Measurement of Higgs boson decay into b-quarks in associated production with a top-quark pair in pp collisions at $\sqrt{s} = 13$ TeV with the ATLAS detector

The ATLAS collaboration

DOI
10.1007/JHEP06(2022)097

Publication date
2022

Document Version
Final published version

Published in
Journal of High Energy Physics

License
CC BY

Citation for published version (APA):
https://doi.org/10.1007/JHEP06(2022)097

General rights
It is not permitted to download or to forward/distribute the text or part of it without the consent of the author(s) and/or copyright holder(s), other than for strictly personal, individual use, unless the work is under an open content license (like Creative Commons).

Disclaimer/Complaints regulations
If you believe that digital publication of certain material infringes any of your rights or (privacy) interests, please let the Library know, stating your reasons. In case of a legitimate complaint, the Library will make the material inaccessible and/or remove it from the website. Please Ask the Library: https://uba.uva.nl/en/contact, or a letter to: Library of the University of Amsterdam, Secretariat, Singel 425, 1012 WP Amsterdam, The Netherlands. You will be contacted as soon as possible.

UvA-DARE is a service provided by the library of the University of Amsterdam (https://dare.uva.nl)

Download date:16 Oct 2023
Measurement of Higgs boson decay into $b$-quarks in associated production with a top-quark pair in $pp$ collisions at $\sqrt{s} = 13$ TeV with the ATLAS detector

The ATLAS collaboration

E-mail: atlas.publications@cern.ch

ABSTRACT: The associated production of a Higgs boson and a top-quark pair is measured in events characterised by the presence of one or two electrons or muons. The Higgs boson decay into a $b$-quark pair is used. The analysed data, corresponding to an integrated luminosity of 139 fb$^{-1}$, were collected in proton-proton collisions at the Large Hadron Collider between 2015 and 2018 at a centre-of-mass energy of $\sqrt{s} = 13$ TeV. The measured signal strength, defined as the ratio of the measured signal yield to that predicted by the Standard Model, is $0.35^{+0.36}_{-0.34}$. This result is compatible with the Standard Model prediction and corresponds to an observed (expected) significance of 1.0 (2.7) standard deviations. The signal strength is also measured differentially in bins of the Higgs boson transverse momentum in the simplified template cross-section framework, including a bin for specially selected boosted Higgs bosons with transverse momentum above 300 GeV.

KEYWORDS: Hadron-Hadron Scattering

ArXiv ePrint: 2111.06712
1 Introduction

The Higgs boson [1–3] was discovered by the ATLAS and CMS collaborations [4, 5] in 2012, with a mass of around 125 GeV [6]. Since then, the analysis of proton-proton (pp) collision data at centre-of-mass energies of 7 TeV, 8 TeV and 13 TeV delivered by the Large Hadron Collider (LHC) [7] has led to more detailed measurements of its properties and couplings, testing the predictions of the Standard Model (SM). Of particular interest is the Yukawa coupling to the top quark, the heaviest elementary particle in the SM, which could be very sensitive to effects of physics beyond the SM (BSM) [8].

Both the ATLAS and CMS collaborations have observed the interaction of the Higgs boson with third-generation fermions of the SM. The coupling to \( \tau \)-leptons was measured in the observation of \( H \rightarrow \tau \tau \) decays [9–11], while the observation of the decay of the Higgs boson into \( b \)-quark pairs provided direct evidence for the Yukawa coupling to down-type quarks [12, 13]. The interaction of the Higgs boson with the top quark was indirectly constrained (assuming no BSM phenomena in the loops) from measurements of gluon-gluon fusion Higgs production and decay to \( \gamma \gamma \) [9], before being directly measured in the observation of Higgs boson production in association with a pair of top quarks (\( t\bar{t}H \)) [14, 15].
Figure 1. Representative tree-level Feynman diagrams for the production of a Higgs boson in association with a top-quark pair ($t\bar{t}H$) in (a) the t-channel and (b) the s-channel and the subsequent decay of the Higgs boson into $b\bar{b}$, and for (c) the main $t\bar{t} + b\bar{b}$ background.

The $t\bar{t}H$ production mode, in which the top quark couples to the Higgs boson at tree level, is the most favourable to extract direct information on the top-quark Yukawa coupling without assumptions about the potential presence of BSM physics [16–19].

Although this production mode contributes only around 1% of the total Higgs boson production cross-section at the LHC [20], the top quarks in the final state offer a distinctive signature and allow access to many Higgs boson decay modes. The decay into two $b$-quarks ($H \rightarrow b\bar{b}$) is the most probable, with a SM branching fraction of about 58% [20]. Furthermore, in the $H \rightarrow b\bar{b}$ decay mode the reconstruction of the Higgs boson kinematics is possible, which allows the extraction of additional information about the structure of the top-Higgs interaction [21–24]. This analysis therefore aims to select events with a Higgs boson produced in association with a pair of top quarks and decaying into a pair of $b$-quarks ($t\bar{t}H(b\bar{b})$), in which one or both top quarks decay semileptonically, producing an electron or a muon, collectively referred to as leptons ($\ell$). With many final-state particles, the main challenges are the low efficiency to reconstruct and identify all of them, the large combinatorial ambiguities when trying to match the observed objects to the decay products of the Higgs boson and top quarks, and the large background of $t\bar{t} +$ jets processes (in particular when these jets originate from $b$- or $c$-quarks), which have a much larger production cross-section than the signal. Representative Feynman diagrams for the $t\bar{t}H(b\bar{b})$ signal and dominant $t\bar{t} + b\bar{b}$ background are shown in figure 1.

The ATLAS Collaboration searched for $t\bar{t}H(b\bar{b})$ production at $\sqrt{s} = 8$ TeV in final states with at least one [25] or no lepton [26], and at $\sqrt{s} = 13$ TeV with at least one lepton in the final state with data collected in 2015 and 2016, corresponding to an integrated luminosity of 36.1 fb$^{-1}$ [27]. A combined signal strength (defined as the ratio of the measured signal yield to that predicted by the Standard Model) of $0.84^{+0.64}_{-0.61}$ was measured, with an observed (expected) significance of 1.4 (1.6) standard deviations. This result was combined with analyses of Higgs boson decays into massive vector bosons, $\tau$-leptons, or photons to reach observation of the $t\bar{t}H$ production mode [14]. The CMS Collaboration searched for the same processes using 35.9 fb$^{-1}$ of data collected at $\sqrt{s} = 13$ TeV in 2016, in events with at least one lepton [28] or no lepton [29], and measured a signal strength of $0.72 \pm 0.45$ and $0.9 \pm 1.5$, respectively. These results also contributed to the observation of the $t\bar{t}H$ production mode [15].

In this paper, a measurement of the $t\bar{t}H$ production cross-section in the $H \rightarrow b\bar{b}$ channel is performed using the LHC Run 2 pp collision data collected by the ATLAS detector, \[1^\text{Electrons and muons from the decay of a $\tau$-lepton itself originating from a $W$ boson are included.}\]
corresponding to an integrated luminosity of 139 fb$^{-1}$ at $\sqrt{s} = 13$ TeV. Events with either one lepton or two leptons are analysed separately in exclusive single-lepton or dilepton categories depending on the number of leptons, the number of jets and the number of jets identified as originating from $b$-hadrons ($b$-jets). In the single-lepton channel (but not in the dilepton channel because the expected number of events is small), a specific category, referred to as ‘boosted’ in the following, is designed to select events in which the Higgs boson is produced with high transverse momentum ($p_T$). The non-boosted categories, where the Higgs boson decay products are less collimated, are referred to as ‘resolved’. Machine-learning algorithms are used to classify events into signal-rich categories, which are analysed together with the signal-depleted ones in a combined profile likelihood fit. The output distributions of these multivariate algorithms are used as the main discriminant to extract the signal. This signal extraction fit simultaneously determines the event yields for the signal and for the most important background components, while constraining the overall background model within the assigned systematic uncertainties. In addition, making use of the possibility to reconstruct the Higgs boson kinematics in the $H \rightarrow b \bar{b}$ channel, the cross-section is measured as a function of the Higgs boson ‘truth’ transverse momentum $\hat{p}_T^H$, as obtained from Monte Carlo simulation, in the simplified template cross-sections (STXS) formalism [20], which aims to reduce the theory dependencies that are folded into the measurements.

Several aspects of the analysis have been improved relative to the previous version [27]. The full Run 2 dataset is analysed, with improved reconstruction algorithms and detector calibrations, which in turn enhance the performance of the $b$-tagging algorithm. This allows the use of tighter selection criteria to reject events in poorly modelled regions of phase space, and to define analysis regions differential in Higgs boson $p_T$. An enhanced model is adopted for the $t\bar{t} +$ jets background, with updated event generator versions, a higher-order precision prediction for the $t\bar{t} + b\bar{b}$ process, an increased number of simulated events, and an improved assessment of uncertainties. Finally, a new deep neural network improves the performance of the boosted channel.

This paper is organised as follows. The ATLAS detector is described in section 2. The signal and background modelling is presented in section 3. Section 4 summarises the selection criteria for reconstructed objects and events, while section 5 describes the analysis strategy. Systematic uncertainties are discussed in section 6. Results are presented in section 7, and the conclusions in section 8.

2 ATLAS detector

The ATLAS experiment [30–33] at the LHC is a multipurpose particle detector with a forward-backward symmetric cylindrical geometry and a near $4\pi$ coverage in solid angle.$^2$

---

$^2$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 ($r, \phi$) are used in the transverse plane, $\phi$ being the azimuthal angle around the z-axis. The rapidity is defined as $y = (1/2) \ln[(E + p_z)/(E - p_z)]$ where $E$ is the energy and $p_z$ is the longitudinal component of the momentum along the beam pipe. The pseudorapidity is defined in terms of the polar angle $\theta$ as $\eta = -\ln \tan(\theta/2)$. Angular distance is measured in units of $\Delta R \equiv \sqrt{(\Delta \eta)^2 + (\Delta \phi)^2}$. 

---
It consists of an inner tracking detector surrounded by a thin superconducting solenoid providing a 2 T axial magnetic field, electromagnetic and hadron calorimeters, and a muon spectrometer. The inner tracking detector (ID) covers the pseudorapidity range $|\eta| < 2.5$. It consists of silicon pixel, silicon microstrip, and transition radiation tracking detectors. Lead/liquid-argon (LAr) sampling calorimeters provide electromagnetic (EM) energy measurements with high granularity. A steel/scintillator-tile hadron calorimeter covers the central pseudorapidity range $|\eta| < 1.7$. The endcap and forward regions are instrumented with LAr calorimeters for both the EM and hadronic energy measurements up to $|\eta| = 4.9$. The muon spectrometer surrounds the calorimeters and is based on three large air-core toroidal superconducting magnets with eight coils each. The field integral of the toroids ranges between 2 and 6 T m across most of the detector. The inner tracking detector includes a system of precision chambers for tracking and fast detectors for triggering. A two-level trigger system is used to select events. The first-level trigger is implemented in hardware and uses a subset of the detector information to accept events at a rate below 100 kHz. This is followed by a software-based trigger that reduces the accepted event rate to 1 kHz on average depending on the data-taking conditions [34]. An extensive software suite [35] is used 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 Signal and background modelling

Simulated event samples are used to model the $t\bar{t}H$ signal and the background processes. The numbers of selected events and the distribution shapes of variables used to discriminate between signal and background enter into the signal extraction fit, which also constrains the modelling of the background processes. The Monte Carlo (MC) samples were produced using either the full ATLAS detector simulation [36] based on GEANT4 [37] or a faster simulation where the full GEANT4 simulation of the calorimeter response is replaced by a detailed parameterisation of the shower shapes [36]. For the observables used in the analysis, the two simulations were found to give similar modelling. To simulate the effects of multiple interactions in the same and neighbouring bunch crossings (pile-up), additional interactions were generated using PYTHIA 8.186 [38] with a set of tuned parameters called the A3 tune [39] and overlaid onto the simulated hard-scatter event. Simulated events are reweighted to match the pile-up conditions observed in the full Run 2 dataset, with a mean number of $pp$ interactions per bunch crossing of 34. All simulated events are processed through the same reconstruction algorithms and analysis chain as the data.

In all samples where the parton shower (PS), hadronisation, and multi-parton interactions (MPI) were generated with either PYTHIA 8 or HERWIG 7 [40, 41], the decays of $b$- and $c$-hadrons were simulated using the EVTGEN 1.6.0 program [42]. The $b$-quark mass was set to $m_b = 4.80 \text{ GeV}$ (4.50 GeV) for samples using PYTHIA 8 (HERWIG 7). For PYTHIA 8, the A14 tune [43] and the NNPDF2.3lo parton distribution function (PDF) set [44] were used. For HERWIG 7, the H7UE tune [41] was used with the MMHT2014lo PDF set [45]. The Higgs boson mass was set to $m_H = 125.0 \text{ GeV}$, and the top-quark mass to $m_t = 172.5 \text{ GeV}$. The precision of the matrix element (ME) generators is next-to-leading order (NLO) in
quantum chromodynamics (QCD) for most samples. Some samples are normalised to higher precision in QCD (next-to-next-to-leading order, NNLO, or next-to-next-to-leading logarithm, NNLL) or with electroweak (EW) corrections. A summary of all generated samples is presented in table 1, which includes both the samples used for nominal predictions and other samples used to assess systematic uncertainties. Further details are provided in the following subsections.

3.1 Signal modelling

In \( t\bar{t}H \) events the production and decays were modelled in the five-flavour scheme using the Powheg Box [61–65] generator at NLO in QCD with the NNPDF3.0nlo [66] PDF set. The \( h_{\text{damp}} \) parameter\(^3\) was set to 0.75 \times (m_t+m_t+m_H) = 352.5 \text{ GeV}, and the functional form of the renormalisation and factorisation scales were both set to \( \sqrt{m_T(t) \cdot m_T(\bar{t}) \cdot m_T(H)} \) (where \( m_T = \sqrt{m^2 + p_T^2} \) is the transverse mass of a particle). The events were showered by PYTHIA 8 and all Higgs boson decay modes are considered. The samples are normalised to the fixed-order cross-section calculation, \( \sigma_{t\bar{t}H} = 507^{+35}_{-50} \text{ fb} \), which includes NLO QCD and EW corrections [20] for a Higgs boson mass of 125 GeV.

3.2 \( t\bar{t} + \) jets background

Simulated \( t\bar{t} + \) jets events are categorised according to the flavour of additional jets in the event, using the procedure described in ref. [25]. For this purpose, jets are reconstructed from stable particles (mean lifetime \( \tau > 3 \times 10^{-11} \text{ s} \)) using the anti-\( k_t \) algorithm [67, 68], and the number of \( b\)- or \( c\)-hadrons within \( \Delta R = 0.4 \) of the jet axis is considered (with \( p_T > 5 \text{ GeV} \) for the leading \( b\)- or \( c\)-hadron around the jet), excluding particles produced by the top-quark decay. Events are labelled as \( t\bar{t} + \geq 1b \) if at least one \( b\)-flavour jet is identified, or else as \( t\bar{t} + \geq 1c \) if at least one \( c\)-flavour jet is identified, and otherwise as \( t\bar{t} + \) light. Where necessary, the \( t\bar{t} + \geq 1b \) events are further separated into \( t\bar{t} + 1b \) (where exactly one jet is matched to at least one \( b\)-hadron) and \( t\bar{t} + \geq 2b \) (all remaining events).

To accurately model the dominant \( t\bar{t} + \geq 1b \) background, a sample with \( t\bar{t} + bb \) MEs was produced at NLO QCD accuracy in the four-flavour scheme with the Powheg Box RES [69] generator and OpenLoops [70, 71], using a pre-release of the implementation of this process in Powheg Box RES provided by the authors [72], with the NNPDF3.0nlo nf4 [66] PDF set, and using PYTHIA 8 for the PS and hadronisation. The factorisation scale was set to 0.5 \times \sum_{i=t,\bar{t},b,\bar{b}} m_T(i) \) (where \( j \) stands for extra partons), the renormalisation scale was set to \( \sqrt{m_T(t) \cdot m_T(\bar{t}) \cdot m_T(b) \cdot m_T(\bar{b})} \), and the \( h_{\text{damp}} \) parameter was set to 0.5 \times \sum_{i=t,\bar{t},b,\bar{b}} m_T(i) \). The mass of the two \( b\)-quarks produced in the ME in association with the two top quarks was set to the same value as the mass of the \( b\)-quarks from the top-quark decays, \( m_b = 4.95 \text{ GeV} \).

Inclusive \( t\bar{t} + \) jets events were also generated with \( t\bar{t} \) MEs in the five-flavour scheme using the Powheg Box v2 generator at NLO in QCD, using PYTHIA 8 for the PS and hadronisation. Here, the \( h_{\text{damp}} \) parameter was set to 1.5 \( m_t \) [60], and the functional form of

\(^3\)The \( h_{\text{damp}} \) parameter controls the \( p_T \) of the first additional emission beyond the leading-order Feynman diagram in the PS and therefore regulates the high-\( p_T \) emission against which the \( t\bar{t}H \) system recoils.
<table>
<thead>
<tr>
<th>Process</th>
<th>ME generator</th>
<th>ME PDF</th>
<th>PS</th>
<th>Normalisation</th>
</tr>
</thead>
<tbody>
<tr>
<td>Higgs boson</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>$t\bar{t}H$</td>
<td>Powheg Box v2</td>
<td>NNPDF3.0nlo</td>
<td>Pythia 8.230</td>
<td>NLO+NLO (EW)</td>
</tr>
<tr>
<td></td>
<td>Powheg Box v2</td>
<td>NNPDF3.0nlo</td>
<td>Herwig 7.04</td>
<td>NLO+NLO (EW)</td>
</tr>
<tr>
<td></td>
<td>MadGraph5_aMC@NLO</td>
<td>NNPDF3.0nlo</td>
<td>Pythia 8.230</td>
<td>NLO+NLO (EW)</td>
</tr>
<tr>
<td>$tHjb$</td>
<td>MadGraph5_aMC@NLO</td>
<td>NNPDF3.0nlo</td>
<td>Pythia 8.230</td>
<td></td>
</tr>
<tr>
<td>$tWH$</td>
<td>MadGraph5_aMC@NLO</td>
<td>NNPDF3.0nlo</td>
<td>Pythia 8.235</td>
<td></td>
</tr>
<tr>
<td>$t\bar{t}+\text{jets}$ and single-top</td>
<td>Powheg Box v2</td>
<td>NNPDF3.0nlo</td>
<td>Pythia 8.230</td>
<td>NLO+NLL [46–52]</td>
</tr>
<tr>
<td></td>
<td>Powheg Box v2</td>
<td>NNPDF3.0nlo</td>
<td>Herwig 7.04</td>
<td>NLO+NLL [46–52]</td>
</tr>
<tr>
<td></td>
<td>MadGraph5_aMC@NLO</td>
<td>NNPDF3.0nlo</td>
<td>Pythia 8.230</td>
<td>NLO+NLL [46–52]</td>
</tr>
<tr>
<td></td>
<td>MadGraph5_aMC@NLO</td>
<td>NNPDF3.0nlo</td>
<td>Pythia 8.230</td>
<td>NLO+NLL [46–52]</td>
</tr>
<tr>
<td></td>
<td>MadGraph5_aMC@NLO</td>
<td>NNPDF3.0nlo</td>
<td>Pythia 8.230</td>
<td>NLO+NLL [46–52]</td>
</tr>
<tr>
<td></td>
<td>MadGraph5_aMC@NLO</td>
<td>NNPDF3.0nlo</td>
<td>Pythia 8.230</td>
<td>NLO+NLL [46–52]</td>
</tr>
<tr>
<td>s-channel</td>
<td>Powheg Box v2</td>
<td>NNPDF3.0nlo</td>
<td>Pythia 8.230</td>
<td>NLO+NLL [46–52]</td>
</tr>
<tr>
<td></td>
<td>Powheg Box v2</td>
<td>NNPDF3.0nlo</td>
<td>Herwig 7.04</td>
<td>NLO+NLL [46–52]</td>
</tr>
<tr>
<td></td>
<td>MadGraph5_aMC@NLO</td>
<td>NNPDF3.0nlo</td>
<td>Pythia 8.230</td>
<td>NLO+NLL [46–52]</td>
</tr>
<tr>
<td></td>
<td>MadGraph5_aMC@NLO</td>
<td>NNPDF3.0nlo</td>
<td>Pythia 8.230</td>
<td>NLO+NLL [46–52]</td>
</tr>
<tr>
<td></td>
<td>MadGraph5_aMC@NLO</td>
<td>NNPDF3.0nlo</td>
<td>Pythia 8.230</td>
<td>NLO+NLL [46–52]</td>
</tr>
<tr>
<td>Other</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>$W+\text{jets}$</td>
<td>Sherpa 2.2.1</td>
<td>NNPDF3.0nlo</td>
<td>Sherpa</td>
<td>NNLO [57]</td>
</tr>
<tr>
<td></td>
<td>Sherpa 2.2.1</td>
<td>NNPDF3.0nlo</td>
<td>Sherpa</td>
<td>NNLO [57]</td>
</tr>
<tr>
<td>$Z+\text{jets}$</td>
<td>Sherpa 2.2.1</td>
<td>NNPDF3.0nlo</td>
<td>Sherpa</td>
<td>NNLO [57]</td>
</tr>
<tr>
<td>$VV$ (had.)</td>
<td>Sherpa 2.2.1</td>
<td>NNPDF3.0nlo</td>
<td>Sherpa</td>
<td></td>
</tr>
<tr>
<td>$VV$ (lep.)</td>
<td>Sherpa 2.2.2</td>
<td>NNPDF3.0nlo</td>
<td>Sherpa</td>
<td></td>
</tr>
<tr>
<td>$VV$ (lep.) + jj</td>
<td>Sherpa 2.2.2</td>
<td>NNPDF3.0nlo</td>
<td>Sherpa</td>
<td></td>
</tr>
<tr>
<td>$t\bar{t}W$</td>
<td>MadGraph5_aMC@NLO</td>
<td>NNPDF3.0nlo</td>
<td>Pythia 8.210</td>
<td>NLO+NLO (EW)</td>
</tr>
<tr>
<td></td>
<td>Sherpa 2.0.0</td>
<td>NNPDF3.0nlo</td>
<td>Sherpa</td>
<td>NLO+NLO (EW)</td>
</tr>
<tr>
<td></td>
<td>Sherpa 2.0.0</td>
<td>NNPDF3.0nlo</td>
<td>Sherpa</td>
<td>NLO+NLO (EW)</td>
</tr>
<tr>
<td>$t\bar{t}Z$ (qq, vv)</td>
<td>MadGraph5_aMC@NLO</td>
<td>NNPDF3.0nlo</td>
<td>Pythia 8.210</td>
<td>NLO+NLO (EW)</td>
</tr>
<tr>
<td></td>
<td>Sherpa 2.0.0</td>
<td>NNPDF3.0nlo</td>
<td>Sherpa</td>
<td>NLO+NLO (EW)</td>
</tr>
<tr>
<td></td>
<td>Sherpa 2.0.0</td>
<td>NNPDF3.0nlo</td>
<td>Sherpa</td>
<td>NLO+NLO (EW)</td>
</tr>
<tr>
<td>$t\bar{t}q$</td>
<td>MadGraph5_aMC@NLO</td>
<td>NNPDF3.0nlo</td>
<td>Pythia 8.212</td>
<td></td>
</tr>
<tr>
<td>$t\bar{t}Z$</td>
<td>MadGraph5_aMC@NLO</td>
<td>NNPDF3.0nlo</td>
<td>Pythia 8.230</td>
<td>NLO+NLO (EW)</td>
</tr>
</tbody>
</table>

Table 1. Table summarising the generator set-ups for samples used in this analysis. The first row for each sample details the nominal settings used for this process in the analysis. Any additional rows describe samples which are used to evaluate the modelling and performance of the analysis. For overlap between $t\bar{t}$ and $tW$-like diagrams, samples using the diagram removal scheme [59] are listed as [DR] and samples using the diagram subtraction scheme [59, 60] are listed as [DS]. The precision of the ME generator is NLO in QCD if no additional information is provided in parentheses. The higher-order cross-section used to normalise these samples is listed in the last column and refers to the order of QCD processes if no additional information is provided. If no information is present in this column, no higher-order $K$-factor is applied to this process. For the $VV$ Sherpa samples, ‘lep.’ (‘had.’) means that both bosons decay leptonically (one decays leptonically and one hadronically).
the renormalisation and factorisation scales was set to $m_T(t)$.$^4$ The $t\bar{t} + \geq 1c$ and $t\bar{t} + \text{light}$ events using this prediction are combined with the previously described $t\bar{t} + \geq 1b$ sample to form the nominal $t\bar{t}$ model, while the $t\bar{t} + \geq 1b$ events from this five-flavour scheme are used only to assign modelling uncertainties.

### 3.3 Other backgrounds

The QCD $V + \text{jets}$ processes (i.e. $W + \text{jets}$ and $Z + \text{jets}$) were simulated with the SHERPA 2.2.1 generator [73]. In this set-up, NLO-accurate MEs for up to two partons, and leading-order (LO) accurate MEs for up to four partons were calculated with the Comix [74] and OPENLOOPS libraries. They were matched with the SHERPA parton shower [75] by using the MEPS@NLO prescription [76–79] with the set of tuned parameters developed by the SHERPA authors and based on the NNPDF3.0nnlo set of PDFs [66]. A data-driven correction to the normalisation was derived for $Z + \text{jets}$ predictions containing at least two heavy-flavour jets (where heavy-flavour means jets originating from $b$-hadrons and $c$-hadrons). Events are selected, with objects passing the selection discussed in section 4, in dedicated control regions in data defined by requiring at least two jets and two opposite-electric-charge same-flavour leptons ($e^+e^-$ or $\mu^+\mu^-$) with an invariant mass inside the $Z$-boson mass window 83–99 GeV. A 25\% yield increase is found to improve the modelling of $Z + \text{jets}$ events with at least two heavy-flavour jets, consistent with ref. [80].

Diboson events were simulated with the SHERPA 2.2.1 and 2.2.2 generators. In this set-up multiple MEs were matched and merged with the SHERPA PS based on Catani-Seymour dipole factorisation [74, 75] using the MEPS@NLO prescription. For semileptonically and fully leptonically decaying diboson event samples, as well as loop-induced diboson event samples, the virtual QCD correction for MEs at NLO accuracy was provided by the OPENLOOPS library. For EW $VVjj$ production, the calculation was performed in the $G_\mu$ scheme, ensuring an optimal description of pure EW interactions at the EW scale [81–83]. All samples were generated using the NNPDF3.0nnlo PDF set, along with the dedicated set of tuned PS parameters developed by the SHERPA authors.

The associated production of a single top quark and a Higgs boson is treated as background. Samples for two subprocesses, $tHjb$ and $tWH$ production, were generated using the MADGRAPH5_AMC@NLO 2.6.2 generator at NLO in QCD. The functional form of the renormalisation and factorisation scale was set to $0.5 \times \sum_i m_T(i)$, where the sum runs over all the particles generated from the ME calculation. For $tHjb$ ($tWH$), events were generated in the four-flavour (five-flavour) scheme using the NNPDF3.0nlo nf4 (NNPDF3.0nlo) PDF set, and the diagram removal scheme was employed to handle the interference with $t\bar{t}H$ in the $tWH$ sample [59, 84].

Single-top $t$-channel, $s$-channel, and $tW$ production were modelled using the POWHEG BOX v2 [85–87] generator at NLO in QCD. For $t$-channel production, events were generated in the four-flavour scheme with the NNPDF3.0nlo nf4 PDF set, and the functional form of the renormalisation and factorisation scales was set to $m_T(b)$ following the recommendation of ref. [85]. For $s$-channel and $tW$ production, events were generated

---

$^4$This scale is calculated in the $t\bar{t}$ rest-frame and hence the $p_T$ values of the top quark and top antiquark are equivalent.
in the five-flavour scheme with the NNPDF3.0nlo PDF set, and the functional form of the renormalisation and factorisation scales was set to the top-quark mass. For $tW$ production, the diagram removal scheme [59] was employed to handle the interference with $t\bar{t}$ production [60]. An additional sample which applies the diagram subtraction scheme [59, 60] was used to assess an uncertainty in the modelling of this interference.

The production and decay of a top-quark pair in association with a vector boson (i.e. $t\bar{t}W$ or $t\bar{t}Z$), referred to collectively as $t\bar{t}V$, was modelled using the MadGraph5\_aMC@NLO 2.3.3 generator at NLO in QCD. The functional form of the renormalisation and factorisation scale was set to $0.5 \times \sum_i m_T(i)$, where the sum runs over all the particles generated from the ME calculation.

For events with four top quarks ($t\bar{t}tt$) the production and decays were modelled using the MadGraph5\_aMC@NLO 2.3.3 generator at NLO in QCD with the NNPDF3.1nlo PDF set. The functional form of the renormalisation and factorisation scales was set to $0.25 \times \sum_i m_T(i)$, where the sum runs over all the particles generated from the ME calculation, following the recommendation of ref. [58].

The $tZq$ events were generated using the MadGraph5\_aMC@NLO 2.3.3 generator in the four-flavour scheme at LO in QCD, with the CTEQ6L1 [88] PDF set. Following the recommendations of ref. [85], the renormalisation and factorisation scales were set to $4 \times m_T(b)$, where the $b$-quark is the one coming from the gluon splitting.

The $tWZ$ sample was produced using the MadGraph5\_aMC@NLO 2.3.3 generator at NLO in QCD with the NNPDF3.0nlo PDF set. The renormalisation and factorisation scales were set to the top-quark mass. The diagram removal scheme was employed to handle the interference between $tWZ$ and $t\bar{t}Z$.

In the single-lepton channel, a negligible contribution is found from events with a jet or a photon misidentified as a lepton, or from events with a non-prompt lepton. In the dilepton channel, the contribution from events with non-prompt leptons is estimated from simulated events (from $t\bar{t}$, $t\bar{t}Z$, $t\bar{t}W$, diboson, $Z$+jets, $W$+jets, $tW$, s-channel and t-channel single top production) where at least one of the reconstructed leptons does not match a true lepton in the event record. These events are collectively referred to as fakes.

4 Object and event selection

Events are selected from $pp$ collisions at $\sqrt{s} = 13$ TeV recorded by the ATLAS detector between 2015 and 2018, corresponding to an integrated luminosity of 139 fb$^{-1}$. Only events for which the LHC beams were in stable-collision mode and all relevant subsystems were operational are considered [89]. Events are required to have at least one primary vertex with two or more tracks with $p_T > 0.5$ GeV. If more than one vertex is found, the hard-scattering primary vertex is selected as the one with the highest sum of squared transverse momenta of associated tracks [90].

Events were recorded using single-lepton triggers with either a low $p_T$ threshold and a lepton isolation requirement, or a higher threshold but a looser identification criterion and without any isolation requirement. The lowest $p_T$ threshold in the single-muon trigger was 20 (26) GeV [91] for data taken in 2015 (2016–2018), while in the single-electron trigger it was 24 (26) GeV [92].
Electrons are reconstructed from tracks in the ID associated with topological clusters of energy depositions in the calorimeter [93] and are required to have $p_T > 10 \text{ GeV}$ and $|\eta| < 2.47$. Candidates in the calorimeter barrel-endcap transition region ($1.37 < |\eta| < 1.52$) are excluded. Electrons must satisfy the *Medium* likelihood identification criterion. Muon candidates are identified by matching ID tracks to full tracks or track segments reconstructed in the muon spectrometer, using the *Loose* identification criterion [94]. Muons are required to have $p_T > 10 \text{ GeV}$ and $|\eta| < 2.5$. Lepton tracks must match the primary vertex of the event, i.e. they have to satisfy $|z_0 \sin(\theta)| < 0.5 \text{ mm}$ and $|d_0/\sigma(d_0)| < 5 (3)$ for electrons (muons), where $z_0$ is the longitudinal impact parameter relative to the primary vertex and $d_0$ (with uncertainty $\sigma(d_0)$) is the transverse impact parameter relative to the beam line.

Jets are reconstructed from noise-suppressed topological clusters of calorimeter energy depositions [95] calibrated at the electromagnetic scale [96], using the anti-$k_t$ algorithm with a radius parameter of 0.4. These are referred to as small-$R$ jets. The average energy contribution from pile-up is subtracted according to the jet area and jets are calibrated as described in ref. [96] with a series of simulation-based corrections and in situ techniques. Jets are required to satisfy $p_T > 25 \text{ GeV}$ and $|\eta| < 2.5$. The effect of pile-up is reduced by an algorithm requiring that the calorimeter-based jets are consistent with originating from the primary vertex using tracking information [97].

Jets containing $b$-hadrons are identified ($b$-tagged) with the MV2c10 multivariate algorithm [98], which combines information about the transverse impact parameters of displaced tracks and the topological properties of secondary and tertiary decay vertices reconstructed within the jet. Four working points, defined by different thresholds for the MV2c10 discriminant, are used in this analysis, corresponding to efficiencies ranging from 85% to 60% for $b$-jets with $p_T > 20 \text{ GeV}$ as determined in simulated $t\bar{t}$ events. The corresponding rejection rates are in the range 2–22 for $c$-jets (containing $c$-hadrons and no $b$-hadrons) and 27–1150 for light-jets. Jets are then assigned a pseudo-continuous $b$-tagging value [98] according to the tightest working point they satisfy. Correction factors are applied to the simulated events to compensate for differences between data and simulation in the $b$-tagging efficiency for $b$-, $c$-, and light-jets [98–100].

Hadronically decaying $\tau$-leptons ($\tau_{\text{had}}$) are distinguished from jets using their track multiplicity and a multivariate discriminant based on calorimetric shower shapes and tracking information [101]. They are required to have $p_T > 25 \text{ GeV}$ and $|\eta| < 2.5$, and to pass the *Medium* $\tau$-identification working point.

An overlap removal procedure is applied to prevent double-counting of objects. The closest jet within $\Delta R_y = \sqrt{(\Delta y)^2 + (\Delta \phi)^2} = 0.2$ of a selected electron is removed. If the nearest jet surviving that selection is within $\Delta R_y = 0.4$ of the electron, the electron is discarded. Muons are usually removed if they are separated from the nearest jet by $\Delta R_y < 0.4$, since this reduces the background from heavy-flavour decays inside jets. However, if this jet has fewer than three associated tracks, the muon is kept and the jet is removed instead; this avoids an inefficiency for high-energy muons undergoing significant energy loss in the calorimeter. A $\tau_{\text{had}}$ candidate is rejected if it is separated by $\Delta R_y < 0.2$ from any selected electron or muon. No overlap removal is performed between jets and $\tau_{\text{had}}$ candidates.
The missing transverse momentum (with magnitude $E_{\text{T}}^{\text{miss}}$) is reconstructed as the negative vector sum of the $p_T$ of all the selected electrons, muons, $\tau_{\text{had}}$ and jets described above, with an extra ‘soft term’ built from additional tracks associated with the primary vertex, to make it resilient to pile-up contamination [102]. The missing transverse momentum is not used for event selection but is included in the inputs to the multivariate discriminants that are built in the most sensitive analysis categories (see section 5.2).

For the boosted category, the small-$R$ jets are reclustered [103] using the anti-$k_t$ algorithm with a radius parameter of 1.0, resulting in a collection of large-$R$ jets (referred to as RC jets). These RC jets are required to have a reconstructed invariant mass higher than 50 GeV, $p_T > 200$ GeV and at least two small-$R$ constituent jets. RC jets are used as input to a deep neural network (DNN), described in section 5.2, to identify high-$p_T$ (boosted) top-quark and Higgs boson candidates decaying into collimated hadronic final states. The event is flagged as containing a boosted Higgs boson candidate if one of the RC jets has a high probability of originating from a Higgs boson, as estimated by the DNN.

Events are required to have exactly one lepton in the single-lepton channels and exactly two leptons with opposite electric charge in the dilepton channel. At least one of the leptons must have $p_T > 27$ GeV and match a corresponding object at trigger level. In the $ee$ and $\mu\mu$ channels, the dilepton invariant mass must be above 15 GeV and outside the $Z$-boson mass window 83–99 GeV. To maintain orthogonality with other $t\bar{t}H$ channels [104], events are vetoed if they contain one or more (two or more) $\tau_{\text{had}}$ candidates in the dilepton (single-lepton) channel. Leptons are further required to satisfy additional identification and isolation criteria to increase background rejection: electrons (muons) must pass the Tight (Medium) identification criterion and the Gradient (FixedCutTightTrackOnly) isolation criteria [93, 105].

5 Analysis strategy

In order to target the $t\bar{t}H(b\bar{b})$ final state, events are categorised into mutually exclusive regions defined by the number of leptons, the number of jets, the number of $b$-tagged jets at different $b$-tagging efficiencies (60%, 70%, 77%, or 85%) and the number of boosted Higgs boson candidates (see section 5.1).

The analysis regions with higher signal-to-background ratio are referred to as ‘signal regions’ (SR). Higgs boson candidates are reconstructed using a DNN in the boosted category, and boosted decision trees (BDT) referred to as ‘reconstruction BDTs’ in the resolved categories, aiming at associating the reconstructed jets to the final state partons. Kinematic variables of these Higgs boson candidates, such as their transverse momentum $p_T^H$, are computed and used together with several other variables as input to other BDTs. These, referred to as ‘classification BDTs’, are then employed to separate signal from background, in each of the SRs (see section 5.2). The remaining analysis regions, depleted in signal, are referred to as ‘control regions’ (CR). They provide stringent constraints on the normalisation of the backgrounds and on systematic uncertainties when used in a combined fit with the signal regions.
5.1 Analysis regions

Table 2 defines the 16 regions into which the events are classified: 11 SRs (dilepton SR$_{\geq 4j, \geq 4b}$, single-lepton SR$_{\geq 6j, \geq 4b}$, and SR$_{\text{boosted}}$, split according to the reconstructed $p_T^H$ into four, five and two regions, respectively), and five CRs.

In the single-lepton channel, events are assigned to the boosted category if they contain at least four jets $b$-tagged at the 85% working point, one boosted Higgs boson candidate, and at least two jets not belonging to the boosted Higgs boson candidate which are $b$-tagged at the 77% working point. The boosted Higgs boson candidate must satisfy $p_T > 300$ GeV, have an invariant mass in the range 100–140 GeV, a DNN score $P(H) > 0.6$ (see section 5.2), and exactly two jet constituents $b$-tagged at the 85% working point. If more than one candidate is identified, the one with a mass closest to the Higgs boson mass is chosen. The selected RC jet is used to determine the kinematic properties of the boosted Higgs boson candidate (reconstructed $p_T^H$, $m_{b\bar{b}}$, etc.). All other selected events belong to the resolved categories.

The analysis regions are further split according to the reconstructed $p_T^H$ to allow the extraction of multiple signal parameters, sensitive to new physics effects. The $p_T^H$ ranges are the same as used to define STXS bins with ‘truth’ $\hat{p}_T^H$ (where the ‘truth’ $\hat{p}_T^H$ is taken from the MC event record before the Higgs boson decays), and were chosen to minimise the correlation among signal strengths in different STXS bins. The split goes as follows:

- The single-lepton resolved signal region, SR$_{\geq 6j, \geq 4b}$, is split in five reconstructed $p_T^H$ regions: 0–120 GeV, 120–200 GeV, 200–300 GeV, 300–450 GeV and $\geq 450$ GeV.
- The dilepton signal region, SR$_{\geq 4j, \geq 4b}$, is split in a similar manner, with the two highest $p_T^H$ bins merged because only a small number of events are expected.
- The boosted signal region, SR$_{\text{boosted}}$, is split into two reconstructed $p_T^H$ regions: 300–450 GeV and $\geq 450$ GeV.
- Control regions are inclusive in reconstructed $p_T^H$ to maintain the constraints on the background composition.

After these selections are applied, $t\bar{t} +$ heavy-flavour jets dominate the background composition and the $t\bar{t}H$ selection efficiency is 1.2%. Although all Higgs boson decay modes are considered, $H \rightarrow b\bar{b}$ events account for at least 94% of $t\bar{t}H$ events selected in the signal regions. In the SRs the shape and normalisation of a multivariate discriminant distribution is used in the statistical analysis, except in the highest reconstructed $p_T^H$ bin of the single-lepton resolved analysis, where only the event yield is used.

In the dilepton CRs, only the event yield is used to correct the amount of $t\bar{t} + \geq 1c$ predicted from the inclusive $t\bar{t} +$ jets sample. In the single-lepton channel the distribution shape and normalisation of the average $\Delta R$ for all possible combinations of $b$-tagged jet pairs, $\Delta R_{bb}^{\text{avg}}$, is used in the CRs to help better constrain the background contributions and correct their shape. Control regions have different ratios of $t\bar{t} + \geq 1b$ to $t\bar{t} + \geq 1c$ events: regions labelled ‘hi’, referring to a higher $b$-tagging probability, are enriched in $t\bar{t} + \geq 1b$, while in regions labelled ‘lo’, the proportion of $t\bar{t} + \geq 1c$ events is increased. The different
proportions of $t\bar{t} + \geq 1b$ and $t\bar{t} + \geq 1c$ in the control regions allow the signal extraction fit to better constrain the relative fractions of these processes in the signal regions.

### 5.2 Multivariate analysis

Multivariate classifiers are used in two parts of this analysis: identifying Higgs boson candidate objects and classifying $t\bar{t}H$ signal events. In all SRs of the resolved categories, the multivariate classifiers are constructed analogously to the reconstruction and classification BDTs used in the previous analysis [27] and trained with TMVA [106]. The training for the reconstruction BDTs is identical to this previous analysis, matching reconstructed jets to the partons emitted from top-quark and Higgs boson decays. For this purpose, W-boson, top-quark and Higgs boson candidates are built from combinations of jets and leptons. The $b$-tagging information is used to discard combinations containing jet-parton assignments inconsistent with the correct parton candidate flavour. The combination of jets with the highest reconstruction BDT score is selected, allowing the computation of kinematic properties of the Higgs boson candidate (reconstructed $p_T^H$, $m_{\text{Higgs}}$, etc.) by summing the four-momenta of the two jets associated with the Higgs boson candidate in this combination.

In the boosted channel the Higgs boson candidate is found using a DNN with a three-node softmax output layer trained with Keras [107] and a TensorFlow backend [108] on a sample of RC jets from signal events. The DNN quantifies the probability that an RC jet is originated from a Higgs boson (labelled $P(H)$ and shown in figure 2), top quark or any other process (mostly multijet production). The most important DNN input variables for identifying a Higgs boson candidate jet are built from the small-$R$ jet constituent masses.
Figure 2. Comparison between data and prediction for the DNN $P(H)$ output for the Higgs boson candidate prior to any fit to the data in the single-lepton boosted channel for (a) $300 \leq p_{T}^{H} < 450$ GeV and (b) $p_{T}^{H} \geq 450$ GeV. The dashed line shows the $t\bar{t}H$ signal distribution normalised to the total background prediction. The uncertainty band includes all uncertainties and their correlations.

and pseudo-continuous $b$-tagging values, while substructure variables [109] also contribute. The Higgs boson jets are correctly identified 76% of the time, while top-quark jets are identified in 67% of the cases, in signal events.

The performance of the reconstruction in the three channels is reported in figure 3, which shows the migration matrix for Higgs boson candidates between reconstructed $p_{T}^{H}$ bins and ‘truth’ $p_{T}^{H}$ bins, for all reconstructed Higgs boson candidates. When applying the reconstruction BDT or the DNN (which are trained with $t\bar{t}H$ events to match $b$-jets to a Higgs boson candidate) to the $t\bar{t} + \geq 1b$ sample, the two selected jets do not originate from the top-quark decay products about half (two-thirds) of the time in the lower (higher) $p_{T}^{H}$ bin. The number of events with both jets originating from top-quark decays is negligible.

The classification BDTs have been retrained since the previous analysis to profit from the improved background modelling and from the increased number of simulated events, using the signal and components of the nominal background model presented in this paper. The dilepton BDT is trained only against $t\bar{t} + b\bar{b}$ events (as it constitutes most of the background), the single-lepton resolved BDT is trained against $tt + \geq 1c$ and $tt + \text{light}$ events also contribute) and the single-lepton boosted BDT is trained against all background processes.

These BDTs are built by combining kinematic variables, such as invariant masses and angular separations of pairs of reconstructed jets and leptons, outputs of the reconstruction discriminants, as well as the pseudo-continuous $b$-tagging discriminant of selected jets. The complete lists of input variables can be found in appendix A. The reconstruction discriminants provide their own output value as well as variables derived from the selected combination of jets with the highest reconstruction BDT score in the resolved channels.
In the single-lepton resolved channel, a likelihood discriminant method that combines the signal and background probabilities of all possible jet combinations in each event is also used as input to the classification BDT [27]. In the boosted channel, information from the DNN is used as input to the classification BDT, including the Higgs probability, $P(H)$. Distributions of the output of these BDT classifiers serve as SR inputs to the signal extraction fit.

The most important variables entering the dilepton BDT include the reconstruction BDT score for the Higgs boson candidate identified using Higgs boson information, the average $\Delta \eta$ between $b$-tagged jets, the minimum invariant mass of a $b$-tagged jet pair, the $\Delta R$ between the Higgs candidate and the $t\bar{t}$ candidate system, and the number of $b$-tagged jet pairs with an invariant mass within 30 GeV of 125 GeV. The most important variables entering the single-lepton resolved BDT include the likelihood discriminant, the average $\Delta R$ for all possible combinations of two $b$-tagged jets ($\Delta R_{bb}^{\text{avg}}$), the invariant mass of the two $b$-tagged jets with the smallest $\Delta R$, the reconstruction BDT score for the Higgs boson candidate identified using Higgs boson information, and the $\Delta R$ between the two highest-$p_T$ $b$-tagged jets. The most important variables entering the single-lepton boosted BDT include the DNN $P(H)$ output for the Higgs boson candidate, the sum of $b$-tagging discriminants of small-$R$ jets from Higgs, hadronic top and leptonically top candidates, the hadronic top candidate’s invariant mass, the small-$R$ jet multiplicity and, in the sum of $b$-tagging discriminants, the fraction due to all jets not associated with the Higgs or hadronic top candidates.

6 Systematic uncertainties

Many sources of systematic uncertainty affect this analysis, impacting the categorisation of events as well as the shape and normalisation of the final discriminants used in the signal extraction fit. All sources of experimental uncertainty considered, with the exception of the
uncertainty in the luminosity, affect both the normalisations and shapes of distributions in all the simulated event samples. Uncertainties related to the theoretical modelling of the signal and background processes affect both the normalisations and shapes of the distributions, with the exception of cross-section and normalisation uncertainties which only affect the overall yield of the considered sample. Nonetheless, the normalisation uncertainties modify the relative fractions of the different samples leading to a shape uncertainty in the distribution of the final discriminant for the total prediction in the different analysis regions.

A single independent nuisance parameter is assigned to each source of systematic uncertainty. Some of the systematic uncertainties, in particular most of the experimental uncertainties, are decomposed into several independent sources, as specified in the following. Each individual source then has a correlated effect across all the channels, analysis regions, and signal and background samples. Modelling uncertainties are typically broken down into components which target specific physics effects in the event generation, such as scale variations or changing the hadronisation model, and are uncorrelated between different samples.

6.1 Experimental uncertainties

The uncertainty in the combined 2015–2018 integrated luminosity is 1.7% [110], obtained using the LUCID-2 detector [111] for the primary luminosity measurement. An uncertainty associated with the modelling of pile-up in the simulation is included to cover the difference between the predicted and measured inelastic cross-section values [112].

The jet energy scale uncertainty is derived by combining information from test-beam data, LHC collision data and simulation, and the jet energy resolution uncertainty is obtained by combining dijet $p_T$-balance measurements and simulation [96]. Additional considerations related to jet flavour, pile-up corrections, $\eta$ dependence and high-$p_T$ jets are included. These uncertainties are further propagated into the single-lepton boosted analysis by applying the reclustering described in section 4 with systematically varied inputs. A total of 40 independent contributions are considered. While the uncertainties are not large, varying between 1% and 5% per jet (depending on the jet $p_T$), the effects are amplified by the large number of jets considered in the final state. The efficiency to identify and remove jets from pile-up is measured with $Z \to \mu^+ \mu^-$ events in data using techniques similar to those used in ref. [97]. All small-$R$ jet constituent uncertainties are propagated to RC jets.

The efficiency to correctly tag $b$-jets is measured using dileptonic $t\bar{t}$ events [98]. The mis-tag rate for $c$-jets is measured using single-lepton $t\bar{t}$ events, exploiting the $c$-jets from the hadronic $W$-boson decays using techniques similar to those in ref. [99]. The mis-tag rate for light-jets is measured using a negative-tag method similar to that in ref. [100] applied to $Z$+ jets events. The uncertainty in tagging $b$-jets is 2%–10% depending on the working point and jet $p_T$. The uncertainty in mis-tagging $c$-jets (light-jets) is 10%–25% (15%–50%) depending on the working point and jet $p_T$. For the calibration of the four working points used in this analysis, a large number of uncertainty components are considered, yielding 45, 20, and 20 uncorrelated sources of uncertainty for $b$-, $c$- and light-jets, respectively.

Uncertainties associated with leptons arise from the trigger, reconstruction, identification, and isolation, as well as the lepton momentum scale and resolution. Efficiencies are measured and scale and resolution calibrations are performed using leptons in $Z \to \ell^+ \ell^-$.
and $J/\psi \rightarrow \ell^+\ell^-$ events [93, 105]. Systematic uncertainties in these measurements account for 22 independent sources but have only a small impact on the final result.

All uncertainties related to the energy scales or resolution of the reconstructed objects are propagated to the calculation of the missing transverse momentum. Three additional uncertainties associated with the scale and resolution of the soft term are also included. As the missing transverse momentum is only used for event reconstruction and not for event selection, these uncertainties have a minimal impact.

6.2 Theoretical modelling uncertainties

Uncertainties in the predicted SM $t\bar{t}H$ signal cross-section are evaluated with a particular focus on the impact on STXS bins. An uncertainty of $\pm 3.6\%$ from varying the PDF and $\alpha_s$ in the fixed-order calculation is applied [20, 113–117]. The effect of PDF variations on the shape of the distributions considered in this analysis is found to be negligible. Uncertainties in the Higgs boson branching fractions are also considered, and amount to 2.2% for the $b\bar{b}$ decay mode [20]. An uncertainty related to the amount of initial-state radiation (ISR) is estimated by simultaneously varying the renormalisation and factorisation scales in the ME and $\alpha_s^{\text{ISR}}$ in the PS [118], while an uncertainty related to final-state radiation (FSR) is estimated by varying $\alpha_s^{\text{FSR}}$ in the PS. The nominal POWHEG BOX + PYTHIA 8 sample is also compared with the POWHEG BOX + HERWIG 7 sample to assess an uncertainty related to the choice of PS and hadronisation model, and with the MADGRAPH5_AMC@NLO + PYTHIA 8 sample to assess an uncertainty arising from changing the NLO matching procedure (sample details are given in table 1). Uncertainties due to missing higher-order terms in the perturbative QCD calculations affecting the total cross-section and event migration between STXS bins are estimated by varying the renormalisation and factorisation scales independently by a factor of two, as well as evaluating the ISR and FSR uncertainties. The largest effect was found to originate from the ISR uncertainty, corresponding to a 9.2% variation of the total cross-section, leading to an uncertainty of 10%–17% in bin migrations estimated using the Stewart-Tackmann procedure [119]. All signal uncertainties are correlated across STXS bins, with the exception of bin migration uncertainties.

The systematic uncertainties affecting the $t\bar{t} + \text{jets}$ background modelling are summarised in table 3. An uncertainty of 6% is assumed for the inclusive $t\bar{t}$ production cross-section predicted at NNLO+NNLL, including effects from varying the factorisation and renormalisation scales, the PDFs, $\alpha_s$, and the top-quark mass [46–52]. This uncertainty is applied to $t\bar{t} + \text{light}$ samples only, since this component is dominant in $t\bar{t}$ production in the full phase-space. An uncertainty of 100% in the normalisation of $t\bar{t} + \geq 1c$ events is applied, motivated by the fitted value of this normalisation in the previous analysis [27]. The normalisation of $t\bar{t} + \geq 1b$ is allowed to float freely in the signal extraction fit. The $t\bar{t} + \geq 1b$, $t\bar{t} + \geq 1c$ and $t\bar{t} + \text{light}$ processes are affected by different types of uncertainties: $t\bar{t} + \text{light}$ profits from relatively precise measurements in data; $t\bar{t} + \geq 1b$ and $t\bar{t} + \geq 1c$ can have similar or different diagrams depending on the precision of the ME and the flavour scheme used for the PDF, and the different masses of the $c$- and $b$-quarks contribute to additional differences between these two processes. For these reasons, all uncertainties in the $t\bar{t} + \text{jets}$ background
Table 3. Summary of the sources of systematic uncertainty for $t\bar{t} + \text{jets}$ modelling. The systematic uncertainties listed in the second section of the table are evaluated in such a way as to have no impact on the normalisation of the three $t\bar{t} + \geq 1b$, $t\bar{t} + \geq 1c$, and $t\bar{t} + \text{light}$ components in the phase-space selected in this analysis. The last column of the table indicates the $t\bar{t} + \text{jets}$ components to which a systematic uncertainty is assigned. All systematic uncertainty sources are treated as uncorrelated across the three components.

Modelling are assigned independent nuisance parameters for the $t\bar{t} + \geq 1b$, $t\bar{t} + \geq 1c$ and $t\bar{t} + \text{light}$ processes. The effect of PDF uncertainties was found to be negligible.

Systematic uncertainties in the acceptance and the distribution shapes are extracted from comparisons between the nominal prediction and different samples or settings. The fraction of $t\bar{t} + \geq 1b$ events in the selected phase-space in all alternative samples is reweighted to match the fraction in the nominal sample. This is to allow the normalisation of $t\bar{t} + \geq 1b$ to be driven solely by the free-floating parameter in the signal extraction fit to data. The systematic uncertainties related to varying the amount of ISR, the amount of FSR, the PS and hadronisation model, and the NLO matching procedure are estimated using the same procedure as for $t\bar{t}H$, comparing the nominal prediction with alternative samples. In the specific case of $t\bar{t} + \geq 1b$, relative uncertainties are used to estimate the effect of changing the PS and hadronisation model or the NLO matching procedure by comparing predictions from the NLO $t\bar{t}$ generators (see table 3). These comparisons are made using predictions in which the additional $b$-quarks were generated at leading-log precision from gluon splitting. A check was performed, comparing the nominal $t\bar{t} + \geq 1b$ prediction with a smaller $t\bar{t} + b\bar{b}$ sample produced with SHERPA 2.2.1. The size of the difference estimated from $t\bar{t}$ samples was observed to be generally the same as or larger than the difference between the two $t\bar{t}+b\bar{b}$ NLO predictions. The impact of these uncertainties on the final results is reported in section 7.

Special consideration is given to the correlation of modelling uncertainties across different $p_T^H$ bins, in order to provide the fit with enough flexibility to cover background mismodelling without biasing the signal extraction. The $t\bar{t} + \geq 1b$ NLO matching uncertainty is shown to depend on $p_T^H$ and is therefore decorrelated across $p_T$ bins in the SRs. The NLO matching uncertainty and the PS and hadronisation uncertainties for $t\bar{t} + \geq 1b$ are further
Figure 4. Pre-fit distributions of the reconstructed Higgs boson candidate $p_T^H$ for the (a) dilepton SR$_{≥4b}^{≥4j}$, (b) single-lepton resolved SR$_{≥6j}^{≥4b}$ and (c) single-lepton boosted SR$_{boosted}^{≥4b}$ signal regions. The dashed line shows the $ttH$ signal distribution normalised to the total background prediction. The uncertainty band includes all uncertainties and their correlations, except for the uncertainty in the $k(tt + ≥1b)$ normalisation factor which is not defined pre-fit. The last bin includes the overflow.

decorrelated between the single-lepton and dilepton channels in order to avoid the transfer of constraints from the single-lepton resolved channel to the dilepton channel, which is less sensitive to the high-$p_T^H$ regime and produces less additional radiation. The pre-fit distributions of the reconstructed $p_T^H$ are shown in figure 4. An additional uncertainty is derived for the $tt + ≥1b$ sample to cover the mismodelling observed in this distribution. After removing the overall normalisation difference by scaling the $tt + ≥1b$ background in the dilepton SR$_{≥4b}^{≥4j}$ (single-lepton SR$_{≥6j}^{≥4b}$), a weight is computed in each reconstructed $p_T^H$ bin of the dilepton SR$_{≥4b}^{≥4j}$ (single-lepton SR$_{≥6j}^{≥4b}$), which corrects the predicted $tt + ≥1b$ contribution so that the data and background model yields agree in each $p_T^H$ bin. The derived weights are not applied to the nominal sample: instead they define the $+1σ$ variation of an additional uncertainty in the $tt + ≥1b$ sample in each reconstructed $p_T^H$ bin. The weights derived in the single-lepton resolved channel are also applied in the boosted channel. This uncertainty enters the signal extraction fit as a single nuisance parameter ($p_T^{bb}$ shape), correlated across all channels, such that a pull of $+1σ$ corresponds to applying this weight, effectively correcting the reconstructed $p_T^H$ spectrum.

To account for variations in the $tt + ≥1b$ subcomponent fractions found in different predictions, an additional nuisance parameter is introduced to cover the largest discrepancy between two models for the fraction of $tt + 1b$ and $tt + ≥2b$. The $1σ$ variation of this nuisance parameter corresponds to reducing the amount of $tt + ≥2b$ by $13\%$ and increasing the amount of $tt + 1b$ by $22\%$. This uncertainty is correlated across all regions, and impacts each region differently due to the varying compositions of $tt + ≥1b$.

An uncertainty of $5\%$ is considered for the cross-sections of the three single-top production modes [55, 56, 120, 121]. Uncertainties associated with the PS and hadronisation model, and with the NLO matching scheme, are evaluated by comparing, for each process,
the nominal Powheg Box + Pythia 8 sample with a sample produced using Powheg Box + Herwig 7 and MadGraph5_AMC@NLO + Pythia 8, respectively. The uncertainty associated with the interference between $tW$ and $t\bar{t}$ production at NLO [59] is assessed by comparing the nominal Powheg Box + Pythia 8 sample produced using the diagram removal scheme with the alternative sample produced with the same generator but using the diagram subtraction scheme.

An uncertainty of 40% is assumed for the $W + \text{jets}$ cross-section, with an additional 30% normalisation uncertainty used for $W +$ heavy-flavour jets, taken as uncorrelated between events with two and more than two heavy-flavour jets. These uncertainties are based on variations of the factorisation and renormalisation scales and of the Sherpa matching parameters. An uncertainty of 35% is applied to the $Z + \text{jets}$ normalisation, uncorrelated across jet bins, to account for both the variations of the scales and Sherpa matching parameters and the uncertainty in the extraction from data of the correction factor for the heavy-flavour component.

The uncertainty in the $t\bar{t}V$ NLO cross-section prediction is 15% [122], split into PDF and scale uncertainties as for $t\bar{t}H$. An additional $t\bar{t}V$ modelling uncertainty, related to the choice of PS and hadronisation model and NLO matching scheme is assessed by comparing the nominal MadGraph5_AMC@NLO + Pythia 8 samples with alternative ones generated with Sherpa.

The uncertainty in the diboson background is assumed to be 50%, which includes uncertainties in the inclusive cross-section and additional jet production [123]. A 50% normalisation uncertainty is considered for the four-top background, covering effects from varying the factorisation and renormalisation scales, the PDFs and $\alpha_s$ [124].

The small backgrounds from $tZq$ and $tWZ$ are each assigned cross-section uncertainties; for $tZq$ two uncertainties are used, 7.9% accounting for factorisation and renormalisation scale variations and 0.9% accounting for PDFs, and for $tWZ$ a single uncertainty of 50% is used [124]. Uncertainties in the associated production of a single top quark and a Higgs boson include factorisation and renormalisation scale variations as well as PDF uncertainties: they amount to $+6.5/-14.9\%$ $(+6.5/-6.7\%)$ and $\pm3.7\%$ $(\pm6.3\%)$ for $tHjb$ ($tWH$), respectively [20]. Finally, a 25% normalisation uncertainty is considered for the fake lepton background in the dilepton channel.

7 Results

The distributions of the classification BDT in the signal regions, and the event yield or the $\Delta R_{bb}^{\text{avg}}$ distributions in the dilepton or single-lepton control regions, respectively, are combined in a profile likelihood fit to extract the signal, while simultaneously determining the yields and constraining the normalisation and shape of differential distributions of the most important background components. The binning of these distributions is optimised to maximise the analysis sensitivity while keeping the total MC statistical uncertainty in each bin to a level adjusted to avoid biases due to fluctuations in the predicted number of events.

The statistical analysis is based on a binned likelihood function $L(\mu, \theta)$ constructed as a product of Poisson probability terms over all bins considered in the analysis. The
likelihood function depends on the signal-strength parameter $\mu$, defined as $\mu = \sigma/\sigma_{SM}$ (where $\sigma$ is the measured cross-section and $\sigma_{SM}$ is the Standard Model prediction), and a set of nuisance parameters $\theta$ which characterise the effects of systematic uncertainties in the signal and background expectations. They are implemented in the likelihood function as Gaussian or Poisson priors, with the exception of the unconstrained normalisation factor $k(tt + \geq 1b)$ for the $tt + \geq 1b$ background, of which no prior knowledge from theory or subsidiary measurements is assumed. The statistical uncertainty in the prediction, which incorporates the statistical uncertainty arising from the limited number of simulated events, is included in the likelihood in the form of additional nuisance parameters, one for each of the considered bins. In the statistical analysis, the number of events expected in a given bin depends on $\mu$ and $\theta$. The nuisance parameters $\theta$ adjust the expectations for signal and background according to the corresponding systematic uncertainties, and their fitted values correspond to the amount that best fits the data. The test statistic $t_\mu$ is defined as the profile likelihood ratio: $t_\mu = -2 \ln(L(\mu, \hat{\theta}_\mu)/L(\hat{\mu}, \hat{\theta}))$, where $\hat{\mu}$ and $\hat{\theta}$ are the values of the parameters that maximise the likelihood function, and $\hat{\theta}_\mu$ are the values of the nuisance parameters that maximise the likelihood function for a given value of $\mu$ [125]. This test statistic, as implemented in the RooStat framework [126, 127], is used to assess the compatibility of the observed data with the background-only hypothesis (i.e. for $\mu = 0$) and to make statistical inferences about $\mu$. The uncertainty in the best-fit value of the signal strength is extracted by finding the values of $\mu$ that correspond to varying $t_\mu$ by one unit.

Tables 4 and 5 show the observed and predicted signal and background event yields in all SRs and CRs before and after the inclusive fit to data. Post-fit values are summarised in figure 5, where the precision increases post-fit due to profiling and the uncertainties can be observed to increase as a function of $p_{H_T}$, ranging from 2% to 12%. The SR BDT distributions are presented in figures 6 and 7, while the $\Delta R_{bb}$ distributions in the single-lepton resolved CRs are shown in figure 8. All distributions are compatible with the data. The normalisation factor for the $tt + \geq 1b$ background is found to be $k(tt + \geq 1b) = 1.28 \pm 0.08$.

The best-fit $\mu$ value is

$$\mu = 0.35 \pm 0.20 \text{ (stat.)} \pm^{0.30}_{0.28} \text{ (syst.)} = 0.35^{+0.36}_{-0.34},$$

corresponding to an observed (expected) significance of 1.0 (2.7) standard deviations with respect to the background-only hypothesis. This observed inclusive $\mu$ and its uncertainty, common to all channels, cannot affect the shape of the signal distributions but only their normalisation, leading to post-fit signal yield uncertainties very close to 100%.

The statistical uncertainty is obtained by repeating the fit to data after fixing all nuisance parameters to their post-fit values, with the exception of the free normalisation factors in the fit: $k(tt + \geq 1b)$ and $\mu$. The total systematic uncertainty is obtained by subtracting the statistical variance from the total variance, i.e. $\sigma_{\text{syst}} = \sqrt{\sigma_{\text{tot}}^2 - \sigma_{\text{stat}}^2}$. The expected significance is computed from a fit to a pseudo-dataset, built using the pulls from the nominal fit when fixing $\mu = 1$. Figure 9 shows the event yield in data compared with the post-fit prediction for all events entering the analysis selection, grouped and ordered by the signal-to-background ratio of the corresponding final-discriminant bins.
Table 4. Pre-fit and post-fit event yields in the dilepton channel (top) signal regions and (bottom) control regions. Post-fit yields are after the inclusive fit in all channels. All uncertainties are included, taking into account correlations in the post-fit case. The $k(t\ell + \geq 1b)$ uncertainty is not defined pre-fit and therefore only included in the post-fit uncertainties. For the $t\bar{t}H$ signal, the pre-fit yield values correspond to the theoretical prediction and corresponding uncertainties, while the post-fit yield and uncertainties correspond to those in the inclusive signal-strength measurement. ‘Other sources’ refers to s-channel, t-channel, $tW$, $tWZ$, $tZq$, $Z +$ jets and diboson events.

The global goodness of fit, including all input variables to the classification BDTs and to a fit using the saturated model [128], is 92%, validating the good post-fit modelling achieved. Some of the most important variables used in the classification BDTs are shown in figure 10 for the dilepton channel, figure 11 for the single-lepton resolved channel and figure 12 for the single-lepton boosted channel. The probability of obtaining a level of agreement worse than observed between the fitted predictions and data was evaluated by calculating the $\chi^2$ for a given number of degrees of freedom and integrating the cumulative probability distribution to +$\infty$. For all channels combined, the mean probability is 80% for classification BDT outputs and 60% for classification BDT training variables in all SRs. Post-fit distributions of the reconstructed Higgs boson candidate mass in the three SRs are shown in figure 13.
Table 5. Pre-fit and post-fit event yields in the single-lepton (top) resolved signal regions and (bottom) boosted signal regions and control regions. Post-fit yields are after the inclusive fit in all channels. All uncertainties are included, taking into account correlations in the post-fit case. The $k(tt+ \geq 1b)$ uncertainty is not defined pre-fit and therefore only included in the post-fit uncertainties. For the $ttH$ signal, the pre-fit yield values correspond to the theoretical prediction and corresponding uncertainties, while the post-fit yield and uncertainties correspond to those in the inclusive signal-strength measurement. ‘Other top sources’ refers to s-channel, t-channel, $tWZ$ and $tZq$ events.

The probability of the obtained signal strength being compatible with the SM prediction is 8.5%, estimated by redoing the fit while fixing $\mu = 1$. The measured signal strength is compatible with that obtained previously [27] from part of the dataset, and the impact of systematic uncertainties has been reduced by about a factor of two. The main improvements come from improved theoretical knowledge in $tt+ \geq 1b$ modelling, from the much larger size of simulated event samples for systematic uncertainty estimation as well as from the refined $b$-tagging scale factors and jet energy scale and resolution measurements.
The fit is repeated with three independent signal strengths, one each for the dilepton, single-lepton resolved and single-lepton boosted channels. Figure 14 shows the $\mu$ value obtained for each channel and the single signal strength from the previous fit. In this fit the normalisation factor for the $t\bar{t} + \geq 1b$ background is found to be $k(t\bar{t} + \geq 1b) = 1.27 \pm 0.08$, compatible with the value in the single $\mu$ fit. The probability of obtaining a discrepancy between these three $\mu$ values equal to or larger than the one observed is 90%.

The measurement is largely dominated by systematic uncertainties. Their contributions to the fit to $\mu$ are reported in table 6. The dominant impact comes from the modelling of the $t\bar{t} + \geq 1b$ background, followed by the signal modelling, $tW$ modelling and $b$-tagging efficiency uncertainties. The largest observed pull on systematic uncertainties, as shown in figure 15, is seen in the $t\bar{t} + \geq 1b$ ISR uncertainty, and is about 1.2$\sigma$, mostly driven by the
Uncertainty

Figure 7. Comparison between data and prediction for the BDT discriminant in the single-lepton SRs after the inclusive fit to the data. The resolved channel is shown for (a) $0 \leq p_T^H < 120\,\text{GeV}$, (b) $120 \leq p_T^H < 200\,\text{GeV}$, (c) $200 \leq p_T^H < 300\,\text{GeV}$, (d) $300 \leq p_T^H < 450\,\text{GeV}$ and (e) $p_T^H \geq 450\,\text{GeV}$ (yield only). The boosted channel is shown for (f) $300 \leq p_T^H < 450\,\text{GeV}$ and (g) $p_T^H \geq 450\,\text{GeV}$. The $ttH$ signal yield (solid red) is normalised to the fitted $\mu$ value from the inclusive fit. The dashed line shows the $ttH$ signal distribution normalised to the total background prediction. The uncertainty band includes all uncertainties and their correlations.

Figure 8. Comparison between data and prediction for $\Delta R_{bb}^{\text{avg}}$ after the inclusive fit to the data in the single-lepton (a) $CR_{\geq 4b_{\text{lo}}}$ and (b) $CR_{\geq 4b_{\text{hi}}}$ control regions. The $ttH$ signal yield (solid red) is normalised to the fitted $\mu$ value from the inclusive fit. The dashed line shows the $ttH$ signal distribution normalised to the total background prediction. The uncertainty band includes all uncertainties and their correlations. The first (last) bin includes the underflow (overflow).
Figure 9. Post-fit yields of signal (S) and total background (B) as a function of log(S/B), compared with data. Final-discriminant bins in all dilepton and single-lepton analysis regions are combined into bins of log(S/B), with the signal normalised to the SM prediction used for the computation of log(S/B). The signal is then shown normalised to the best-fit value and the SM prediction. The lower frame reports the ratio of data to background, and this is compared with the expected ttH-signal-plus-background yield divided by the background-only yield for the best-fit signal strength (solid red line) and the SM prediction (dashed orange line).

The lower frame reports the ratio of data to background, and this is compared with the expected $t\bar{t}H$-signal-plus-background yield divided by the background-only yield for the best-fit signal strength (solid red line) and the SM prediction (dashed orange line).

Another large pull is on the reconstructed $p_T^H$ shape uncertainty in the $t\bar{t} + \geq 1b$ background, as expected from the pre-fit mismodelling (see figure 4) and how this uncertainty is defined: a $+1\sigma$ variation corresponds to correcting the reconstructed $p_T^H$ shape such that it agrees between data and the background model (see section 6.2). The sensitivity of the result to this uncertainty was tested by replacing the data-driven mismodelling with decorrelated free-floating $t\bar{t} + \geq 1b$ normalisation factors across the STXS bins and analysis regions, and no bias was observed on the fitted signal strength. The reconstructed $p_T^H$ distributions display good post-fit agreement, as shown in figure 17 (to be compared with the pre-fit discrepancies shown in figure 4).
Figure 10. Comparison between data and prediction for (a) the reconstruction BDT score for the Higgs boson candidate identified using Higgs boson information, and (b) the average $\Delta y$ between $b$-tagged jets, after the inclusive fit to the data in the dilepton resolved channel for $0 \leq p_T < 120\text{ GeV}$. The $ttH$ signal yield (solid red) is normalised to the fitted $\mu$ value from the inclusive fit. The dashed line shows the $ttH$ signal distribution normalised to the total background prediction. The uncertainty band includes all uncertainties and their correlations.

Figure 11. Comparison between data and prediction for (a) the likelihood discriminant, and (b) the average $\Delta R$ for all possible combinations of $b$-tagged jet pairs, after the inclusive fit to the data in the single-lepton resolved channel for $0 \leq p_T^H < 120\text{ GeV}$. The $ttH$ signal yield (solid red) is normalised to the fitted $\mu$ value from the inclusive fit. The dashed line shows the $ttH$ signal distribution normalised to the total background prediction. The uncertainty band includes all uncertainties and their correlations.
Figure 12. Comparison between data and prediction for the DNN $P(H)$ output for the Higgs boson candidate after the inclusive fit to the data in the single-lepton boosted channel for (a) $300 \leq p_T^H < 450$ GeV and (b) $p_T^H \geq 450$ GeV. The $t\bar{t}H$ signal yield (solid red) is normalised to the fitted $\mu$ value from the inclusive fit. The dashed line shows the $t\bar{t}H$ signal distribution normalised to the total background prediction. The uncertainty band includes all uncertainties and their correlations.

Figure 13. Post-fit distributions of the reconstructed Higgs boson candidate mass for the (a) dilepton SR$\geq4l$, (b) single-lepton resolved SR$\geq6j$ and (c) single-lepton boosted SR$\text{boosted}$ signal regions. The $t\bar{t}H$ signal yield (solid red) is normalised to the fitted $\mu$ value from the inclusive fit. The dashed line shows the $t\bar{t}H$ signal distribution normalised to the total background prediction. The uncertainty band includes all uncertainties and their correlations. The first (last) bin includes the underflow (overflow).
Additionally, a large pull is observed in the $t\bar{t} + \geq 1b$ NLO matching uncertainty in the dilepton channel in the $0 \leq p_T^H < 120$ GeV bin, as shown in figure 15 where it is ranked sixth for the impact on this measurement. Due to the large number of nuisance parameters entering into the likelihood fit, there is a non-negligible probability that any one of the $t\bar{t} + \geq 1b$ NLO matching uncertainties could be pulled to a value at least as extreme as that observed from a purely statistical standpoint. The probability is calculated to be 17%. Despite the large differences between models (see section 6.2), the varying $t\bar{t} + \geq 1b$ composition in the CRs and SRs allows the $t\bar{t} + \geq 1b$ subcomponent fraction systematic uncertainty to be constrained. In general the fit mostly constrains the $t\bar{t} + \geq 1b$ modelling uncertainties and the normalisation of the $t\bar{t} + \geq 1c$ background, which is pulled to 0.6σ and agrees very well with the previous publication [27]. To further validate the robustness of the fit, a pseudo-data sample was built from simulated events by replacing the nominal $t\bar{t} + \geq 1b$ background with the alternative $t\bar{t} + b\bar{b}$ MC sample generated with SHERPA 2.2.1. The fit to this pseudo-data sample did not reveal any significant bias in the signal extraction, the signal strength being compatible with unity within uncertainties.

The fit is also performed with multiple signal strengths corresponding to the five ‘truth’ $\hat{p}_T^H$ bins. Figure 18 shows the values obtained. The normalisation factor is found to be $k(t\bar{t} + 1b) = 1.28 \pm 0.08$, in perfect agreement with the single $\mu$ fit value. The global goodness of fit is 88%, summarising the good post-fit modelling obtained. The probability that the obtained signal strengths are compatible with the SM predictions is 45%, estimated by redoing the fit while fixing the $\mu$ value to 1 in the five bins. Overall, pulls and constraints similar to those in the inclusive measurement are observed. The measurement is dominated by systematic uncertainties in the lowest bin of ‘truth’ $\hat{p}_T^H$ (mostly from the $t\bar{t} + \geq 1b$ background modelling), and by statistical uncertainties in the upper three bins.

Cross-section upper limits are also derived in the STXS framework. In this case, the likelihood function is slightly different from the one used to extract signal strengths: the effects of signal scale and PDF uncertainties on the predicted cross-section are not included.
<table>
<thead>
<tr>
<th>Uncertainty source</th>
<th>$\Delta \mu$</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Process modelling</strong></td>
<td></td>
</tr>
<tr>
<td>$ttH$ modelling</td>
<td>+0.13</td>
</tr>
<tr>
<td>$t\bar{t} + \geq 1b$ modelling</td>
<td>$-0.05$</td>
</tr>
<tr>
<td>$t\bar{t} + \geq 1b$ NLO matching</td>
<td>+0.21</td>
</tr>
<tr>
<td>$t\bar{t} + \geq 1b$ fractions</td>
<td>$-0.12$</td>
</tr>
<tr>
<td>$t\bar{t} + \geq 1b$ FSR</td>
<td>+0.10</td>
</tr>
<tr>
<td>$t\bar{t} + \geq 1b$ PS &amp; hadronisation</td>
<td>+0.09</td>
</tr>
<tr>
<td>$t\bar{t} + \geq 1b p_{T}^{bb}$ shape</td>
<td>+0.04</td>
</tr>
<tr>
<td>$t\bar{t} + \geq 1b$ ISR</td>
<td>+0.04</td>
</tr>
<tr>
<td>$t\bar{t} + \geq 1c$ modelling</td>
<td>+0.03</td>
</tr>
<tr>
<td>$t\bar{t} +$ light modelling</td>
<td>$-0.03$</td>
</tr>
<tr>
<td>$tW$ modelling</td>
<td>+0.08</td>
</tr>
<tr>
<td><strong>Background-model statistical uncertainty</strong></td>
<td>+0.04</td>
</tr>
<tr>
<td><strong>b-tagging efficiency and mis-tag rates</strong></td>
<td>$-0.05$</td>
</tr>
<tr>
<td><strong>Jet energy scale and resolution</strong></td>
<td></td>
</tr>
<tr>
<td>$b$-jet energy scale</td>
<td>+0.00</td>
</tr>
<tr>
<td>Jet energy scale (flavour)</td>
<td>$-0.01$</td>
</tr>
<tr>
<td>Jet energy scale (pile-up)</td>
<td>+0.00</td>
</tr>
<tr>
<td>Jet energy scale (remaining)</td>
<td>$-0.01$</td>
</tr>
<tr>
<td>Jet energy resolution</td>
<td>+0.02</td>
</tr>
<tr>
<td>Luminosity</td>
<td>+0.01</td>
</tr>
<tr>
<td>Other sources</td>
<td>+0.03</td>
</tr>
<tr>
<td><strong>Total systematic uncertainty</strong></td>
<td>+0.30</td>
</tr>
<tr>
<td>$t\bar{t} + \geq 1b$ normalisation</td>
<td>+0.04</td>
</tr>
<tr>
<td><strong>Total statistical uncertainty</strong></td>
<td>+0.20</td>
</tr>
<tr>
<td><strong>Total uncertainty</strong></td>
<td>+0.36</td>
</tr>
</tbody>
</table>

**Table 6.** Breakdown of the contributions to the uncertainties in $\mu$. The contributions from the different sources of uncertainty are evaluated after the fit. The $\Delta \mu$ values are obtained by repeating the fit after having fixed a certain set of nuisance parameters corresponding to a group of systematic uncertainties, and then evaluating $(\Delta \mu)^2$ by subtracting the resulting squared uncertainty of $\mu$ from its squared uncertainty found in the full fit. The same procedure is followed when quoting the effect of the $t\bar{t} + \geq 1b$ normalisation. The total uncertainty is different from the sum in quadrature of the different components due to correlations between nuisance parameters existing in the fit.
Figure 15. Ranking of the 20 nuisance parameters with the largest post-fit impact on $\mu$ in the fit. Nuisance parameters corresponding to statistical uncertainties in the simulated event samples are not included. The empty blue rectangles correspond to the pre-fit impact on $\mu$ and the filled blue ones to the post-fit impact on $\mu$, both referring to the upper scale. The impact of each nuisance parameter, $\Delta \mu$, is computed by comparing the nominal best-fit value of $\mu$ with the result of the fit when fixing the considered nuisance parameter to its best-fit value, $\hat{\theta}$, shifted by its pre-fit (post-fit) uncertainties $\pm \Delta \theta$ ($\pm \Delta \hat{\theta}$). The black points show the pulls of the nuisance parameters relative to their nominal values, $\theta_0$. These pulls and their relative post-fit errors, $\Delta \hat{\theta}/\Delta \theta$, refer to the lower scale. The ‘ljets’ (‘dilep’) label refers to the single-lepton (dilepton) channel.

because, while affecting the signal-strength measurements, they do not affect the cross-section measurements. Scale effects are still present in the statistical model though, via the ISR uncertainty, but with no impact on the overall cross-section. The inclusive cross-section of 507 fb is used to calculate these limits, scaled by the fraction of events in each $p_T^H$ bin to establish the fiducial cross-section for each STXS bin. The measured 95% confidence level (CL) cross-section upper limits in each STXS bin are shown in figure 19, where the hatched uncertainty bands correspond to the theoretical uncertainty in the fiducial cross-section prediction in each bin.
Figure 16. Pre-fit (top) and post-fit (bottom) distributions of the number of jets in the (a, d) dilepton SR$_{4j}^{2b}$, (b, e) single-lepton resolved SR$_{6j}^{2b}$, and (c, f) single-lepton boosted SR$_{4j}^{2b}$ signal regions. The $t\bar{t}H$ signal yield (solid red) is normalised to the Standard Model expectation (the fitted $\mu$ value from the inclusive fit) in the pre-fit (post-fit) distributions. The dashed line shows the $t\bar{t}H$ signal distribution normalised to the total background prediction. The uncertainty band includes all uncertainties and their correlations, except in the pre-fit distributions where the uncertainty in the $k(t\bar{t} + \geq 1b)$ normalisation factor is not defined.
**Figure 17.** Post-fit distributions of the reconstructed Higgs boson candidate $p_T^H$ for the (a) dilepton SR$_{≥4b}^{≥2}$ (b) single-lepton resolved SR$_{≥2}$ and (c) single-lepton boosted SR$_{boosted}$ signal regions. The $t\bar{t}H$ signal yield (solid red) is normalised to the fitted $\mu$ value from the inclusive fit. The uncertainty band includes all uncertainties and their correlations. The last bin includes the overflow.

**Figure 18.** Signal-strength measurements in the individual STXS $\rho_1^{H}$ bins, as well as the inclusive signal strength.
8 Conclusion

Measurements of a Standard Model Higgs boson produced in association with a pair of top quarks and decaying into a pair of $b$-quarks are performed. The results are based on the Run 2 dataset of $pp$ collision data collected at $\sqrt{s} = 13$ TeV by the ATLAS detector at the LHC, corresponding to an integrated luminosity of $139 \text{fb}^{-1}$. The event selection targets top-quark pair decays to a final state containing one or two leptons. Machine-learning techniques are used to discriminate between signal and background events, the latter being dominated by $t\bar{t} + \text{jets}$ production. The measured signal strength is $0.35^{+0.36}_{-0.34}$, corresponding to an observed (expected) significance of 1.0 (2.7) standard deviations. The probability of the obtained signal strength being compatible with the SM prediction is 8.5%. The measurement uncertainty is dominated by systematic uncertainties, despite significant improvement relative to the previous measurement [27], especially regarding theoretical knowledge of the $t\bar{t} + \geq 1b$ background process, which still drives the sensitivity. To further test the Standard Model, the first differential measurement of the $ttH$ signal strength was performed in five bins of Higgs boson transverse momentum in the STXS framework, including a bin for specially selected boosted Higgs bosons with transverse momentum above 300 GeV. Upper limits on the STXS cross-sections are also derived. Observed results are compatible with Standard Model expectations within uncertainties.

Acknowledgments

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, Azerbaijan; STFC, Belarus; CNPq and FAPESP, Brazil; NSERC, NRC and CFI, Canada; CERN; ANID, Chile; CAS, MOST and NSFC, China; Minciencias, Colombia; MEYS CR, Czech Republic; DNRF and DNSRC, Denmark; IN2P3-CNRS and CEA-DRF/IRFU, France; SRNSFG, Georgia; BMBF, HGF and MPG, Germany; GSRI, Greece; RGC and Hong Kong SAR, China; ISF and Benoziyo Center, Israel; INFN, Italy; MEXT and JSPS, Japan; CNRST, Morocco; NWO, Netherlands; RCN, Norway; MEiN, Poland; FCT, Portugal; MNE/IFA, Romania; JINR; MES of Russia and NRC KI, Russian Federation; MESTD, Serbia; ARRS and MIZŠ, Slovenia; DSI/NRF, South Africa; MICINN, Spain; SRC and Wallenberg Foundation, Sweden; SERI, SNSF and Cantons of Bern and Geneva, Switzerland; MOST, Taiwan; TAEK, Turkey; STFC, United Kingdom; DOE and NSF, United States of America. In addition, individual groups and members have received support from BCKDF, CANARIE, Compute Canada and CRC, Canada; COST, ERC, ERDF, Horizon 2020 and Marie Skłodowska-Curie Actions, European Union; Investissements d’Avenir Labex, Investissements d’Avenir Idex and ANR, France; DFG and AvH Foundation, Germany; Herakleitos, Thales and Aristeia programmes co-financed by EU-ESF and the Greek NSRF, Greece; BSF-NSF and GIF, Israel; Norwegian Financial Mechanism 2014–2021, Norway; NCN and NAWA, Poland; La Caixa Banking Foundation, CERCA Programme Generalitat de Catalunya and PROMETEO and GenT Programmes Generalitat Valenciana, Spain; Göran Gustafssons Stiftelse, Sweden; The Royal Society and Leverhulme Trust, United Kingdom.

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 (U.K.) and BNL (U.S.A.), the Tier-2 facilities worldwide and large non-WLCG resource providers. Major contributors of computing resources are listed in ref. [130].

A Input variables to the classification BDTs

In this appendix, the full list of variables used as inputs to the classification BDT, described in section 5.2, in each of the signal regions is reported. Variables are listed separately in table 7 for the dilepton channel, in table 8 for the resolved single-lepton channel and in table 9 for the single-lepton boosted channel. Variables from the reconstruction BDT exploit the chosen jet-parton assignment (see section 5.2). The $b$-tagging discriminant assigned to each jet and the large-$R$ jets used to build the Higgs-boson and top-quark candidates in the boosted category are defined in section 4. Some kinematic and topological variables are built considering only $b$-tagged jets in the event. In the resolved and boosted single-lepton channels, $b$-tagged jets are defined as the four jets with the largest value of the $b$-tagging discriminant. If two jets have the same $b$-tagging discriminant value, they are ordered by decreasing jet $p_T$ value. In the dilepton channel, jets are required to pass the 70% working point to be $b$-tagged.
<table>
<thead>
<tr>
<th>Variable</th>
<th>Definition</th>
</tr>
</thead>
<tbody>
<tr>
<td>$m_{bb}^{\text{min}}$</td>
<td>Minimum invariant mass of a $b$-tagged jet pair</td>
</tr>
<tr>
<td>$m_{bb}^{\text{min}} \Delta R$</td>
<td>Invariant mass of the $b$-tagged jet pair with minimum $\Delta R$</td>
</tr>
<tr>
<td>$m_{jj}^{\text{max}} p_{\text{T}}$</td>
<td>Invariant mass of the jet pair with maximum $p_{\text{T}}$</td>
</tr>
<tr>
<td>$m_{bb}^{\text{max}} p_{\text{T}}$</td>
<td>Invariant mass of the $b$-tagged jet pair with maximum $p_{\text{T}}$</td>
</tr>
<tr>
<td>$\Delta \eta_{bb}^{\text{avg}}$</td>
<td>Average $\Delta \eta$ for all $b$-tagged jet pairs</td>
</tr>
<tr>
<td>$N_{bb}^{\text{Higgs 30}}$</td>
<td>Number of $b$-tagged jet pairs with invariant mass within $30\text{GeV}$ of the Higgs boson mass</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Variables from reconstruction BDT</th>
</tr>
</thead>
<tbody>
<tr>
<td>BDT outputs</td>
</tr>
<tr>
<td>$m_{bb}^{\text{Higgs}}$</td>
</tr>
<tr>
<td>$\Delta R_{H,t\bar{t}}$</td>
</tr>
<tr>
<td>$\Delta R_{H,t\bar{t}}^{\text{min}}$</td>
</tr>
<tr>
<td>$\Delta R_{H,b}^{\text{min}}$</td>
</tr>
</tbody>
</table>

**Table 7.** Variables used in the classification BDTs in the dilepton channel. For variables depending on $b$-tagged jets, only jets $b$-tagged using the 70% working point are considered. For variables from the reconstruction BDT, those with a $\dagger$ are from the BDT using Higgs boson information, those with no sign are from the BDT without Higgs boson information while for those with a $\ddagger$ both versions are used.
### Variable Definition

**General kinematic variables**

- $\Delta R_{avg}^{bb}$: Average $\Delta R$ for all $b$-tagged jet pairs
- $\Delta R_{\max}^{bb} p_T$: $\Delta R$ between the two $b$-tagged jets with the largest vector sum $p_T$
- $\Delta \eta_{ij}$: Maximum $\Delta \eta$ between any two jets
- $m_{bb}^{\min} \Delta R$: Mass of the combination of two $b$-tagged jets with the smallest $\Delta R$
- $N_{bb}^{Higgs 30}$: Number of $b$-tagged jet pairs with invariant mass within 30 GeV of the Higgs boson mass
- $Aplanarity$: $1.5 \lambda_2$, where $\lambda_2$ is the second eigenvalue of the momentum tensor
  
- $H_1$: Second Fox-Wolfram moment computed using all jets and the lepton

### Variables from reconstruction BDT

- $BDT$ output: Output of the reconstruction BDT
- $m_{bb}^{Higgs}$: Higgs candidate mass
- $m_{H,b_{lep, top}}$: Mass of Higgs candidate and $b$-jet from leptonic top candidate
- $\Delta R_{bb}^{Higgs}$: $\Delta R$ between $b$-jets from the Higgs candidate
- $\Delta R_{H,tf}$: $\Delta R$ between Higgs candidate and $t\bar{t}$ candidate system
  
  [†] $\Delta R_{H,lep, top}$: $\Delta R$ between Higgs candidate and leptonic top candidate

### Variables from likelihood calculations

- LHD: Likelihood discriminant

### Variables from $b$-tagging

- $w_{bb}^{Higgs, b-tag}$: Sum of $b$-tagging discriminants of $b$-tagged jets from best Higgs candidate from the reconstruction BDT
- $B_{3^{\text{rd}}}^{jet}$: 3$^{\text{rd}}$ largest jet $b$-tagging discriminant
- $B_{4^{\text{th}}}^{jet}$: 4$^{\text{th}}$ largest jet $b$-tagging discriminant
- $B_{5^{\text{th}}}^{jet}$: 5$^{\text{th}}$ largest jet $b$-tagging discriminant

**Table 8.** Input variables to the classification BDTs in the single-lepton resolved channel. For variables depending on $b$-tagged jets, jets are sorted by their pseudo-continuous $b$-tag score, and by their $p_T$ when they have the same pseudo-continuous $b$-tag score. For variables from the reconstruction BDT, those with a † are from the BDT using Higgs boson information, those with no † are from the BDT without Higgs boson information.
<table>
<thead>
<tr>
<th>Variable</th>
<th>Definition</th>
</tr>
</thead>
<tbody>
<tr>
<td>( m_{Hbb} )</td>
<td>Higgs candidate mass</td>
</tr>
<tr>
<td>( p_T^H )</td>
<td>Higgs candidate transverse momentum</td>
</tr>
<tr>
<td>( \eta_{lep} )</td>
<td>( \eta ) of the Higgs candidate relative to the lepton</td>
</tr>
<tr>
<td>( P(H) )</td>
<td>DNN Higgs probability for the Higgs candidate</td>
</tr>
<tr>
<td>( m_{\text{had top}} )</td>
<td>Hadronic top candidate mass</td>
</tr>
<tr>
<td>( p_T^{\text{had top}} )</td>
<td>Hadronic top candidate transverse momentum</td>
</tr>
<tr>
<td>( \eta_{\text{had top}} )</td>
<td>( \eta ) of the hadronic top candidate relative to the lepton</td>
</tr>
<tr>
<td>( B_i^{\text{had top}} )</td>
<td>( i )th largest jet ( b )-tagging discriminant associated to the hadronic top candidate</td>
</tr>
<tr>
<td>( m_{\text{lep top}} )</td>
<td>Leptonic top candidate mass</td>
</tr>
<tr>
<td>( p_T^{\text{lep top}} )</td>
<td>Leptonic top candidate transverse momentum</td>
</tr>
<tr>
<td>( B_{\text{lep top}} )</td>
<td>( b )-tagging discriminant of the jet associated to the leptonic top candidate</td>
</tr>
<tr>
<td>( n_{\text{jets}} )</td>
<td>Small-( R ) jets multiplicity</td>
</tr>
<tr>
<td>( \Delta R_{H,\text{had top}} )</td>
<td>( \Delta R ) between the Higgs and the hadronic top candidates</td>
</tr>
<tr>
<td>( \Delta R_{H,\text{lep top}} )</td>
<td>( \Delta R ) between the Higgs and the leptonic top candidates</td>
</tr>
<tr>
<td>( \Delta R_{\text{had top,lep top}} )</td>
<td>( \Delta R ) between the hadronic top and the leptonic top candidates</td>
</tr>
<tr>
<td>( p_T^{t\bar{t}H} )</td>
<td>( t\bar{t}H ) system transverse momentum</td>
</tr>
<tr>
<td>( p_T^{t\bar{t}} )</td>
<td>( t\bar{t} ) system transverse momentum</td>
</tr>
<tr>
<td>( w_{\text{sum}}^{b\text{-tag}} )</td>
<td>Sum of ( b )-tagging discriminants of jets from Higgs, hadronic and leptonic top candidates</td>
</tr>
<tr>
<td>( w_{\text{add jet}}^{b\text{-tag}} )</td>
<td>Fraction of the sum of ( b )-tagging discriminants of all jets not associated to Higgs or hadronic top candidates</td>
</tr>
</tbody>
</table>

Table 9. Input variables to the classification BDT in the single-lepton boosted channel. For variables depending on \( b \)-tagged jets, jets are sorted by their pseudo-continuous \( b \)-tag score, and by their \( p_T \) when they have the same pseudo-continuous \( b \)-tag score. The \( i \) index runs from zero to two.
References


[27] CMS collaboration, Search for $\tilde{t}\tilde{H}$ production in the $H \rightarrow b\bar{b}$ decay channel with leptonic $\tilde{t}$ decays in proton-proton collisions at $\sqrt{s} = 13$ TeV, *JHEP* **03** (2019) 026 [arXiv:1804.03682] [inSPIRE].


[29] ATLAS collaboration, The ATLAS Experiment at the CERN Large Hadron Collider, 2008 *JINST* **3** S08003 [inSPIRE].


[58] R. Frederix, D. Pagani and M. Zaro, Large NLO corrections in $t\bar{t}W^\pm$ and $t\bar{t}t\bar{t}$ hadroproduction from supposedly subleading EW contributions, JHEP 02 (2018) 031 [arXiv:1711.02116] [nSPIRE].


\textsuperscript{a} Also at Institute of Physics, Azerbaijan Academy of Sciences, Baku; Azerbaijan
\textsuperscript{aa} Also at Institute of Theoretical Physics, Ilia State University, Tbilisi; Georgia
\textsuperscript{ab} Also at Instituto de Fisica Teorica, IFT-UAM/CSIC, Madrid; Spain
\textsuperscript{ac} Also at Istanbul University, Dept. of Physics, Istanbul; Turkey
\textsuperscript{ad} Also at Joint Institute for Nuclear Research, Dubna; Russia
\textsuperscript{ae} Also at Moscow Institute of Physics and Technology State University, Dolgoprudny; Russia
\textsuperscript{af} Also at National Research Nuclear University MEPhI, Moscow; Russia
\textsuperscript{ag} Also at Physics Department, An-Najah National University, Nablus; Palestine
\textsuperscript{ah} Also at Physikalisches Institut, Albert-Ludwigs-Universität Freiburg, Freiburg; Germany
\textsuperscript{ai} Also at The City College of New York, New York NY; United States of America
\textsuperscript{aj} Also at TRIUMF, Vancouver BC; Canada
\textsuperscript{ak} Also at Università di Napoli Parthenope, Napoli; Italy
\textsuperscript{al} Also at University of Chinese Academy of Sciences (UCAS), Beijing; China
\textsuperscript{am} Also at Yeditepe University, Physics Department, Istanbul; Turkey
\textsuperscript{*} Deceased