identified</td>
<td>1330 ± 250</td>
<td>73 ± 33</td>
</tr>
<tr>
<td>( Z \to \ell \ell ,(d.d.) )</td>
<td>1700 ± 80</td>
<td>220 ± 15</td>
</tr>
<tr>
<td>Other</td>
<td>7130 ± 100</td>
<td>1003 ± 33</td>
</tr>
<tr>
<td>Total background</td>
<td>7128</td>
<td>982</td>
</tr>
<tr>
<td>Data</td>
<td>1180 ± 110</td>
<td>25 ± 4</td>
</tr>
</tbody>
</table>

Table 5

Event yields and predictions as determined by the background-only fit in different signal regions of the \( H \to \mu \tau \) analysis. Uncertainties include both the statistical and systematic contributions. "Other" contains diboson, \( Z \to \ell \ell \), \( H \to \tau \tau \) and \( H \to WW \) backgrounds. The uncertainty of the total background includes all correlations between channels. The normalizations of top-quark \((\ell \tau \text{had} \text{ channel only})\) and \( Z \to \tau \tau \) background components are determined by the fit, while the expected signal event yields are given by \( B(H \to \mu \tau) = 1\% \).

<table>
<thead>
<tr>
<th>( \ell \tau \text{had} )</th>
<th>VBF</th>
<th>non-VBF</th>
</tr>
</thead>
<tbody>
<tr>
<td>( \epsilon_{\mu \tau} )</td>
<td>287 ± 23</td>
<td>14.6 ± 1.9</td>
</tr>
<tr>
<td>( Z \to \tau \tau )</td>
<td>1860 ± 130</td>
<td>144 ± 26</td>
</tr>
<tr>
<td>Top-quark</td>
<td>1340 ± 210</td>
<td>41 ± 21</td>
</tr>
<tr>
<td>Mis-identified</td>
<td>1180 ± 140</td>
<td>168 ± 18</td>
</tr>
<tr>
<td>Other</td>
<td>5684 ± 100</td>
<td>743 ± 29</td>
</tr>
<tr>
<td>Total background</td>
<td>1200 ± 200</td>
<td>25 ± 5</td>
</tr>
<tr>
<td>Data</td>
<td>184500 ± 1200</td>
<td>580 ± 30</td>
</tr>
</tbody>
</table>

which leading jets are produced at a lower order of the perturbative expansion of the scattering process, a combined normalization factor across the two categories is used in the \( \ell \tau \text{had} \) channel.

Top-quark \( CRs \) are almost exclusively composed of top-quark backgrounds: the purity is 95% across both searches and categories, with \( t \bar{t} \) process accounting for more than 90% of the top-quark backgrounds. The \( Z \to \tau \tau \) \( CRs \) achieved a purity of \( \sim 80\% \) in the non-VBF categories, while a lower purity of \( \sim 60\% \) is observed in the VBF categories. The contributions of all other background components are normalized to their SM predictions when the likelihood fit (Section 7) is applied.

The shape and normalization of diboson and \( Z \to \mu \tau \) background distributions are validated with data in dedicated regions where their contributions are enhanced. The latter process only contributes sizeably in the \( \ell \tau \text{had} \) channel, where it represents up to 10% of the total background.
Another source of background comes from $W + \text{jets}$, top-quark and multi-jet events, where jets are misidentified as leptons. This background is estimated directly from OS data events where an inverted isolation requirement is imposed on the subleading lepton [17]. Normalization factors are applied to correct for the inverted isolation requirement. The normalization factors are derived in a dedicated region where the leptons are required to have same-sign (SS) charges. Additional corrections are made by reweighting the MC distributions of $\Delta \phi (\ell_1, E^{\text{miss}}, \ell_2)$ and $\Delta \phi (\ell_2, E^{\text{miss}})$ to data in the SS region, which improves the modelling of non-collinear jet-lepton separation. A similar improvement is observed in the nominal OS region. In most of the cases, the misidentified jet mimics the lepton of lower $p_T$, $\ell_2$, while the fraction of events where both leptons are misidentified varies between 2% to 8% across categories. The systematic uncertainties of the estimation of the misidentified lepton background include contributions from closure tests in SS and OS regions enriched with misidentified leptons, from the corrections made to the $\Delta \phi$ distributions, and from the composition of the misidentified lepton background.

5.2. $\ell_{\text{had}}$ channel

The main background contributions come from the $Z \to \tau \tau$ process and events where either a jet or an electron is misidentified as $\ell_{\text{had}}$-vis. The shape of the $Z \to \tau \tau$ background distribution is modelled by simulation, and the corresponding normalization factors are determined from the simultaneous fit of the event yields in all signal and control regions. The $Z \to \tau \tau$ normalization factors are fully correlated with those of the $\ell_{\text{had}}$-vis channel, in each VBF and non-VBF category. Top-quark production represents less than 1% of the total background in the $\ell_{\text{had}}$ channel and is determined by simulation, including its normalization, which is kept fixed in the fit.

The main contributions to jets misidentified as $\ell_{\text{had}}$-vis come from multi-jet events and $W$-boson production in association with jets, and a fake-factor method is used to estimate the contribution of each component separately. A fake factor is defined as the ratio of the number of events where the highest-$p_T$ jet is identified as a tight $\ell_{\text{had}}$-vis candidate to the number of events where the highest-$p_T$ jet fails to satisfy this $\tau$-ID criterion but satisfies a looser criterion. The procedure, including systematic uncertainties, is described in Ref. [17]. Since a different $\tau$-ID working point
6. Systematic uncertainties

The systematic uncertainties affect the normalization of signal and background, and/or the shape of their corresponding final discriminant distributions. Each source of systematic uncertainty is considered to be uncorrelated with the other sources. The effect of each systematic uncertainty is fully considered in each category, including control regions. Correlations of each systematic uncertainty are maintained across processes, channels, categories and regions. The size of the systematic uncertainties and their impact on the fitted branching ratio are discussed in Section 7. The main sources of systematic uncertainties are related to the estimation of the backgrounds originating from mis-identified leptons/jets and to the jet energy scale uncertainties.

Experimental uncertainties include those originating from the reconstruction, identification, tagging and triggering efficiencies of all physics objects as well as their momentum scale and resolution. These include effects from leptons [57–59], hadrons [63, 64,77] and E\(_T\) [78]. Uncertainties affecting the kinematics of the physics objects are propagated to the BDT input variables. The corresponding shape and normalization variations of the BDT discriminant are considered in the statistical analysis. Uncertainties of the luminosity measurement [79], pile-up modelling and uncertainties specific to mis-identified background estimation techniques mentioned in Section 5 are included.

The procedures to estimate the uncertainty of the Higgs boson production cross-sections follow the recommendations of the LHC Higgs Cross-Section Working Group [80]. Theoretical uncertainties affecting the ggF signal originate from nine sources [24]. Two sources account for yield uncertainties, which are evaluated by an overall variation of all relevant scales and are correlated across all bins of the BDT discriminant distribution [81]. Another two sources account for migration uncertainties of zero to one jet and one to at least two jets in the event [81–83], two for Higgs boson p\(_T\) shape uncertainties, one for the treatment of the top-quark mass in the loop corrections, and two for the acceptance uncertainties of ggF production in the VBF phase space from selecting exactly two and at least three jets, respectively [84,85]. For VBF and \(W, Z\) production cross-sections, the uncertainties due to missing higher-order QCD corrections are estimated by varying the factorization and renormalization scales up and down by factors of two around the nominal scale. For all signal samples, PDF uncertainties are estimated using 30 eigenvector variations and two \(\alpha_s\) variations using the default PDF set PDF4LHC15 [32]. Uncertainties related to the simulation of the underlying event, hadronization and parton shower are estimated by comparing the acceptances when using Pythia 8.212 [25] or Herwig 7.0.3 [86,87].

The sources of modelling uncertainties considered for the \(Z \rightarrow \tau \tau\) process are the same as in Ref. [17] and their effect on the event migrations between categories and on the shape of the BDT discriminant are considered, since the overall normalizations are determined from data in the statistical analysis. These systematic uncertainties include variations of PDF sets, factorization and renormalization scales, CKKW matching [88], resummation scale and parton shower modelling. The other background processes are either normalized using data (processes with top-quarks and mis-identified leptons and \(t\bar{t}\) candidates) or their cross-section uncertainties have negligible impact and therefore are not included. The shape uncertainties of these backgrounds originate from experimental uncertainties only.

7. Statistical analysis

The searches for \(H \rightarrow e\tau\) and \(H \rightarrow \mu\tau\) are treated independently. For each search, the analysis exploits the four signal regions and the two control regions specified in Table 2. The BDT score distributions of all signal regions are analysed to test the presence of a signal, simultaneously with the event yields in control regions, which are included to constrain the normalizations of the major backgrounds estimated from simulation. The statistical analysis uses a binned likelihood function \(L(\mu, \theta)\), constructed as a product of Poisson probability terms over all bins considered in the search. This function depends on the parameter \(\mu\), defined as the branching ratio \(B(H \rightarrow e\tau)\), and a set of nuisance parameters \(\theta\) that encode the effect of systematic uncertainties in the signal and background expectations. All nuisance parameters are implemented in the likelihood function as Gaussian or log-normal constraints. The normalization factors of the single-top-quark and \(t\bar{t}\) backgrounds in the \(\ell\tau\_\tau\) channel and of the \(Z \rightarrow \tau \tau\) background component are unconstrained parameters of the fit. Estimates of the parameters of interest are calculated with the profile-likelihood-ratio test statistic \(q_{\text{LLR}}\) [89], and the upper limits on the branching ratios are derived by using \(q_{\text{LLR}}\) and the CL\(_S\) method [90].

The discriminant distributions after the fit in each channel are shown in Figs. 2 and 3. Good agreement between data and the background expectation is observed. The event yields after the background-only fit are summarized in Tables 4 and 5. In the non-VBF category, the yields in the \(t\bar{t}\) channel are larger than in the \(\ell\tau\_\tau\) channel due to the looser selection criteria defined for the former channel (Section 4). Table 6 shows a summary of the uncertainties of \(B(H \rightarrow \ell\tau)\). The uncertainties associated with mis-identified leptons and \(t\bar{t}\) candidates and those related to the jet energy scale and resolution exhibit the highest impact on the best-fit branching ratios in both searches. The combined impact from all systematic uncertainties and the data statistics ranges from 0.17% to 0.19%.

8. Results

The best-fit branching ratios and upper limits are computed while assuming \(B(H \rightarrow \mu\tau) = 0\) for the \(H \rightarrow e\tau\) search and \(B(H \rightarrow e\tau) = 0\) for the \(H \rightarrow \mu\tau\) search. The best-fit values of the LFV Higgs boson branching ratios are equal to \((1.5^{+0.18}_{-0.17})\%\) and \((-0.22 \pm 0.19)\%\) for the \(H \rightarrow e\tau\) and \(H \rightarrow \mu\tau\) search, respectively.
Fig. 2. Distributions of the BDT score after the background-signal fit in each signal region of the $\ell \tau$ search, with the LFV signal overlaid, normalized with $B(H \rightarrow \ell \tau) = 1\%$ and enhanced by a factor 10 for visibility. The top and bottom plots display $\sigma_{\text{mis}}$ and $\sigma_{\text{full}}$ BDT scores respectively. The left (right) column corresponds to the non-VBF (VBF) category. The size of the combined statistical, experimental and theoretical uncertainties of the background is indicated by the hatched bands. The binning is shown as in the statistical analysis.

Table 6
Summary of the systematic uncertainties and their impact on the best-fit value of $B$ in the $H \rightarrow \ell \tau$ and $H \rightarrow \mu \tau$ searches. The measured values are obtained by the fit to data, while the expected values are determined by the fit to a background-only sample.

<table>
<thead>
<tr>
<th>Source of uncertainty</th>
<th>Impact on $B(H \rightarrow \ell \tau)$ [%]</th>
<th>Impact on $B(H \rightarrow \mu \tau)$ [%]</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Measured</td>
<td>Expected</td>
</tr>
<tr>
<td>Electron</td>
<td>+0.05 / -0.05</td>
<td>+0.06 / -0.06</td>
</tr>
<tr>
<td>Muon</td>
<td>+0.04 / -0.04</td>
<td>+0.04 / -0.04</td>
</tr>
<tr>
<td>$t_{\text{had+-}}$</td>
<td>+0.02 / -0.02</td>
<td>+0.02 / -0.02</td>
</tr>
<tr>
<td>Jet</td>
<td>+0.05 / -0.08</td>
<td>+0.09 / -0.09</td>
</tr>
<tr>
<td>$E_{\text{miss}}$</td>
<td>+0.02 / -0.02</td>
<td>+0.03 / -0.03</td>
</tr>
<tr>
<td>$b$-tag</td>
<td>+0.02 / -0.03</td>
<td>+0.02 / -0.03</td>
</tr>
<tr>
<td>$t_{\text{had}}$</td>
<td>+0.08 / -0.07</td>
<td>+0.09 / -0.08</td>
</tr>
<tr>
<td>$t_{\text{full}}$</td>
<td>+0.12 / -0.11</td>
<td>+0.11 / -0.12</td>
</tr>
<tr>
<td>Pileup modelling</td>
<td>+0.02 / -0.01</td>
<td>+0.01 / -0.01</td>
</tr>
<tr>
<td>Luminosity</td>
<td>&lt; 0.01</td>
<td>&lt; 0.01</td>
</tr>
<tr>
<td>Background norm.</td>
<td>+0.05 / -0.04</td>
<td>+0.05 / -0.03</td>
</tr>
<tr>
<td>Theor. uncert. (back.)</td>
<td>+0.04 / -0.03</td>
<td>+0.04 / -0.03</td>
</tr>
<tr>
<td>Theor. uncert. (signal)</td>
<td>+0.01 / -0.01</td>
<td>+0.01 / -0.01</td>
</tr>
<tr>
<td>MC statistics</td>
<td>+0.04 / -0.04</td>
<td>+0.03 / -0.03</td>
</tr>
<tr>
<td>Full systematic</td>
<td>+0.17 / -0.16</td>
<td>+0.17 / -0.17</td>
</tr>
<tr>
<td>Data statistics</td>
<td>+0.07 / -0.07</td>
<td>+0.07 / -0.07</td>
</tr>
<tr>
<td>Total</td>
<td>+0.18 / -0.17</td>
<td>+0.18 / -0.18</td>
</tr>
</tbody>
</table>
In the absence of a significant excess, upper limits on the LFV branching ratios are set for a Higgs boson with $m_H = 125$ GeV. The observed (median expected) 95% CL upper limits are 0.47% ($0.34^{+0.13}_{-0.10} \%$) and 0.28% ($0.37^{+0.14}_{-0.10} \%$) for the $H \rightarrow \ell\tau$ and $H \rightarrow \ell\mu$ searches, respectively. These limits are significantly lower than the corresponding Run 1 limits of Refs. [14,15]. The breakdown of contributions from different signal regions is shown in Fig. 4.

The branching ratio of the LFV Higgs boson decay is related to the non-diagonal Yukawa coupling matrix elements [91] by the formula

$$|Y_{\ell\tau}|^2 + |Y_{\ell\mu}|^2 = \frac{8\pi}{m_H} \frac{B(H \rightarrow \ell\tau)}{1 - B(H \rightarrow \ell\ell)} \Gamma(H)(SM),$$

where $\Gamma_{\mu}(SM) = 4.07$ MeV [92] stands for the Higgs boson width as predicted by the Standard Model. Thus, the observed limits on the branching ratio correspond to the following limits on the coupling matrix elements: $|Y_{\ell\tau}|^2 + |Y_{\ell\mu}|^2 < 0.0020$, and $\sqrt{|Y_{e\tau}|^2 + |Y_{e\mu}|^2} < 0.0015$. Fig. 5 shows the limits on the individual coupling matrix elements $Y_{\ell\tau}$ and $Y_{\ell\mu}$ together with the limits from the ATLAS Run 1 analysis and from $\tau \rightarrow \ell\gamma$ searches [91,93].

9. Conclusions

Direct searches for the decays $H \rightarrow \ell\tau$ and $H \rightarrow \ell\mu$ are performed with proton–proton collisions recorded by the ATLAS detector at the LHC corresponding to an integrated luminosity of 36.1 fb$^{-1}$ at a centre-of-mass energy of $\sqrt{s} = 13$ TeV. No significant excess is observed above the expected background from Standard Model processes. The observed (expected) upper limits at 95% confidence level on the branching ratios of $H \rightarrow \ell\tau$ and $H \rightarrow \ell\mu$ are 0.47% ($0.34^{+0.13}_{-0.10} \%$) and 0.28% ($0.37^{+0.14}_{-0.10} \%$), respectively. These limits are more stringent by a factor of 2 (5) than the corresponding limits for the $H \rightarrow \ell\tau$ ($H \rightarrow \ell\mu$) decay determined by ATLAS at $\sqrt{s} = 8$ TeV.

Acknowledgements

We thank CERN for the very successful operation of the LHC, as well as the support staff from our institutions without whom ATLAS could not be operated efficiently.

We acknowledge the support of ANPCyT, Argentina; YerPhI, Armenia; ARC, Australia; BMWFW and FWF, Austria; ANAS, Azer-
Fig. 4. Upper limits at 95% CL on the LFV branching ratios of the Higgs boson, $H \rightarrow e\tau$ (left) and $H \rightarrow \mu\tau$ (right), indicated by solid and dashed lines. Best-fit values of the branching ratios ($\hat{\mu}$) are also given, in %. The limits are computed while assuming that either $B(H \rightarrow e\tau) = 0$ (left) or $B(H \rightarrow \mu\tau) = 0$ (right). First, the results of the fits are shown, when only the data of an individual channel or of an individual category are used; in these cases the signal and control regions from all other channels/categories are removed from the fit. These results are finally compared with the full fit displayed in the last row.

Fig. 5. Upper limits on the absolute value of the couplings $Y_{\ell\ell}$ and $Y_{\ell\ell'}$ together with the limits from the ATLAS Run 1 analysis (light grey line) and the most stringent indirect limits from $\tau \rightarrow e\gamma$ searches (dark purple region). Also indicated are limits corresponding to different branching ratios (0.01%, 0.1%, 1%, 10% and 50%) and the naturalness limit (denoted n.l.) $|Y_{\ell\ell}| \lesssim \frac{\sqrt{2}}{\lambda}$, where $\lambda$ is the vacuum expectation value of the Higgs field.

The crucial computing support from all WLCG partners is acknowledged gratefully, in particular from CERN, the ATLAS Tier-1 facilities at TRIUMF (Canada), NDGF (Denmark, Norway, Sweden), CC-IN2P3 (France), KIT/GridKA (Germany), INFN-CNAF (Italy), NL-T1 (Netherlands), PIC (Spain), ASGC (Taiwan), RAL (UK) and BNL (USA), the Tier-2 facilities worldwide and large non-WLCG resource providers. Major contributors of computing resources are listed in Ref. [94].

References


The ATLAS Collaboration

S. Wenig 36, N. Wermers 24, M.D. Weston 78, M. Wessels 61a, T.D. Weston 20, K. Whalen 131
N.L. Whallon 148, A.M. Wharton 89, A.S. White 105, A. White 8, M.J. White 1, D. Whiteson 171
B.W. Whitmore 89, W. Wiedemann 181, M. Wielers 144, N. Wiesesette 99, C. Wiglesworth 40,
L.A.M. Wiik-Fuchs 52, F. Wilk 100, H.G. Wilkens 36, L.J. Wilkins 93, H.H. Williams 137, S. Williams 32,
C. Willis 106, S. Willcock 102, J.A. Wilson 21, I. Wingert-Seez 3, E. Winkels 156, F. Winkmeier 131,
O.J. Winston 156, B.T. Winter 32, M. Wittgen 153, M. Wobisch 95, A. Wolf 39, T.M.H. Wolf 120, R. Wolff 101,
R.W. Wölker 135, J. Wollrath 52, M.W. Wolter 55, H. Wolters 140a, 140c, V.W.S. Wong 175, N.L. Woods 146,
S.D. Worm 21, B.K. Wosiek 84, K.W. Woźniak 84, K. Wraight 57, S.L. Wu 181, X. Wu 54, Y. Wu 60a,
T.R. Wyatt 100, B.M. Wynne 50, S. Xella 40, Z. Xi 105, L. Xia 178, X. Xiao 105, D. Xu 15a, H. Xu 60a, L. Xu 29,
T. Xu 145, W. Xu 105, Z. Xu 60b, Z. Xu 153, B. Yabsley 157, S. Yacoob 35a, K. Yajima 133, D.P. Yallup 94,
D. Yamaguchi 165, Y. Yamaguchi 165, A. Yamamoto 81, M. Yamatani 163, T. Yamazaki 163, Y. Yamazaki 82,
Z. Yan 55, H.J. Yang 60a, 60d, H.T. Yang 18, S. Yang 77, X. Yang 60b, 8, Y. Yang 163, W-M. Yao 18, Y.C. Yap 46,
K. Yoshihara 154, C.J.S. Young 26, C. Young 153, J. Yu 78, R. Yuan 60b, J. Xue 61a, S.P.Y. Yuen 24,
B. Zabinski 84, G. Zacharias 19, E. Zaffaroni 54, J. Zahreddine 136, A.M. Zaitsev 123,am, T. Zakarevichvili 159b,
N. Zakharchuk 34, S. Zambito 39, D. Zanzi 36, D.R. Zaripov 57, S.V. Zeißen 47, C. Zeitnitz 152,
G. Zhang 15b, H. Zhang 15c, J. Zhang 5, L. Zhang 60a, M. Zhang 173, R. Zhang 24, X. Zhang 60b,
Y. Zhang 15a, 15d, Z. Zhang 63a, Z. Zhang 132, P. Zhao 49, Y. Zhao 60b, Z. Zhao 60a, A. Zhemchugov 79,
Z. Zheng 105, D. Zhong 173, B. Zhou 105, C. Zhou 181, M.S. Zhou 15a, 15d, M. Zhou 155, N. Zhou 60c, Y. Zhou 7,
C.G. Zhu 60b, H.L. Zhu 60a, H. Zhu 15, Y. Zhu 15, Y. Zhu 15b, D. Zhu 15a, K. Zhukov 110,
V. Zhulanov 122, 122a, D. Zieminska 65, N.I. Zimine 79, S. Zimmermann 52, Z. Zinonos 115, M. Ziolkowski 151,
G. Zobernig 181, A. Zoccoli 23b, 23a, K. Zoch 53, T.G. Zorbas 149, R. Zou 37, L. Zwalinski 36
Also at CERN, Geneva; Switzerland.

Also at CPPM, Aix-Marseille Université, CNRS/IN2P3, Marseille; France.

Also at Département de Physique Nucléaire et Corpusculaire, Université de Genève, Genève; Switzerland.

Also at Departament de Física de la Universitat Autònoma de Barcelona, Barcelona; Spain.

Also at Departamento de Física, Instituto Superior Técnico, Universidade de Lisboa, Lisboa; Portugal.

Also at Department of Applied Physics and Astronomy, University of Sharjah, Sharjah; United Arab Emirates.

Also at Department of Financial and Management Engineering, University of the Aegean, Chios; Greece.

Also at Department of Physics and Astronomy, Michigan State University, East Lansing MI; United States of America.

Also at Department of Physics and Astronomy, University of Louisville, Louisville, KY; United States of America.

Also at Department of Physics, California State University, East Bay; United States of America.

Also at Department of Physics, California State University, Fresno; United States of America.

Also at Department of Physics, California State University, Sacramento; United States of America.

Also at Department of Physics, King’s College London, London; United Kingdom.

Also at Department of Physics, St. Petersburg State Polytechnical University, St. Petersburg; Russia.

Also at Department of Physics, Stanford University, Stanford CA; United States of America.

Also at Department of Physics, University of Adelaide, Adelaide; Australia.

Also at Department of Physics, University of Fribourg, Fribourg; Switzerland.

Also at Department of Physics, University of Michigan, Ann Arbor MI; United States of America.

Also at Faculty of Physics, M.V. Lomonosov Moscow State University, Moscow; Russia.

Also at Giresun University, Faculty of Engineering, Giresun; Turkey.

Also at Graduate School of Science, Osaka University, Osaka; Japan.

Also at Hellenic Open University, Patras; Greece.

Also at Instituto Catalana de Recerca i Estudis Avancats, ICREA, Barcelona; Spain.

Also at Institut für Experimentalphysik, Universität Hamburg, Hamburg; Germany.

Also at Institute for Mathematics, Astrophysics and Particle Physics, Radboud University Nijmegen/Nikhef, Nijmegen; Netherlands.

Also at Institute for Nuclear Research and Nuclear Energy (INBNE) of the Bulgarian Academy of Sciences, Sofia; Bulgaria.

Also at Institute for Particle and Nuclear Physics, Wigner Research Centre for Physics, Budapest; Hungary.

Also at Institute of Particle Physics (IPP); Canada.

Also at Institute of Physics, Academia Sinica, Taipei; Taiwan.

Also at Institute of Physics, Azerbaijan Academy of Sciences, Baku; Azerbaijan.

Also at Institute of Theoretical Physics, Ilia State University, Tbilisi; Georgia.

Also at Instituto de Física Teórica, IFT-UAM/CSIC, Madrid; Spain.

Also at Istanbul University, Dept. of Physics, Istanbul; Turkey.

Also at Joint Institute for Nuclear Research, Dubna; Russia.

Also at LAL, Université Paris-Sud, CNRS/IN2P3, Université Paris-Saclay, Orsay; France.

Also at Louisiana Tech University, Ruston LA; United States of America.

Also at LPNHE, Sorbonne Université, Paris Diderot Sorbonne Paris Cité, CNRS/IN2P3, Paris; France.

Also at Manhattan College, New York NY; United States of America.

Also at Moscow Institute of Physics and Technology State University, Dolgoprudny; Russia.

Also at National Research Nuclear University MEPhI, Moscow; Russia.

Also at Physics Department, An-Najah National University, Nablus; Palestine.

Also at Physics Dept, University of South Africa, Pretoria; South Africa.

Also at Physikalisches Institut, Albert-Ludwigs-Universität Freiburg, Freiburg; Germany.

Also at School of Physics, Sun Yat-sen University, Guangzhou; China.

Also at The City College of New York, New York NY; United States of America.

Also at The Collaborative Innovation Center of Quantum Matter (CICQM), Beijing; China.

Also at Tomsk State University, Tomsk, and Moscow Institute of Physics and Technology State University, Dolgoprudny; Russia.

Also at TRIUMF, Vancouver BC; Canada.

Also at Università di Napoli Parthenope, Napoli; Italy.

* Deceased.