Search for the Standard Model Higgs boson produced in association with top quarks and decaying into $b\bar{b}$ in $pp$ collisions at $\sqrt{s} = 8$ TeV with the ATLAS detector

The ATLAS Collaboration

Abstract

A search for the Standard Model Higgs boson produced in association with a pair of top quarks, $t\bar{t}H$, is presented. The analysis uses 20.3 fb$^{-1}$ of $pp$ collision data at $\sqrt{s} = 8$ TeV, collected with the ATLAS detector at the Large Hadron Collider during 2012. The search is designed for the $H \rightarrow b\bar{b}$ decay mode and uses events containing one or two electrons or muons. In order to improve the sensitivity of the search, events are categorised according to their jet and $b$-tagged jet multiplicities. A neural network is used to discriminate between signal and background events, the latter being dominated by $t\bar{t}+jets$ production. In the single-lepton channel, variables calculated using a matrix element method are included as inputs to the neural network to improve discrimination of the irreducible $t\bar{t}+b\bar{b}$ background. No significant excess of events above the background expectation is found and an observed (expected) limit of 3.4 (2.2) times the Standard Model cross section is obtained at 95% confidence level. The ratio of the measured $t\bar{t}H$ signal cross section to the Standard Model expectation is found to be $\mu = 1.5 \pm 1.1$ assuming a Higgs boson mass of 125 GeV.

© 2015 CERN for the benefit of the ATLAS Collaboration.
Reproduction of this article or parts of it is allowed as specified in the CC-BY-3.0 license.
Search for the Standard Model Higgs boson produced in association with top quarks and decaying into $b\bar{b}$ in $pp$ collisions at $\sqrt{s} = 8$ TeV with the ATLAS detector

The ATLAS Collaboration

Abstract

A search for the Standard Model Higgs boson produced in association with a pair of top quarks, $t\bar{t}H$, is presented. The analysis uses 20.3 fb$^{-1}$ of $pp$ collision data at $\sqrt{s} = 8$ TeV, collected with the ATLAS detector at the Large Hadron Collider during 2012. The search is designed for the $H \rightarrow b\bar{b}$ decay mode and uses events containing one or two electrons or muons. In order to improve the sensitivity of the search, events are categorised according to their jet and $b$-tagged jet multiplicities. A neural network is used to discriminate between signal and background events, the latter being dominated by $t\bar{t}+\text{jets}$ production. In the single-lepton channel, variables calculated using a matrix element method are included as inputs to the neural network to improve discrimination of the irreducible $t\bar{t}+b\bar{b}$ background. No significant excess of events above the background expectation is found and an observed (expected) limit of 3.4 (2.2) times the Standard Model cross section is obtained at 95% confidence level. The ratio of the measured $t\bar{t}H$ signal cross section to the Standard Model expectation is found to be $\mu = 1.5 \pm 1.1$ assuming a Higgs boson mass of 125 GeV.

1 Introduction

The discovery of a new particle in the search for the Standard Model (SM) Higgs boson [1–3] at the LHC was reported by the ATLAS [8] and CMS [9] collaborations in July 2012. There is now clear evidence of this particle in the $H \rightarrow \gamma\gamma$, $H \rightarrow ZZ^{(*)} \rightarrow 4\ell$, $H \rightarrow WW^{(*)} \rightarrow \ell\nu\ell\nu$ and $H \rightarrow \tau\tau$ decay channels, at a mass of around 125 GeV, which have strengthened the SM Higgs boson hypothesis [10–15] of the observation. To determine all properties of the new boson experimentally, it is important to study it in as many production and decay modes as possible. In particular, its coupling to heavy quarks is a strong focus of current experimental searches. The SM Higgs boson production in association with a top-quark pair ($t\bar{t}H$) [16–19] with subsequent Higgs decay into bottom quarks ($H \rightarrow b\bar{b}$) addresses heavy-quark couplings in both production and decay. Due to the large measured mass of the top quark, the Yukawa coupling of the top quark ($y_t$) is much stronger than that of other quarks. The observation of the $t\bar{t}H$ production mode would allow for a direct measurement of this coupling, to which other Higgs production modes are only sensitive through loop effects. Since $y_t$ is expected to be close to unity, it is also argued to be the quantity that might give insight into the scale of new physics [20].

The $H \rightarrow b\bar{b}$ final state is the dominant decay mode in the SM for a Higgs boson with a mass of 125 GeV. So far, this decay mode has not yet been observed. While a search for this decay via the gluon fusion process is precluded by the overwhelming multijet background, Higgs boson production in association with a vector boson ($VH$) [21–23] or a top-quark pair significantly improves the signal-to-background ratio for this decay.

This paper describes a search for the SM Higgs boson in the $t\bar{t}H$ production mode and is designed to be primarily sensitive to the $H \rightarrow b\bar{b}$ decay, although other Higgs boson decay modes are also treated as signal. Figs. 1(a) and 1(b) show two examples of tree-level diagrams for $t\bar{t}H$ production with a subsequent $H \rightarrow b\bar{b}$ decay. A search for the associated production of the Higgs boson with a top-quark pair using several Higgs decay modes (including $H \rightarrow b\bar{b}$) has recently been published by the CMS Collaboration [24] quoting a ratio of the measured $t\bar{t}H$ signal cross section to the SM expectation for a Higgs boson mass of 125.6 GeV of $\mu = 2.8 \pm 1.0$. 
The main source of background to this search comes from $t\bar{t}$ pairs produced in association with additional jets. The dominant source is $t\bar{t}+bb$ production, resulting in the same final-state signature as the signal. An example is shown in Fig. 1(c). A second contribution arises from $t\bar{t}$ production in association with light-quark ($u$, $d$, $s$) or gluon jets, referred to as $t\bar{t}$+light background, and from $t\bar{t}$ production in association with $c$-quarks, referred to as $t\bar{t}+c\bar{c}$. The size of the second contribution depends on the misidentification rate of the algorithm used to identify $b$-quark jets.

The search presented in this paper uses 20.3 fb$^{-1}$ of data collected with the ATLAS detector in $pp$ collisions at $\sqrt{s} = 8$ TeV during 2012. The analysis focuses on final states containing one or two electrons or muons from the decay of the $t\bar{t}$ system. Selected events are classified into exclusive categories, referred to as “regions”, according to the number of reconstructed jets and $b$-quark jets identified by the $b$-tagging algorithm ($b$-tagged jets or $b$-jets for short). Neural networks (NN) are employed in the regions with a significant expected contribution from the $t\bar{t}H$ signal to separate it from the background. Simpler kinematic variables are used in regions that are depleted of the $t\bar{t}H$ signal, and primarily serve to constrain uncertainties on the background prediction. A combined fit to signal-rich and signal-depleted regions is performed to search for the signal while simultaneously obtaining a background prediction.

2 ATLAS detector

The ATLAS detector [25] consists of four main subsystems: an inner tracking system, electromagnetic and hadronic calorimeters, and a muon spectrometer. The inner detector provides tracking information from pixel and silicon microstrip detectors in the pseudorapidity.\(^1\)

\(^1\) 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 coinciding with the axis of the beam pipe. The $x$-axis points from the IP to the centre of the LHC ring, and the $y$-axis points upward. Cylindrical coordinates ($r,\phi$) are used in the transverse plane, $\phi$ being the azimuthal angle around the beam pipe. The pseudorapidity is defined in terms of the polar angle $\theta$ as $\eta = -\ln \tan(\theta/2)$. Transverse momentum and energy are defined as $p_T = p \sin \theta$ and $E_T = E \sin \theta$, respectively.

3 Object reconstruction

The main physics objects considered in this search are electrons, muons, jets and $b$-jets.

Electron candidates [26] are reconstructed from energy deposits (clusters) in the electromagnetic calorimeter that are matched to a reconstructed track in the inner detector. To reduce the background from non-prompt electrons, i.e. from decays of hadrons (in particular heavy flavour) produced in jets, electron candidates are required to be isolated. In the single-lepton channel, where such background is significant, an $\eta$-dependent isolation cut is made, based on the sum of transverse energies of cells around the direction of each candidate, in a cone of size $\Delta R = \sqrt{(\Delta \phi)^2 + (\Delta \eta)^2} = 0.2$. This energy sum excludes cells associated with the beam pipe. The $x$-axis points from the IP to the centre of the LHC ring, and the $y$-axis points upward. Cylindrical coordinates ($r,\phi$) are used in the transverse plane, $\phi$ being the azimuthal angle around the beam pipe. The pseudorapidity is defined in terms of the polar angle $\theta$ as $\eta = -\ln \tan(\theta/2)$. Transverse momentum and energy are defined as $p_T = p \sin \theta$ and $E_T = E \sin \theta$, respectively.
electron and is corrected for leakage from the electron cluster itself. A further isolation cut is made on the scalar sum of the track $p_T$ around the electron in a cone of size $\Delta R = 0.3$ (referred to as $p_T^{\text{cone30}}$). The longitudinal impact parameter of the electron track with respect to the selected event primary vertex, $z_0$, is required to be less than 2 mm. To increase efficiency in the dilepton channel, the electron selection is optimised by using an improved electron identification method based on a likelihood variable \cite{27} and the electron isolation. The ratio of $p_T^{\text{cone30}}$ to the $p_T$ of the electron is required to be less than 0.12, i.e. $p_T^{\text{cone30}}/p_T < 0.12$. The optimised selection improves the efficiency by roughly 7% per electron.

Muon candidates are reconstructed from track segments in the muon spectrometer, and matched with tracks found in the inner detector \cite{28}. The final muon candidates are refitted using the complete track information from both detector systems, and are required to satisfy $|y| < 2.5$. Additionally, muons are required to be separated by $\Delta R > 0.4$ from any selected jet (see below for details on jet reconstruction and selection). Furthermore, muons must satisfy a $p_T$-dependent track-based isolation requirement that has good performance under conditions with a high number of jets from other $pp$ interactions within the same bunch crossing, known as “pileup”, or in boosted configurations where the muon interactions within the same bunch crossing, known as “pileup”, or in boosted configurations where the muon is close to a jet: the track $p_T$ scalar sum in a cone of variable size $\Delta R < 10 \text{ GeV}/p_T$ around the muon must be less than 5% of the muon $p_T$. The longitudinal impact parameter of the muon track with respect to the primary vertex, $z_0$, is required to be less than 2 mm.

Jets are reconstructed from calibrated clusters \cite{25} built from energy deposits in the calorimeters, using the anti-$k_t$ algorithm \cite{30,32} with a radius parameter $R = 0.4$. Prior to jet finding, a local cluster calibration scheme \cite{33,34} is applied to correct the cluster energies for the effects of dead material, non-compensation and out-of-cluster leakage. The jets are calibrated using energy- and $\eta$-dependent calibration factors, derived from simulations, to the mean energy of stable particles inside the jets. Additional corrections to account for the difference between simulation and data are applied \cite{35}. After energy calibration, jets are required to have $p_T > 25 \text{ GeV}$ and $|y| < 2.5$. To reduce the contamination due to pileup jets, the scalar sum of the $p_T$ of tracks matched to the jet and originating from the primary vertex must be at least 50% of the scalar sum of the $p_T$ of all tracks matched to the jet. This is referred to as the jet vertex fraction. This criterion is only applied to jets with $p_T < 50 \text{ GeV}$ and $|\eta| < 2.4$.

During jet reconstruction, no distinction is made between identified electrons and jet candidates. Therefore, if any of the jets lie $\Delta R < 0.2$ from a selected electron, the single closest jet is discarded in order to avoid double-counting of electrons as jets. After this, electrons which are $\Delta R < 0.4$ from a jet are removed to further suppress background from non-isolated electrons.

Jets are identified as originating from the hadronisation of a $b$-quark via an algorithm \cite{36} that uses multivariate techniques to combine information from the impact parameters of displaced tracks with topological properties of secondary and tertiary decay vertices reconstructed within the jet. The working point used for this search corresponds to a 70% efficiency to tag a $b$-quark jet, with a light-jet mistag rate of 1%, and a charm-jet mistag rate of 20%, as determined for $b$-tagged jets with $p_T > 20 \text{ GeV}$ and $|y| < 2.5$ in simulated $tt$ events. Tagging efficiencies in simulation are corrected to match the results of the calibrations performed in data \cite{37}.

4 Event selection and classification

For this search, only events collected using a single-electron or single-muon trigger under stable beam conditions and for which all detector subsystems were operational are considered. The corresponding integrated luminosity is 20.3 fb$^{-1}$. Triggers with different $p_T$ thresholds are combined in a logical OR in order to maximise the overall efficiency. The $p_T$ thresholds are 24 or 60 GeV for electrons and 24 or 36 GeV for muons. The triggers with the lower $p_T$ threshold include isolation requirements on the lepton candidate, resulting in inefficiency at high $p_T$ that is recovered by the triggers with higher $p_T$ threshold. The triggers use selection criteria looser than the final reconstruction requirements.

Events accepted by the trigger are required to have at least one reconstructed vertex with at least five associated tracks, consistent with the beam collision region in the $x$–$y$ plane. If more than one such vertex is found, the vertex candidate with the largest sum of squared transverse momenta of its associated tracks is taken as the hard-scatter primary vertex.

In the single-lepton channel, events are required to have exactly one identified electron or muon with $p_T > 25 \text{ GeV}$ and at least four jets, at least two of which are $b$-tagged. The selected lepton is required to match, with $\Delta R < 0.15$, the lepton reconstructed by the trigger.

In the dilepton channel, events are required to have exactly two leptons of opposite charge and at least two $b$-jets. The leading and subleading lepton must have $p_T > 25 \text{ GeV}$ and $p_T > 15 \text{ GeV}$, respectively. Events in the single-lepton sample with additional leptons passing
this selection are removed from the single-lepton sample to avoid statistical overlap between the channels. In the dilepton channel, events are categorised into $ee$, $\mu\mu$ and $e\mu$ samples. In the $e\mu$ category, the scalar sum of the transverse energy of leptons and jets, $H_T$, is required to be above 130 GeV. In the $ee$ and $\mu\mu$ event categories, the invariant mass of the two leptons, $m_{\ell\ell}$, is required to be larger than 15 GeV in events with more than two $b$-jets, to suppress contributions from the decay of hadronic resonances such as the $J/\psi$ and $\Upsilon$ into a same-flavour lepton pair. In events with exactly two $b$-jets, $m_{\ell\ell}$ is required to be larger than 60 GeV due to poor agreement between data and prediction at lower $m_{\ell\ell}$. A further cut on $m_{\ell\ell}$ is applied in the $ee$ and $\mu\mu$ categories to reject events close to the $Z$ boson mass: $|m_{\ell\ell} - m_Z| > 8$ GeV.

After all selection requirements, the samples are dominated by $t\bar{t}+jets$ background. In both channels, selected events are categorised into different regions. In the following, a given region with $m$ jets of which $n$ are $b$-jets are referred to as “($m_j,nb$)”. The regions with a signal-to-background ratio $S/B > 1\%$, where $S$ and $B$ denote the expected signal for a SM Higgs boson with $m_H = 125$ GeV, and background, respectively, and $S/\sqrt{B} > 0.3$ are referred to as “signal-rich regions”, as they provide most of the sensitivity to the signal. The remaining regions are referred to as “signal-depleted regions”. They are almost purely background-only regions and are used to constrain systematic uncertainties, thus improving the background prediction in the signal-rich regions. The regions are analysed separately and combined statistically to maximise the overall sensitivity. In the most sensitive regions, ($\geq 6j,\geq 4b$) in the single-lepton channel and ($\geq 4j,\geq 4b$) in the dilepton channel, $H \rightarrow b\bar{b}$ decays are expected to constitute about 90% of the signal contribution.

In the single-lepton channel, a total of nine independent regions are considered: six signal-depleted regions, ($4j,2b$), ($4j,3b$), ($4j,4b$), ($5j,2b$), ($5j,3b$), ($\geq 6j,2b$), and three signal-rich regions, ($5j,\geq 4b$), ($\geq 6j,3b$) and ($\geq 6j,\geq 4b$). In the dilepton channel, a total of six independent regions are considered. The signal-rich regions are ($\geq 4j,3b$) and ($\geq 4j,\geq 4b$), while the signal-depleted regions include ($2j,2b$), ($3j,2b$), ($3j,3b$) and ($\geq 4j,2b$). Figure 2(a) shows the $S/\sqrt{B}$ and $S/B$ ratios for the different regions under consideration in the single-lepton channel based on the simulations described in Sect. 5. The expected proportions of different backgrounds in each region are shown in Fig. 2(b). The same is shown in the dilepton channel in Figs. 3(a) and 3(b).

5 Background and signal modelling

After event preselection, the main background in both the single-lepton and dilepton channels is $t\bar{t}+jets$ production. In the single-lepton channel, additional background contributions come from single top quark production, followed by the production of a $W$ or $Z$ boson in association with jets ($W/Z+Jets$), diboson ($WW$, $WZ$, $ZZ$) production, as well as the associated production of a vector boson and a $t\bar{t}$ pair, $t\bar{t}+V$ ($V=W,Z$). Multijet events also contribute to the selected sample via the misidentification of a jet or a photon as an electron or the presence of a non-prompt lepton, referred to as “Lepton misID” background. The corresponding yield is estimated via a data-driven method known as the “matrix method” [38]. In the dilepton channel, backgrounds containing at least two prompt leptons other than $t\bar{t}+jets$ production arise from $Z+Jets$, diboson, and $Wt$-channel single top quark production, as well as from the $t\bar{t}V$ processes. There are also several processes which may contain either non-prompt leptons that pass the lepton isolation requirements or jets misidentified as leptons. These processes include $W+Jets$, $t\bar{t}$ production with a single prompt lepton in the final state, and single top quark production in $t$- and $s$-channels. Their yield is estimated using simulation and cross-checked with a data-driven technique based on the selection of a same-sign lepton pair. In both channels, the contribution of the misidentified lepton background is negligible after requiring two $b$-tagged jets.

In the following, the simulation of each background and of the signal is described in detail. For all MC samples, the top quark mass is taken to be $m_t = 172.5$ GeV and the Higgs boson mass is taken to be $m_H = 125$ GeV.

5.1 $t\bar{t}+jets$ background

The $t\bar{t}+jets$ sample is generated using the POWHEG-Box 3.0 NLO generator [39][41] with the CT10 parton distribution function (PDF) set [42]. It is interfaced to PYTHIA 6.425 [43] with the CTEQ6L1 PDF set [44] and the Perugia2011C [45] underlying-event tune. The sample is normalised to the top++2.0 [46] theoretical calculation performed at next-to-next-to-leading order (NNLO) in QCD that includes resummation of next-to-next-to-leading logarithmic (NNLL) soft gluon terms [47][51].

The $t\bar{t}+jets$ sample is generated inclusively, but events are categorised depending on the flavour of partons that are matched to particle jets that do not originate from the decay of the $t\bar{t}$ system. The matching procedure is done using the requirement of $\Delta R < 0.4$. Particle jets
**ATLAS Simulation**

\( \sqrt{s} = 8 \text{ TeV}, 20.3 \text{ fb}^{-1} \)

**Single lepton**

\( m_{\ell} = 125 \text{ GeV} \)

---

**Dilepton channel**

- (a) The plots for a specific jet multiplicity (4, 5, \( \geq 6 \)), and the columns show the 6-jet multiplicity (2, 3, \( \geq 4 \)). Signal-rich regions are shaded in dark red, while the rest are shown in light blue. The \( S/B \) ratio for each region is also noted. (b) The fractional contributions of the various backgrounds to the total background prediction in each considered region. The ordering of the rows and columns is the same as in (a).

---

**ATLAS Simulation**

\( \sqrt{s} = 8 \text{ TeV} \)

**Dilepton**

\( m_{\ell} = 125 \text{ GeV} \)

---

**Fig. 3** Dilepton channel: (a) The \( S/\sqrt{B} \) ratio for each of the regions assuming SM cross sections and branching fractions and \( m_{H} = 125 \text{ GeV} \). Each row shows the plots for a specific jet multiplicity (2, 3, \( \geq 4 \)), and the columns show the 6-jet multiplicity (2, 3, \( \geq 4 \)). Signal-rich regions are shaded in dark red, while the rest are shown in light blue. The \( S/B \) ratio for each region is also noted. (b) The fractional contributions of the various backgrounds to the total background prediction in each considered region. The ordering of the rows and columns is the same as in (a).
are reconstructed by clustering stable particles excluding muons and neutrinos using the anti-\( k_T \) algorithm with a radius parameter \( R = 0.4 \), and are required to have \( p_T > 15 \) GeV and \( |\eta| < 2.5 \).

Events where at least one such particle jet is matched to a bottom quark not originating from a top quark decay are labelled as \( tt+bb \) events. Similarly, events which are not already categorised as \( tt+bb \), and where at least one particle jet is matched to a charm quark not originating from a \( W \) boson decay, are labelled as \( tt+c\bar{c} \) events. Events labelled as either \( tt+bb \) or \( tt+c\bar{c} \) are generically referred to as \( tt+HF \) events (HF for “heavy flavour”). The remaining events are labelled as \( tt+light-jet \) events, including those with no additional jets.

Since Powheg only models \( tt+bb \) via the parton shower, an alternative \( tt+\)jets sample is generated with the MadGraph5 1.5.11 LO generator \cite{72} using the CT10 PDF set and interfaced to Pythia 6.425 for showering and hadronisation. It includes tree-level diagrams with up to three extra partons (including \( b \)- and \( c \)-quarks) and uses settings similar to those in Ref. \cite{24}. To avoid double-counting of partonic configurations generated by both the matrix element calculation and the parton-shower evolution, a parton–jet matching scheme (“MLM matching”) \cite{53} is employed.

Fully matched NLO predictions with massive \( b \)-quarks have become available recently \cite{54} within the Sherpa with OpenLoops framework \cite{55,56} referred to in the following as SherpaOL. The SherpaOL NLO sample is generated following the four-flavour scheme using the Sherpa 2.0 pre-release and the CT10 PDF set. The renormalisation scale \( (\mu_R) \) is set to \( \mu_R = \prod_{i=t,b} E_{T,i}^{1/4} \), where \( E_{T,i} \) is the transverse energy of parton \( i \), and the factorisation and resummation scales are both set to \( (E_{T,t} + E_{T,b})/2 \).

For the purpose of comparisons between \( tt+\)jets event generators and the propagation of systematic uncertainties related to the modelling of \( tt+HF \), as described in Sects. 8.3.1, a finer categorisation of different topologies in \( tt+HF \) is made. In particular, the following categories are considered: if two particle jets are both matched to an extra \( b \)-quark or extra \( c \)-quark each, the event is referred to as \( tt+bb \) or \( tt+c\bar{c} \); if a single particle jet is matched to a single \( b(c) \)-quark the event is referred to as \( tt+b(Tt+c) \); if a single particle jet is matched to a \( bb \) or a \( c\bar{c} \) pair, the event is referred to as \( tt+B \) or \( tt+C \), respectively.

Figure 4 shows the relative contributions of the different \( tt+bb \) event categories to the total \( tt+bb \) cross section at generator level for the Powheg+Pythia, Madgraph+Pythia and SherpaOL samples. It demonstrates that Powheg+Pythia is able to reproduce reasonably well the \( tt+HF \) content of the Madgraph \( tt+\)jets sample, which includes a LO \( tt+bb \) matrix element calculation, as well as the NLO SherpaOL prediction.

The relative distribution across categories is such that SherpaOL predicts a higher contribution of the \( tt+B \) category, as well as every category where the production of a second \( bb \) pair is required. The modelling of the relevant kinematic variables in each category is in reasonable agreement between Powheg+Pythia and SherpaOL. Some differences are observed in the very low regions of the mass and \( p_T \) of the \( bb \) pair, and in the \( p_T \) of the top quark and \( tt \) systems.

The prediction from SherpaOL is expected to model the \( tt+bb \) contribution more accurately than both Powheg+Pythia and Madgraph+Pythia. Thus, in the analysis \( tt+bb \) events are reweighted from Powheg+Pythia to reproduce the NLO \( tt+bb \) prediction from SherpaOL for relative contributions of different categories as well as their kinematics. The reweighting is done at generator level using several kinematic variables such as the top quark \( p_T \), \( tt \) system \( p_T \), \( \Delta R \) and \( p_T \) of the dijet system not coming from the top quark decay. In the absence of an NLO calculation of \( tt+c\bar{c} \) production, the Madgraph+Pythia sample is used.

Fig. 4 Relative contributions of different categories of \( tt+bb \) events in Powheg+Pythia, Madgraph+Pythia and SherpaOL samples. Labels “\( tt+MPI \)” and “\( tt+FSR \)” refer to events where heavy flavour is produced via multiparton interaction (MPI) or final state radiation (FSR), respectively. These contributions are not included in the SherpaOL calculation. An arrow indicates that the point is off-scale. Uncertainties are from the limited MC sample sizes.
to evaluate systematic uncertainties on the $t\bar{t}+c\bar{c}$ background.

Since achieving the best possible modelling of the $t\bar{t}+\text{jets}$ background is a key aspect of this analysis, a separate reweighting is applied to $t\bar{t}+\text{light}$ and $t\bar{t}+c\bar{c}$ events in POWHEG+PYTHIA based on the ratio of measured differential cross sections at $\sqrt{s} = 7$ TeV in data and simulation as a function of top quark $p_T$ and $t\bar{t}$ system $p_T$ [57]. It was verified using the simulation that the ratio derived at $\sqrt{s} = 7$ TeV is applicable to $\sqrt{s} = 8$ TeV simulation. It is not applied to the $t\bar{t}+bb$ component since that component was corrected to match the best available theory calculation. Moreover, the measured differential cross section is not sensitive to this component. The reweighting significantly improves the agreement between simulation and data in the total number of jets (primarily due to the $t\bar{t}$ system $p_T$ reweighting) and jet $p_T$ (primarily due to the top quark $p_T$ reweighting). This can be seen in Fig. 5 where the number of jets and the scalar sum of the jet $p_T$ ($H^\text{had}_T$) distributions in the exclusive 2-tag region are plotted in the single-lepton channel before and after the reweighting is applied.

5.2 Other backgrounds

The $W/Z+\text{jets}$ background is estimated from simulation reweighted to account for the difference in the $W/Z$ $p_T$ spectrum between data and simulation [58]. The heavy-flavour fraction of these simulated backgrounds, i.e. the sum of $W/Z+bb$ and $W/Z+c\bar{c}$ processes, is adjusted to reproduce the relative rates of $Z$ events with no $b$-tags and those with one $b$-tag observed in data. Samples of $W/Z+\text{jets}$ events, and diboson production in association with jets, are generated using the ALPGEN 2.14 [59] leading-order (LO) generator and the CTEQ6L1 PDF set. Parton showers and fragmentation are modelled with PYTHIA 6.425 for $W/Z+\text{jets}$ production and with HERWIG 6.520 [60] for diboson production. The $W+\text{jets}$ samples are generated with up to five additional partons, separately for $W+\text{light}$-jets, $Wb\bar{b}+\text{jets}$, $Wc\bar{c}+\text{jets}$, and $W+c\bar{c}$-jets. Similarly, the $Z+\text{jets}$ background is generated with up to five additional partons separated in different parton flavours.

Both are normalised to the respective inclusive NNLO theoretical cross section [61]. The overlap between $WQQ$ ($ZQQ$) ($Q = b, c$) events generated from the matrix element calculation and those from parton-shower evolution in the $W+\text{light}$-jet ($Z+\text{light}$-jet) samples is removed by an algorithm based on the angular separation between the extra heavy quarks: if $\Delta R(Q, \bar{Q}) > 0.4$, the matrix element prediction is used, otherwise the parton shower prediction is used.

The diboson+jets samples are generated with up to three additional partons and are normalised to their respective NLO theoretical cross sections [62].

Samples of single top quark backgrounds are generated with POWHEG-BOX 3.0 using the CT10 PDF set. The samples are interfaced to PYTHIA 6.425 with the CTEQ6L1 set of parton distribution functions and Perugia2011C underlying-event tune. Overlaps between the $t\bar{t}$ and $Wt$ final states are removed [63]. The single top quark samples are normalised to the approximate NNLO theoretical cross sections [64,66] using the MSTW2008 NNLO PDF set [67,68].

Samples of $t\bar{t}+V$ are generated with MADGRAPH 5 and the CTEQ6L1 PDF set. PYTHIA 6.425 with the AUET2B tune [69] is used for showering. The $t\bar{t}V$ samples are normalised to the NLO cross-section predictions [70,71].

5.3 Signal model

The $t\bar{t}H$ signal process is modelled using NLO matrix elements obtained from the HELAC-Oneloop package [72]. POWHEG-BOX serves as an interface to shower Monte Carlo programs. The samples created using this approach are referred to as PowHel samples. They are inclusive in Higgs boson decays and are produced using the CT10NLO PDF set and factorisation ($\mu_F$) and renormalisation scales set to $\mu_T = \mu_R = m_t + m_H/2$.

The PowHel $t\bar{t}H$ sample is showered with PYTHIA 8.1 [73] with the CTEQ6L1 PDF and the AU2 underlying-event tune [74]. The $t\bar{t}H$ cross section and Higgs boson decay branching fractions are taken from (N)NLO theoretical calculations [19,75–81], collected in Ref. [82].

In Appendix A the relative contributions of the Higgs boson decay modes are shown for all regions considered in the analysis.

5.4 Common treatment of MC samples

All samples using HERWIG are also interfaced to JIMMY 4.31 [83] to simulate the underlying event. All simulated samples utilise PHOTOS 2.15 [84] to simulate photon radiation and TAUOLA 1.20 [85] to simulate $\tau$ decays.

Events from minimum-bias interactions are simulated with the PYTHIA 8.1 generator with the MSTW2008 LO PDF set and the AUET2 [86] tune. They are superimposed on the simulated MC events, matching the luminosity profile of the recorded data. The contributions from these pileup interactions are simulated both within the same bunch crossing as the hard-scattering process and in neighbouring bunch crossings.
Finally, all simulated MC samples are processed through a simulation of the detector geometry and response either using Geant4 [88], or through a fast simulation of the calorimeter response [89]. All simulated MC samples are processed through the same reconstruction software as the data. Simulated MC events are corrected so that the object identification efficiencies, energy scales and energy resolutions match those determined from data control samples.

Figures 6(a) and 6(b) show a comparison of predicted yields to data prior to the fit described in Sect. 9 in all analysis regions in the single-lepton and dilepton channel, respectively. The data agree with the SM expectation within the uncertainties of 10-30%. Detailed tables of the event yields prior to the fit for the single-lepton and dilepton channels can be found in Appendix 17.

When requiring high jet and b-tag multiplicity in the analysis, the number of available MC events is significantly reduced, leading to large fluctuations in the resulting distributions for certain samples. This can negatively affect the sensitivity of the analysis through the
large statistical uncertainties on the templates and unreliable systematic uncertainties due to shape fluctuations. In order to mitigate this problem, instead of tagging the jets by applying the $b$-tagging algorithm, their probabilities to be $b$-tagged are parameterised as functions of jet flavour, $p_T$, and $\eta$. This allows all events in the sample before $b$-tagging is applied to be used in predicting the normalisation and shape after $b$-tagging \cite{90}. This prediction agrees well with the normalisation and shape obtained by applying the $b$-tagging algorithm directly.

6 Analysis method

In both the single-lepton and dilepton channels, the analysis uses a neural network (NN) to discriminate signal from background in each of the regions with significant expected $t\bar{t}H$ signal contribution since the $S/\sqrt{B}$ is very small and the uncertainty on the background is larger than the signal. Those include (5j, ≥4b), (≥6j, 3b) and (≥6j, ≥4b) in the case of the single-lepton channel, and (≥4j, 3b) and (≥4j, ≥4b) in the case of the dilepton channel. In the dilepton channel, an additional NN is used to separate signal from background in the (3j, 3b) channel. Despite a small expected $S/\sqrt{B}$, it nevertheless adds sensitivity to the signal due to a relatively high expected $S/B$. In the single-lepton channel, a dedicated NN is used in the (5j, 3b) region to separate $t\bar{t}+\text{light}$ from $t\bar{t}+\text{HF}$ backgrounds. The other regions considered in the analysis have lower sensitivity, and use $H_T^{\text{had}}$ in the single-lepton channel, and the scalar sum of the jet and lepton $p_T$ ($H_T$) in the dilepton channel as a discriminant.

The NNs used in the analysis are built using the NeuroBayes \cite{91} package. The choice of the variables that enter the NN discriminant is made through the ranking procedure implemented in this package based on the statistical separation power and the correlation of variables. Several classes of variables were considered: object kinematics, global event variables, event shape variables and object pair properties. In the regions with ≥6 (≥4) jets, a maximum of seven (five) jets are considered to construct the kinematic variables in the single-lepton (dilepton) channel, first using all the $b$-jets, and then incorporating the untagged jets with the highest $p_T$. All variables used for the NN training and their pairwise correlations are required to be described well in simulation in multiple control regions.

In the (5j, 3b) region in the single-lepton channel, the separation between the $t\bar{t}+\text{light}$ and $t\bar{t}+\text{HF}$ events is achieved by exploiting the different origin of the third $b$-jet in the case of $t\bar{t}+\text{light}$ compared to $t\bar{t}+\text{HF}$ events. In both cases, two of the $b$-jets originate from the $t\bar{t}$ decay. However, in the case of $t\bar{t}+\text{HF}$ events, the third $b$-jet is likely to originate from one of the additional heavy-flavour quarks, whereas in the case of $t\bar{t}+\text{light}$ events, the third $b$-jet is often matched to a $c$-quark from the hadronically decaying $W$ boson. Thus, kinematic variables, such as the invariant mass of the two untagged jets with minimum $\Delta R$, provide discrimination between $t\bar{t}+\text{light}$ and $t\bar{t}+\text{HF}$ events, since the lat-
ter presents a distinct peak at the $W$ boson mass which is not present in the former. This and other kinematic variables are used in the dedicated NN used in this region.

In addition to the kinematic variables, two variables calculated using the matrix element method (MEM), detailed in Sect. 2 are included in the NN training in ($\geq 6j, 3b$) and ($\geq 6j, \geq 4b$) regions of the single-lepton channel. These two variables are the Neyman-Pearson likelihood ratio ($D1$) (Eq. (4)) and the logarithm of the summed signal likelihoods (SSLL) (Eq. (2)). The $D1$ variable provides the best separation between $t\bar{t}H$ signal and the dominant $t\bar{t}+bb$ background in the ($\geq 6j, \geq 4b$) region. The SSLL variable further improves the NN performance.

The variables used in the single-lepton and dilepton channels, as well as their ranking in each analysis region, are listed in Tables 1 and 2, respectively. For the construction of variables in the ($\geq 4j, \geq 4b$) region of the dilepton channel, the two $b$-jets that are closest to the leptons are considered to originate from the top quarks, and the other two $b$-jets are assigned to the Higgs candidate.

Figures 7 and 8 show the distribution of the NN discriminant for the $t\bar{t}H$ signal and background in the single-lepton and dilepton channels, respectively, in the signal-rich regions. In particular, Fig. 8(a) shows the separation between the $t\bar{t}+HF$ and $t\bar{t}+light$-jet production achieved by a dedicated NN in the $(5j, 3b)$ region in the single-lepton channel. The distributions in the highest-ranked input variables from each of the NN regions are shown in Appendix C.

For all analysis regions considered in the fit, the $t\bar{t}H$ signal includes all Higgs decay modes. They are also included in the NN training.

The analysis regions have different contributions from various systematic uncertainties, allowing the combined fit to constrain them. The highly populated ($4j, 2b$) and ($2j, 2b$) regions in the single-lepton and dilepton channels, respectively, provide a powerful constraint on the overall normalisation of the $t\bar{t}H$ background. The ($4j, 2b$), ($5j, 2b$) and ($\geq 6j, 2b$) regions in the single-lepton channel and the ($2j, 2b$), ($3j, 2b$) and ($\geq 4j, 2b$) regions in the dilepton channel are almost pure in $t\bar{t}+light$-jets background and provide an important constraint on $t\bar{t}$ modelling uncertainties both in terms of normalisation and shape. Uncertainties on $c$-tagging are reduced by exploiting the large contribution of $W \rightarrow cs$ decays in the $t\bar{t}+light$-jets background populating the ($4j, 3b$) region in the single-lepton channel. Finally, the consideration of regions with exactly 3 and $\geq 4$ $b$-jets in both channels, having different fractions of $t\bar{t}+bb$ and $t\bar{t}+c\bar{c}$ backgrounds, provides the ability to constrain uncertainties on the $t\bar{t}+bb$ and $t\bar{t}+c\bar{c}$ normalisations.

### 7 The matrix element method

The matrix element method [53] has been used by the D0 and CDF collaborations for precision measurements of the top quark mass [44,45] and for the observations of single top quark production [66,67]. Recently this technique has been used for the $t\bar{t}H$ search by the CMS experiment [68]. By directly linking theoretical calculations and observed quantities, it makes the most complete use of the kinematic information of a given event.

The method calculates the probability density function of an observed event to be consistent with physics process $i$ described by a set of parameters $\alpha$. This probability density function $P_i(x|\alpha)$ is defined as

$$P_i(x|\alpha) = \frac{(2\pi)^4}{\sigma^{exp}_i(\alpha)} \int dp_A dp_B f(p_A) f(p_B) \frac{|M_i(y|\alpha)|^2}{\mathcal{F}} W(y|x) d\Phi_N(y)$$

and is obtained by numerical integration over the entire phase space of the initial- and final-state particles. In this equation, $x$ and $y$ represent the four-momentum vectors of all final-state particles at reconstruction and parton level, respectively. The flux factor $\mathcal{F}$ and the Lorentz-invariant phase space element $d\Phi_N$ describe the kinematics of the process. The transition matrix element $M_i$ is defined by the Feynman diagrams of the hard process. The transfer functions $W(y|x)$ map the detector quantities $x$ to the parton level quantities $y$. Finally, the cross section $\sigma^{exp}_i$ normalises $P_i$ to unity taking acceptance and efficiency into account.

The assignment of reconstructed objects to final-state partons in the hard process contains multiple ambiguities. The process probability density is calculated for each allowed assignment permutation of the jets to the final-state quarks of the hard process. A process likelihood function can then be built by summing the process probabilities for the $N_p$ allowed assignment permutations,

$$\mathcal{L}_i(x|\alpha) = \sum_{p=1}^{N_p} P^P_i(x|\alpha).$$

The process probability densities are used to distinguish signal from background events by calculating the
Table 1: Single-lepton channel: the definitions and rankings of the variables considered in each of the regions where an NN is used.

| Variable | Definition | NN rank |
|----------|------------|---------|
| $D1$ | Neyman–Pearson MEM discriminant (Eq. (4)) | $\geq 6j, \geq 4b$ |
| Centrality | Scalar sum of the $p_T$ divided by sum of the $E$ for all jets and the lepton | $\geq 6j, \geq 4b$ |
| $p_T^{\text{jet5}}$ | $p_T$ of the fifth leading jet | $\geq 6j, \geq 4b$ |
| $H1$ | Second Fox–Wolfram moment computed using all jets and the lepton | $\geq 6j, \geq 4b$ |
| $\Delta R_{\text{avg}}^{\text{bb}}$ | Average $\Delta R$ for all $b$-tagged jet pairs | $\geq 6j, 3b$ |
| SSLL | Logarithm of the summed signal likelihoods (Eq. (2)) | $\geq 6j, 3b$ |
| $m_{\text{min}} \Delta R$ | Mass of the combination of the two $b$-tagged jets with the smallest $\Delta R$ | - |
| $m_{\text{max}} \ p_T$ | Mass of the combination of a $b$-tagged jet and any jet with the largest vector sum $p_T$ | - |
| $\Delta R_{\text{max}}^{\text{pr}}$ | $\Delta R$ between the two $b$-tagged jets with the largest vector sum $p_T$ | - |
| $\Delta R_{\text{min}} \Delta R_{\text{lep–bb}}$ | $\Delta R$ between the lepton and the combination of the two $b$-tagged jets with the smallest $\Delta R$ | - |
| $m_{\text{min}} \Delta R$ | Mass of the combination of the two untagged jets with the smallest $\Delta R$ | - |
| Aplan$_{\text{b–jet}}$ | $1.5\lambda_2$, where $\lambda_2$ is the second eigenvalue of the momentum tensor built with only $b$-tagged jets | - |
| $N_{\text{jet}}^{40}$ | Number of jets with $p_T \geq 40$ GeV | - |
| $m_{\text{min}} \Delta R$ | Mass of the combination of a $b$-tagged jet and any jet with the smallest $\Delta R$ | - |
| $m_{\text{max}} \ p_T$ | Mass of the combination of any two jets with the largest vector sum $p_T$ | - |
| $m_{\text{had}}^{\text{jet}}$ | Scalar sum of jet $p_T$ | - |
| $m_{\text{min}} \Delta R$ | Mass of the combination of any two jets with the smallest $\Delta R$ | - |
| $m_{\text{max}} \ p_T$ | Mass of the combination of the two $b$-tagged jets with the largest vector sum $p_T$ | - |
| $m_{\text{bb}}$ | Scalar sum of the $p_T$ of the pair of untagged jets with the smallest $\Delta R$ | - |
| $m_{\text{max}} \ m_{\text{bb}}$ | Mass of the combination of the two $b$-tagged jets with the largest invariant mass | - |
| $\Delta R_{\text{min}}^{\text{bb}}$ | Minimum $\Delta R$ between the two untagged jets | - |
| $m_{\text{min}} \Delta R$ | Mass of the jet triplet with the largest vector sum $p_T$ | - |

Table 1: Single-lepton channel: the definitions and rankings of the variables considered in each of the regions where an NN is used.

The likelihood ratio of the signal and background processes contributing with fractions $f_{\text{bkg}}$, 

$$r_{\text{sig}}(x|\alpha) = \frac{L_{\text{sig}}(x|\alpha)}{\sum_{\text{bkg}} f_{\text{bkg}} L_{\text{bkg}}(x|\alpha)}$$

(3)

This ratio, according to the Neyman–Pearson lemma, is the most powerful discriminant between signal and background processes. In the analysis, this variable is used as input to the NN along with other kinematic variables.

Matrix element calculation methods are generated with MADGRAPH 5 in LO. The transfer functions are obtained from simulation following a similar procedure as described in Ref. [100]. For the modelling of the parton distribution functions the CTEQ6L1 set from the LHAPDF package [101] is used.

The integration is performed using VEGAS [102]. Due to the complexity and high dimensionality, adaptive MC techniques [103], simplifications and approximations are needed to obtain results within a reasonable computing time. In particular, only the numerically most significant contributing helicity states of a process hypothesis for a given event, identified at the start of each integration, are evaluated. This does not perceptibly decrease the separation power but reduces
| Variable         | Definition                                                                 | NN rank |  
|------------------|-----------------------------------------------------------------------------|---------|
| $\Delta_\eta_{\text{max}}$ | Maximum $\Delta\eta$ between any two jets in the event                      | $\geq 4j, \geq 4b$ |  
| $m_{\text{bb}}$   | Mass of the combination of the two $b$-tagged jets with the smallest $\Delta R$ | 2       |
| $m_{\text{bb}}$   | Mass of the two $b$-tagged jets from the Higgs candidate system             | 3       |
| $\Delta R_{\text{min}}$ | $\Delta R$ between the Higgs candidate and the closest lepton              | $\geq 4j, 3b$ |  
| $N_{\text{Higgs}}$ | Number of Higgs candidates within 30 GeV of the Higgs mass of 125 GeV     | 5       |
| $\Delta R_{\text{max}}$ | $\Delta R$ between the two $b$-tagged jets with the largest vector sum $p_T$ | $\geq 4j, 3b$ |  
| $A_{\text{planjet}}$ | $1.5\lambda_2$, where $\lambda_2$ is the second eigenvalue of the momentum tensor built with all jets | 7       |
| $m_{\text{closest}}$ | Minimum dijet mass between any two jets                                    | 8       |
| $H_T$             | Dijet mass between any two jets closest to the Higgs                        | 9       |
| $\Delta R_{\text{max}}$ | $\Delta R$ between the two $b$-tagged jets with the largest invariant mass | $\geq 4j, 3b$ |  
| $m_{\text{closest}}$ | Minimum $\Delta R$ between any two jets and the furthest lepton            | -       |
| $H_T$             | Dijet mass between any two jets closest to the Higgs                        | 10      |
| $m_{\text{closest}}$ | Sum of the $p_T$ divided by sum of the $E$ for all jets and both leptons    | -       |
| $m_{\text{closest}}$ | Mass of the combination of any two jets with the largest vector sum $p_T$ | -       |
| $H_4$             | Fifth Fox–Wolfram moment computed using all jets and both leptons          | -       |
| $p_T^{\text{jet3}}$ | $p_T$ of the third leading jet                                             | -       |

Table 2: Dilepton channel: the definitions and rankings of the variables considered in each of the regions where an NN is used.

the calculation time by more than an order of magnitude. Furthermore, several approximations are made to improve the VEGAS convergence rate. Firstly, the dimensionality of integration is reduced by assuming that the final-state object directions in $\eta$ and $\phi$ as well as charged lepton momenta are well measured, and therefore the corresponding transfer functions are represented by $\delta$ functions. The total momentum conservation and a negligible transverse momentum of the initial-state partons allow for further reduction. Secondly, kinematic transformations are utilised to optimise the integration over the remaining phase space by aligning the peaks of the integrand with the integration dimensions. The narrow-width approximation is applied to the leptonically decaying $W$ boson. This leaves three $b$-quark energies, one light-quark energy, the hadronically decaying $W$ boson mass and the invariant mass of the two $b$-quarks originating from either the Higgs boson for the signal or a gluon for the background as the remaining parameters which define the integration phase space. The total integration volume is restricted based upon the observed values and the width of the transfer functions and of the propagator peaks in the matrix elements. Finally, the likelihood contributions of all allowed assignment permutations are coarsely integrated, and only for the leading twelve assignment permutations is the full integration performed, with a required precision decreasing according to their relative contributions.

The signal hypothesis is defined as a SM Higgs boson produced in association with a top-quark pair as shown in Fig. 1(a), (b). Hence no coupling of the Higgs boson to the $W$ boson is accounted for in $|M_j|^2$ to allow for a consistent treatment when performing the kinematic transformation. The Higgs boson is required to decay into a pair of $b$-quarks, while the top-quark pair decays into the single-lepton channel. For the background hypothesis, only the diagrams of the irreducible $t\bar{t} + bb$ background are considered. Since it dominates the most signal-rich analysis regions, inclusion of other processes does not improve the separation between signal and background. No gluon radiation from the final-state quarks is allowed, since these are kinematically suppressed and difficult to treat in any kinematic trans-
Fig. 7 Single-lepton channel: NN output for the different regions. In the (5j, 3b) region (a), the $t\bar{t}+HF$ production is considered as signal and $t\bar{t}+\text{light}$ as background whereas in the (5j, ≥4b) (b), (≥6j, 3b) (c), and (≥6j, ≥4b) (d) regions the NN output is for the $t\bar{t}H$ signal and total background. The distributions are normalised to unit area.

Fig. 8 Dilepton channel: NN output for the $t\bar{t}H$ signal and total background in the (a) (3j, 3b), (b) (≥4j, 3b), and (c) (≥4j, ≥4b) regions. The distributions are normalised to unit area.
formation aiming for phase-space alignment during the integration process. In the definition of the signal and background hypothesis the LO diagrams are required to have a top-quark pair as an intermediate state resulting in exactly four $b$-quarks, two light quarks, one charged lepton (electron or muon) and one neutrino in the final state. Assuming lepton universality and invariance under charge conjugation, diagrams of only one lepton flavour and of only negative charge (electron) are considered. The probability density function calculation of the signal and background is only performed in the ($\geq 6j$, 3b) and ($\geq 6j$, 4b) regions of the single-lepton channel. Only six reconstructed jets are considered in the calculation: the four jets with the highest value of the probability to be a $b$-jet returned by the $b$-tagging algorithm (i.e. the highest $b$-tagging weight) and two of the remaining jets with an invariant mass closest to the $W$ boson mass of 80.4 GeV. If a jet is $b$-tagged it cannot be assigned to a light quark in the matrix element description. In the case of more than four $b$-tagged jets, only the four with the highest $b$-tagging weight are treated as $b$-tagged. Assignment permutations between the two light quarks of the hadronically decaying $W$ boson and between the two $b$-quarks originating from the Higgs boson or gluon result in the same likelihood value and are thus not considered. As a result there are in total 12 and 36 assignment permutations in the ($\geq 6j$, 4b) and ($\geq 6j$, 3b) region, respectively, which need to be evaluated in the coarse integration phase.

Using the $ttH$ process as the signal hypothesis and the $tt + bb$ process as the background hypothesis, a slightly modified version of Eq. (3) is used to define the likelihood ratio $D1$:

\[
D1 = \frac{L_{ttH}}{L_{ttH} + \alpha \cdot L_{tt + bb}}, \tag{4}
\]

where $\alpha = 0.23$ is a relative normalisation factor chosen to optimise the performance of the discriminant given the finite bin sizes of the $D1$ distribution. In this definition, signal-like and background-like events have $D1$ values close to one and zero, respectively. The logarithm of the summed signal likelihoods defined by Eq. (2) and the ratio $D1$ are included in the NN training in both the ($\geq 6j$, 3b) and ($\geq 6j$, 4b) regions.

### 8 Systematic uncertainties

Several sources of systematic uncertainty are considered that can affect the normalisation of signal and background and/or the shape of their final discriminant distributions. Individual sources of systematic uncertainty are considered uncorrelated. Correlations of a given systematic effect are maintained across processes and channels. Table 3 presents a summary of the sources of systematic uncertainty considered in the analysis, indicating whether they are taken to be normalisation-only, shape-only, or to affect both shape and normalisation. In Appendix D the normalisation impact of the systematic uncertainties are shown on the $t\bar{t}$ background as well as on the $ttH$ signal.

In order to reduce the degradation of the sensitivity of the search due to systematic uncertainties, they are fitted to data in the statistical analysis, exploiting the constraining power from the background-dominated regions described in Sect. 4. Each systematic uncertainty is represented by an independent parameter, referred to as a “nuisance parameter”, and is fitted with a Gaussian prior for the shape differences and a log-normal distribution for the normalisation. They are centred around zero with a width that corresponds to the given uncertainty.

#### 8.1 Luminosity

The uncertainty on the integrated luminosity for the data set used in this analysis is 2.8%. It is derived following the same methodology as that detailed in Ref. 104. This systematic uncertainty is applied to all contributions determined from the MC simulation.

#### 8.2 Uncertainties on physics objects

##### 8.2.1 Leptons

Uncertainties associated with the lepton selection arise from the trigger, reconstruction, identification, isolation and lepton momentum scale and resolution. In total, uncertainties associated with electrons (muons) include five (six) components.

##### 8.2.2 Jets

Uncertainties associated with the jet selection arise from the jet energy scale (JES), jet vertex fraction requirement, jet energy resolution and jet reconstruction efficiency. Among these, the JES uncertainty has the largest impact on the analysis. The JES and its uncertainty are derived combining information from test-beam data, LHC collision data and simulation [35]. The jet energy scale uncertainty is split into 22 uncorrelated sources which can have different jet $p_T$ and $\eta$ dependencies. In this analysis, the largest jet energy scale uncertainty arises from the $\eta$ dependence of the JES calibration in the end-cap regions of the calorimeter. It is the second leading uncertainty.
8.3 Uncertainties on background modelling

8.3.1 \( tt + \text{jets} \) modelling

An uncertainty of +5%/-6% is assumed for the inclusive \( tt \) production cross section. It includes uncertainties from the PDF and \( \alpha_s \) choices and knowledge of the top quark mass. The PDF and \( \alpha_s \) uncertainties are calculated using the PDF4LHC prescription [105] with the MSTW2008 68% CL NNLO, CT10 NNLO [106] and NNPDF2.3 5f FFN [107] PDF sets, and are added in quadrature to the scale uncertainty. Other systematic uncertainties affecting the modelling of \( tt + \text{jets} \) include uncertainties due to the choice of parton shower and hadronisation model, as well as several uncertainties related to the reweighting procedure applied to improve the \( tt \) MC model. Additional uncertainties are assigned to account for limited knowledge of \( tt + \text{HF} \) jets production. They are described later in this section.

As discussed in Sect. [5] to improve the agreement between data and the \( tt \) simulation a reweighting procedure is applied to \( tt \) MC events based on the difference in the top quark \( p_T \) and \( tt \) system \( p_T \) distributions between data and simulation at \( \sqrt{s} = 7 \) TeV [57]. The nine largest uncertainties associated with the experimental measurement of top quark and \( tt \) system \( p_T \) distributions are considered as separate uncertainty sources in the reweighting applied to the MC prediction. The largest uncertainties on the measurement of the differential distributions include radiation modelling in \( tt \) events, the choice of generator to simulate \( tt \) production, uncertainties on the components of jet energy scale and resolution, and flavour tagging.

Because the measurement is performed for the inclusive \( tt \) sample and the size of the uncertainties applicable to the \( tt + \text{HF} \) system is not known, two addi-
tional uncorrelated uncertainties are assigned to \( tt+cc \) events, consisting of the full difference between applying and not applying the weightings of the \( tt \) system \( p_T \) and top quark \( p_T \), respectively.

An uncertainty due to the choice of parton shower and hadronisation model is derived by comparing events produced by POWHEG interfaced with PYTHIA or HERWIG. Effects on the shapes are compared, symmetrised and applied to the shapes predicted by the default model. Given that the change of the parton shower model leads to two separate effects – a change in the number of jets and a change of the heavy-flavour content – the parton shower uncertainty is represented by three parameters, one acting on the \( tt+\text{light} \) contribution and two others on the \( tt+cc \) and \( tt+bb \) contributions. These three parameters are treated as uncorrelated in the fit.

Detailed comparisons of \( tt+bb \) production between POWHEG+PYTHIA and an NLO prediction of \( tt+bb \) production based on SHERPAOL have shown that the cross sections agree within 50% of each other. Therefore, a systematic uncertainty of 50% is applied to the \( tt + bb \) component of the \( tt+\text{jets} \) background obtained from the POWHEG+PYTHIA MC simulation. In the absence of an NLO prediction for the \( tt+cc \) background, the same 50% systematic uncertainty is applied to the \( tt+cc \) component, and the uncertainties on \( tt + bb \) and \( tt+cc \) are treated as uncorrelated. The large available data sample allows the determination of the \( tt+bb \) and \( tt+cc \) normalisations with much better precision, approximately 15% and 30%, respectively (see Appendix B). Thus, the final result does not significantly depend on the exact value of the assumed prior uncertainty, as long as it is larger than the precision with which the data can constrain it. However, even after the reduction, the uncertainties on the \( tt+bb \) and \( tt+cc \) background normalisation are still the leading and the third leading uncertainty in the analysis, respectively.

Four additional systematic uncertainties in the \( tt+cc \) background estimate are derived from the simultaneous variation of factorisation and renormalisation scales, matching threshold and \( c \)-quark mass variations in the MADGRAPH+PYTHIA \( tt \) simulation, and the difference between the \( tt+cc \) simulation in MADGRAPH+PYTHIA and POWHEG+PYTHIA since MADGRAPH+PYTHIA includes the \( tt+cc \) process in the matrix element calculation while it is absent in POWHEG+PYTHIA.

For the \( tt+bb \) background, three scale uncertainties, including changing the functional form of the renormalisation scale to \( \mu_R = (m_t m_b) \), changing the functional form of the factorisation scale \( \mu_F \) and resummation \( \mu_{Q} \) scales to \( p_T = \mu_F = \mu_{Q} = \sum_{i=t,b} E_{T,i}^{1/2} \\) and varying the renormalisation scale \( \mu_R \) by a factor of two up and down are evaluated. Additionally, the shower recoil model uncertainty and two uncertainties due to the PDF choice in the SHERPAOL NLO calculation are quoted. The effect of these variations on the contribution of different \( tt+bb \) event categories is shown in Fig. The renormalisation scale choice and the shower recoil scheme have a large effect on the modelling of \( tt + bb \). The provide large shape variations of the NN discriminants resulting in the fourth and sixth leading uncertainties in this analysis.

Finally, two uncertainties due to \( tt+bb \) production via multiparton interaction and final-state radiation which are not present in the SHERPAOL NLO calculation are applied. Overall, the uncertainties on \( tt + bb \) normalisation and modelling result in about a 55% total uncertainty on the \( tt + bb \) background contribution in the most sensitive \( ( \geq 6j, \geq 4b) \) and \( ( \geq 4j, \geq 4b) \) regions.

8.3.2 The W/Z+Jets Modelling

As discussed in Sect. the \( W/Z+\text{jets} \) contributions are obtained from the simulation and normalised to the inclusive theoretical cross sections, and a reweighting is applied to improve the modelling of the \( W/Z \) boson \( p_T \) spectrum. The full difference between applying and not applying the \( W/Z \) boson \( p_T \) reweighting is taken as a systematic uncertainty, which is then assumed to be symmetric with respect to the central value. Additional uncertainties are assigned due to the extrapolation of the \( W/Z+\text{jets} \) estimate to high jet multiplicity.

8.3.3 Misidentified Lepton Background Modelling

Systematic uncertainties on the misidentified lepton background estimated via the matrix method in the single-lepton channel receive contributions from the limited number of data events, particularly large jet and \( b \)-tag multiplicities, from the subtraction of the prompt-lepton contribution as well as from the uncertainty on the lepton misidentification rates, estimated in different control regions. The statistical uncertainty is uncorrelated among the different jet and \( b \)-tag multiplicity bins. An uncertainty of 50% associated with the lepton misidentification rate measurements is assumed, which is taken as correlated across jet and \( b \)-tag multiplicity bins, but uncorrelated between electron and muon channels. Uncertainty on the shape of the misidentified lepton background arises from the prompt-lepton background subtraction and the misidentified lepton rate measurement.

In the dilepton channel, since the misidentified lepton background is estimated using both the simulation and same-sign dilepton events in data, a 50% normalisation uncertainty is assigned to cover the maximum
difference between the two methods. It is taken as correlated among the different jet and b-tag multiplicity bins. An additional uncertainty is applied to cover the difference in shape between the predictions derived from the simulation and from same-sign dilepton events in data.

8.3.4 Electroweak background modelling

Uncertainties of $+5\%/-4\%$ and $\pm6.8\%$ are used for the theoretical cross sections of single top production in the single-lepton and dilepton channels [64, 65], respectively. The former corresponds to the weighted average of the theoretical uncertainties on $s$, $t$, and $Wt$-channel production, while the latter corresponds to the theoretical uncertainty on $Wt$-channel production, the only single top process contributing to the dilepton final state.

The uncertainty on the diboson background rates includes an uncertainty on the inclusive diboson NLO cross section of $\pm5\%$ [62] and uncertainties to account for the extrapolation to high jet multiplicity.

Finally, an uncertainty of $\pm30\%$ is assumed for the theoretical cross sections of the $tt + V$ [70, 71] background. An additional uncertainty on $tt + V$ modelling arises from variations in the amount of initial-state radiation. The $tt + Z$ background with $Z$ boson decaying into a $bb$ pair is an irreducible background to the $ttH$, $H \rightarrow bb$ signal, and as such, has kinematics and an NN discriminant shape similar to those of the signal. The uncertainty on the $tt + V$ background normalisation is the fifth leading uncertainty in the analysis.

8.4 Uncertainties on signal modelling

Dedicated NLO PowHel samples are used to evaluate the impact of the choice of factorisation and renormalisation scales on the $ttH$ signal kinematics. In these samples the default scale is varied by a factor of two up and down. The effect of the variations on $ttH$ distributions was studied at particle level and the nominal PowHel $ttH$ sample was reweighted to reproduce these variations. In a similar way, the nominal sample is reweighted to reproduce the effect of changing the functional form of the scale. Additional uncertainties on the $ttH$ signal due to the choice of PDF, parton shower and fragmentation model and NLO generator are also considered. The effect of the PDF uncertainty on the $ttH$ signal is evaluated following the recommendation of the PDF4LHC. The uncertainty in the parton shower and fragmentation is evaluated by comparing PowHel+Pythia8 and PowHel+Herwig samples, while the uncertainty due to a generator choice is evaluated by comparing PowHel+Pythia8 with MadGraph5_AMC@NLO [108] interfaced with Herwig++ [109, 110].
9 Statistical methods

The distributions of the discriminants from each of the channels and regions considered are combined to test for the presence of a signal, assuming a Higgs boson mass of $m_H = 125$ GeV. 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 the ratio of the observed/expected cross section to the SM cross section, and $\theta$, denoting the set of nuisance parameters that encode the effects of systematic uncertainties on the signal and background expectations. They are implemented in the likelihood function as Gaussian or log-normal priors. Therefore, the total number of expected events 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. This procedure allows the impact of systematic uncertainties on the search sensitivity to be reduced by taking advantage of the highly populated background-dominated control regions included in the likelihood fit. It requires a good understanding of the systematic effects affecting the shapes of the discriminant distributions.

The test statistic $q_\mu$ is defined as the profile likelihood ratio: $q_\mu = -2 \ln(L(\mu, \hat{\theta}_\mu)/L(\hat{\mu}, \hat{\theta}))$, where $\hat{\mu}$ and $\hat{\theta}$ are the values of the parameters that maximise the likelihood function (with the constraints $0 \leq \hat{\mu} \leq \mu$), and $\hat{\theta}_\mu$ are the values of the nuisance parameters that maximise the likelihood function for a given value of $\mu$. This test statistic is used to measure the compatibility of the observed data with the background-only hypothesis (i.e. for $\mu = 0$), and to make statistical inferences about $\mu$, such as upper limits using the CL$_s$ method as implemented in the RooFit package.

To obtain the final result, a simultaneous fit to the data is performed on the distributions of the discriminants in 15 regions: nine analysis regions in the single-lepton channel and six regions in the dilepton channel. Fits are performed under the signal-plus-background hypothesis, where the signal-strength parameter $\mu$ is the parameter of interest in the fit and is allowed to float freely, but is required to be the same in all 15 fit regions. The normalisation of each background is determined from the fit simultaneously with $\mu$. Contributions from $t\bar{t}$, $W/Z+\text{jets}$ production, single top, diboson and $t\bar{t}V$ backgrounds are constrained by the uncertainties of the respective theoretical calculations, the uncertainty on the luminosity, and the data themselves. Statistical uncertainties in each bin of the discriminant distributions are taken into account by dedicated parameters in the fit.

10 Results

The results of the binned likelihood fit to data described in Sect.9 are presented in this section. Figure 10 shows the yields after the fit in all analysis regions in the single-lepton and dilepton channels. The post-fit event yields are summarised in Appendix A.

Figures 11–13 and 14–15 show a comparison of data and prediction for the discriminating variables (either $H_T^{\text{had}}$, $H_T$, or NN discriminants) for each of the regions considered in the single-lepton and dilepton channels, respectively, both pre- and post-fit to data. The uncertainties decrease significantly in all regions due to constraints provided by data and correlations between different sources of uncertainty introduced by the fit to the data. In Appendix B the most highly discriminating variables in the NN are shown post-fit compared to data.

Table 4 shows the observed $\mu$ values for the single-lepton and dilepton channels, and their combination. The fitted signal strength for $m_H = 125$ GeV is:

$$\mu(m_H = 125 \text{ GeV}) = 1.5 \pm 1.1.$$  

The expected uncertainty for the signal strength ($\mu = 1$) is $\pm 1.1$. The observed (expected) significance of the signal is 1.4 (1.1) standard deviations, which corresponds to an observed (expected) $p$-value of 8% (15%). The probability, $p$, to obtain a result at least as signal-like as observed if no signal is present is calculated using

$$q_0 = -2 \ln(L(0, \hat{\theta}_0)/L(\hat{\mu}, \hat{\theta}))$$

as a test statistic.

| Signal strength | $\mu$ | Uncertainty |
|-----------------|-------|-------------|
| Single lepton   | 1.2   | 1.3         |
| Dilepton        | 2.8   | 2.0         |
| Combination     | 1.5   | 1.1         |

Table 4 The fitted values of signal strength and their uncertainties for the individual channels as well as their combination, assuming $m_H = 125$ GeV. Total uncertainties are shown.

The fitted values of the signal strength and their uncertainties for the individual channels and their combination are shown in Fig. 10.

The observed limits, those expected with and without assuming a SM Higgs boson with $m_H = 125$ GeV, for each channel and their combination are shown in Fig. 17. A signal 3.4 times larger than predicted by the
SM is excluded at 95% CL using the CLs method. A signal 2.2 times larger than for the SM Higgs boson is expected to be excluded in the case of no SM Higgs boson, and 3.1 times larger in the case of a SM Higgs boson. This is also summarised in Table 5.

The fitted value of the signal strength ($\Delta\mu$) is calculated by fixing the corresponding nuisance parameter at $\hat{\theta} \pm \sigma_{\theta}$, where $\hat{\theta}$ is the fitted value of the nuisance parameter and $\sigma_{\theta}$ is its post-fit uncertainty, and performing the fit again. The difference between the default and the modified $\mu$, $\Delta\mu$, represents the effect on $\mu$ of this particular systematic uncertainty. The largest effect arises from the uncertainty in normalisation of the irreducible $t\bar{t}+bb$ background. This uncertainty is reduced by more than one half from the initial 50%. The $t\bar{t}+bb$ background normalisation is pulled up by about 40% in the fit, resulting in an increase in the observed $t\bar{t}+bb$ yield with respect to the Powheg+Pythia prediction. Most of the reduction in uncertainty on the $t\bar{t}+bb$ normalisation is the result of the significant number of data events in the signal-rich regions dominated by $t\bar{t}+bb$ background. With no Gaussian prior considered on the $t\bar{t}+bb$ normalisation, as described in Sect. 8, the fit still prefers an increase in the amount of $t\bar{t}+bb$ background by about 40%.

The $t\bar{t}+bb$ modelling uncertainties affecting the shape of this background also have a significant effect on $\mu$. These systematic uncertainties affect only the $t\bar{t}+bb$ modelling and are not correlated with the other $t\bar{t}+jets$ backgrounds. The largest of the uncertainties is given by the renormalisation scale choice. The uncertainty drastically changes the shape of the NN for the $t\bar{t}+bb$ background, making it appear more signal-like.

The $t\bar{t}+c\bar{c}$ normalisation uncertainty is ranked third (Fig. 19) and its pull is slightly negative, while the post-fit yields for $t\bar{t}+c\bar{c}$ increase significantly in the four- and five-jet regions in the single-lepton channel and in the two- and three-jet regions of the dilepton channel (see Tables 10 and 11 of Appendix E). It was verified that

---

**Fig. 10** Event yields in all analysis regions in (a) the single-lepton channel and (b) the dilepton channel after the combined fit to data under the signal-plus-background hypothesis. The signal, normalised to the fitted $\mu$, is shown both as a filled area stacked on the other backgrounds and separately as a dashed line. The hashed area represents the total uncertainty on the yields.
Fig. 11 Single-lepton channel: comparison between data and prediction for the discriminant variable used in the (4j, 2b) region (a) before the fit and (b) after the fit, in the (4j, 3b) region (c) before the fit and (d) after the fit, in the (4j, 4b) region (e) before the fit and (f) after the fit. The fit is performed on data under the signal-plus-background hypothesis. The last bin in all figures contains the overflow. The bottom panel displays the ratio of data to the total prediction. The hashed area represents the uncertainty on the background. The $t\bar{t}H$ signal yield (solid) is normalised to the SM cross section before the fit and to the fitted $\mu$ after the fit. In several regions, predominantly the control regions, the $t\bar{t}H$ signal yield is not visible on top of the large background.
Fig. 12  Single-lepton channel: comparison of data and prediction for the discriminant variable used in the (5j, 2b) region (a) before the fit and (b) after the fit, in the (5j, 3b) region (c) before the fit and (d) after the fit, in the (5j, ≥4b) region (e) before the fit and (f) after the fit. The fit is performed on data under the signal-plus-background hypothesis. The last bin in all figures contains the overflow. The bottom panel displays the ratio of data to the total prediction. The hashed area represents the uncertainty on the background. The dashed line shows the signal yield (solid) is normalised to the SM cross section before the fit and to the fitted $\mu_t$ after the fit. In several regions, predominantly the control regions, the $tH$ signal yield is not visible on top of the large background.
The uncertainty on the background. The dashed line shows the t bin in all figures contains the overflow. The bottom panel displays the ratio of data to the total prediction. The hashed area (a) before the fit and (b) after the fit, in the (≥ Single-lepton channel: comparison of data and prediction for the discriminant variable used in the (Fig. 13)

The signal yield is not visible on top of the large background. In several regions, predominantly the control regions, the ttH signal yield is not visible on top of the large background.
The uncertainty on the background. The dashed line shows the fit and (f) after the fit. The fit is performed on data under the signal-plus-background hypothesis. The last bin in all 

Fig. 14 Dilepton channel: comparison of data and prediction for the discriminant variable used in the (2j, 2b) region (a) before the fit and (b) after the fit, in the (3j, 2b) region (c) before the fit and (d) after the fit, in the (3j, 3b) region (e) before the fit and (f) after the fit. The fit is performed on data under the signal-plus-background hypothesis. The last bin in all figures contains the overflow. The bottom panel displays the ratio of data to the total prediction. The hashed area represents the uncertainty on the background. The dashed line shows the signal distribution normalised to background yield. The $t\bar{t}H$ signal yield (solid ) is normalised to the SM cross section before the fit and to the fitted $\mu$ after the fit. In several regions, predominantly the control regions, the $t\bar{t}H$ signal yield is not visible on top of the large background.
Fig. 15 Dilepton channel: comparison of data and prediction for the discriminant variable used in the ($\geq 4j$, $2b$) region (a) before the fit and (b) after the fit, in the ($\geq 4j$, $3b$) region (c) before the fit and (d) after the fit, in the ($\geq 4j$, $4b$) region (e) before the fit and (f) after the fit. The fit is performed on data under the signal-plus-background hypothesis. The last bin in all figures contains the overflow. The bottom panel displays the ratio of data to the total prediction. The hashed area represents the uncertainty on the background. The dashed line shows $t\bar{t}H$ signal distribution normalised to background yield. The $t\bar{t}H$ signal yield (solid) is normalised to the SM cross section before the fit and to the fitted $\mu$ after the fit. In several regions, predominantly the control regions, the $t\bar{t}H$ signal yield is not visible on top of the large background.
Table 5  Observed and expected (median, for the background-only hypothesis) 95% CL upper limits on $\sigma(t\bar{t}H)$ relative to the SM prediction, for the individual channels as well as their combination, assuming $m_H = 125$ GeV. The 68% and 95% confidence intervals around the expected limits under the background-only hypothesis are also provided, denoted by ±1σ and ±2σ, respectively. The expected (median) 95% CL upper limits assuming the SM prediction for $\sigma(t\bar{t}H)$ are shown in the last column.

| Combination         | Observed | −2σ | −1σ | Median | +1σ | +2σ | Median ($\mu = 1$) |
|---------------------|----------|-----|-----|--------|-----|-----|-------------------|
| Single lepton       | 3.6      | 1.4 | 1.9 | 2.6    | 3.7 | 4.9 | 3.6               |
| Dilepton            | 6.7      | 2.2 | 3.0 | 4.1    | 5.8 | 7.7 | 4.7               |
| Combination         | 3.4      | 1.2 | 1.6 | 2.2    | 3.0 | 4.1 | 3.1               |

Fig. 16  The fitted values of the signal strength and their uncertainties for the individual channels and their combination. The green line shows the statistical uncertainty on the signal strength.

Fig. 17  95% CL upper limits on $\sigma(t\bar{t}H)$ relative to the SM prediction, $\sigma/\sigma_{SM}$, for the individual channels as well as their combination. The observed limits (solid lines) are compared to the expected (median) limits under the background-only hypothesis and under the signal-plus-background hypothesis assuming the SM prediction for $\sigma(t\bar{t}H)$ and pre-fit prediction for the background. The surrounding shaded bands correspond to the 68% and 95% confidence intervals around the expected limits under the background-only hypothesis, denoted by ±1σ and ±2σ, respectively.

Fig. 18  Event yields as a function of $\log_{10}(S/B)$, where $S$ (signal yield) and $B$ (background yield) are taken from the $H\rightarrow b\bar{b}$, $H_T$, and NN output bin of each event. Events in all fitted regions are included. The predicted background is obtained from the global signal-plus-background fit. The $t\bar{t}H$ signal is shown both for the best fit value ($\mu = 1.5$) and for the upper limit at 95% CL ($\mu = 3.4$).

this effect is caused by the interplay between the $t\bar{t} + c\bar{c}$ normalisation uncertainty and several other systematic uncertainties affecting the $t\bar{t} + c\bar{c}$ background yield.

The noticeable effect of the light-jet tagging (mistag) systematic uncertainty is explained by the relatively large fraction of the $t\bar{t} + \text{light}$ background in the signal region with four $b$-jets in the single-lepton channel. The $t\bar{t} + \text{light}$ events enter the 4-$b$-tag region through a mistag as opposed to the 3-$b$-tag region where tagging a $c$-jet from a $W$ boson decay is more likely. Since the amount of data in the 4-$b$-tag regions is not large this uncertainty cannot be constrained significantly.
Other leading uncertainties include $b$-tagging and some components of the JES uncertainty. The $t\bar{t} + Z$ background with $Z \rightarrow b\bar{b}$ is an irreducible background to the $t\bar{t}H$ signal as it has the same number of $b$-jets in the final state and similar event kinematics. Its normalisation has a notable effect on $\mu$ ($d\mu/d\sigma(t\bar{t}V) = 0.3$) and the uncertainty arising from the $t\bar{t} + V$ normalisation cannot be significantly constrained by the fit.

11 Summary

A search has been performed for the Standard Model Higgs boson produced in association with a pair of top quarks ($t\bar{t}H$) using 20.3 fb$^{-1}$ of $pp$ collision data at $\sqrt{s} = 8$ TeV collected with the ATLAS detector during the first run of the Large Hadron Collider. The search focuses on $H \rightarrow b\bar{b}$ decays, and is performed in events with either one or two charged leptons.

To improve sensitivity, the search employs a likelihood fit to data in several jet and $b$-tagged jet multiplicity regions. Systematic uncertainties included in the fit are significantly constrained by the data. Discrimination between signal and background is obtained in both final states by employing neural networks in the signal-rich regions. In the single-lepton channel, discriminating variables are calculated using the matrix element technique. They are used in addition to kinematic vari-

---

**Fig. 19** The fitted values of the nuisance parameters with the largest impact on the measured signal strength. The points, which are drawn conforming to the scale of the bottom axis, show the deviation of each of the fitted nuisance parameters, $\theta$, from $\theta_0$, which is the nominal value of that nuisance parameter, in units of the pre-fit standard deviation $\Delta\theta$. The error bars show the post-fit uncertainties, $\sigma_\theta$, which are close to 1 if the data do not provide any further constraint on that uncertainty. Conversely, a value of $\sigma_\theta$ much smaller than 1 indicates a significant reduction with respect to the original uncertainty. The nuisance parameters are sorted according to the post-fit effect of each on $\mu$ (hashed blue area) conforming to the scale of the top axis, with those with the largest impact at the top.
ables as input to the neural network. No significant excess of events above the background expectation is found for a Standard Model Higgs boson with a mass of 125 GeV. An observed (expected) 95% confidence-level upper limit of 3.4 (2.2) times the Standard Model cross section is obtained. By performing a fit under the signal-plus-background hypothesis, the ratio of the measured signal strength to the Standard Model expectation is found to be $\mu = 1.5 \pm 1.1$.

Acknowledgements

We honour the memory of our colleague Richard St. Denis, who was a driving force of the work described here for a long time and died shortly before its completion.

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; BMWF and FWF, Austria; ANAS, Azerbaijan; SSTC, Belarus; CNPq and FAPESP, Brazil; NSERC, NRC and CFI, Canada; CERN; CONICYT, Chile; CAS, MOST and NSFC, China; COLCIENCIAS, Colombia; CRC, DAFM and PPARC, Denmark; DST/NRF, South Africa; MINECO, Spain; SRC and Wallenberg Foundation, Sweden; BRF and RCN, Norway; MNiSW and NCN, Poland; FWO, Belgium; CNRST, Morocco; FCT, Portugal; MNE/IFA, Romania; MES and AMO, Russia; MNiSW, Serbia; ARRS and MIZˇS, Slovenia; DST/NRF, South Africa; MINECO, Spain; SRC and Wallenberg Foundation, Sweden; SER, SNSF and Cantons of Bern and Geneva, Switzerland; NSC, Taiwan; TAEK, Turkey; STFC, the Royal Society and Leverhulme Trust, United Kingdom; DOE and NSF, United States of America.

The crucial computing support from all WLCG partners is acknowledged gratefully, in particular from CERN and the ATLAS Tier-1 facilities at TRIUMF (Canada), NDGF (Denmark, Norway, Sweden), CC-IN2P3 (France), KIT/GridKA (Germany), INP–CNAF (Italy), NL–T1 (Netherlands), PIC (Spain), ASGC (Taiwan), RAL (UK) and BNL (USA) and in the Tier-2 facilities worldwide.

References

1. S. L. Glashow, Partial symmetries of weak interactions, Nucl. Phys. 22 (1961) 579.
2. S. Weinberg, A model of leptons, Phys. Rev. Lett. 19 (1967) 1264–1266.
3. A. Salam, Weak and electromagnetic interactions, Proc. of the 8th Nobel Symposium (1969) 367.
4. F. Englert and R. Brout, Broken Symmetry and the Mass of Gauge Vector Mesons, Phys. Rev. Lett. 13 (1964) 321.
5. P. W. Higgs, Broken Symmetries and the Masses of Gauge Bosons, Phys. Rev. Lett. 13 (1964) 508.
6. P. W. Higgs, Broken Symmetries, Massless Particles and Gauge Fields, Phys. Lett. 12 (1964) 132.
7. G. Guralnik, C. Hagen, and T. Kibble, Global Conservation Laws and Mass-less Particles, Phys. Rev. Lett. 13 (1964) 585.
8. ATLAS Collaboration, Observation of a new particle in the search for the Standard Model Higgs boson with the ATLAS detector at the LHC, Phys. Lett. B716 (2012) 1 [arXiv:1207.7214 [hep-ex]].
9. CMS Collaboration, Observation of a new boson at a mass of 125 GeV with the CMS experiment at the LHC, Phys. Lett. B716 (2012) 30 [arXiv:1207.7235 [hep-ex]].
10. ATLAS Collaboration, Measurements of Higgs boson production and couplings in diboson final states with the ATLAS detector at the LHC, Phys. Lett. B726 (2013) 88 [arXiv:1307.1427 [hep-ex]].
11. ATLAS Collaboration, Evidence for the spin-0 nature of the Higgs boson using ATLAS data, Phys. Lett. B726 (2013) 120 [arXiv:1307.1432 [hep-ex]].
12. ATLAS Collaboration, Evidence for the Higgs-boson Yukawa coupling to tau leptons with the ATLAS Detector, arXiv:1501.04943 [hep-ex].
13. CMS Collaboration, Precise determination of the mass of the Higgs boson and tests of compatibility of its couplings with the standard model predictions using proton collisions at 7 and 8 TeV, arXiv:1412.8662 [hep-ex].
14. CMS Collaboration, Evidence for the direct decay of the 125 GeV Higgs boson to fermions, Nature Physics 10 (2014) 557 [arXiv:1401.6527 [hep-ex]].
15. CMS Collaboration, Constraints on the spin-parity and anomalous HVV couplings of the Higgs boson in proton collisions at 7 and 8 TeV, arXiv:1411.3441 [hep-ex].
16. J. C. Ng and P. Zakaranaskas, QCD-parton calculation of conjoined production of Higgs bosons and heavy flavors in p anti-p collisions, Phys. Rev. D 29 (1984) 876.
17. Z. Kunszt, Associated production of heavy Higgs boson with top quarks, Nucl. Phys. B 29 (1984) 876.
18. S. Dawson and L. H. Orr and L. Reina and D. Wackeroth, Associated top quark Higgs boson production the LHC, Phys. Rev. D 67 (2003) 071503 [arXiv:hep-ph/021438].
19. W. Beenakker et al., Higgs radiation off top quarks at the Tevatron and the LHC, Phys.Rev.Lett. 87 (2001) 201805 [arXiv:0107081 [hep-ph]].
20. F. Bezrukov and M. Shaposhnikov, Why should we care about the top quark Yukawa coupling?, arXiv:1411.1923 [hep-ph].
21. ATLAS Collaboration, Search for the bb decay of the Standard Model Higgs boson in associated (W/Z)H production with the ATLAS detector, JHEP 01 (2015) 009 [arXiv:1409.6212 [hep-ex]].
22. CMS Collaboration, Search for the standard model Higgs boson produced in association with a W or a Z boson and decaying to bottom quarks, Phys. Rev. D 89 (2014) 012003, arXiv:1310.3687 [hep-ex].

23. CDF and D0 Collaboration, T. Aaltonen et al., Evidence for a Particle Produced in Association with Weak Bosons and Decaying to a Bottom-Antibottom Quark Pair in Higgs Boson Searches at the Tevatron, Phys. Rev. Lett. 109 (2012) 071804, arXiv:1207.6436 [hep-ex].

24. CMS Collaboration, Search for the associated production of the Higgs boson with a top-quark pair, JHEP 09 (2014) 087, arXiv:1408.1682 [hep-ex].

25. ATLAS Collaboration, The ATLAS Experiment at the CERN Large Hadron Collider, JINST 3 (2008) S08003.

26. ATLAS Collaboration, Electron reconstruction and identification efficiency measurements with the ATLAS detector using the 2011 LHC proton–proton collision data, Eur. Phys. J. C 74 (2014) 2941, arXiv:1404.2240 [hep-ex].

27. ATLAS Collaboration, Electron efficiency measurements with the ATLAS detector using the 2012 LHC proton–proton collision data, ATLAS-CONF-2014-032 (2014).

28. ATLAS Collaboration, Measurement of the muon reconstruction performance of the ATLAS detector using 2011 and 2012 LHC proton–proton collision data, Eur. Phys. J. C 74 (2014) 3130, arXiv:1407.3935 [hep-ex].

29. ATLAS Collaboration, Jet energy measurement with the ATLAS detector in proton–proton collisions at $\sqrt{s} = 7$ TeV, Eur. Phys. J. C 73 (2013) 2304, arXiv:1112.6426 [hep-ex].

30. M. Cacciari, G. P. Salam, and G. Soyez, The anti-k_t jet clustering algorithm, JHEP 04 (2008) 063, arXiv:0802.1189.

31. M. Cacciari and G. P. Salam, Dispelling the N^3 myth for the k_t jet-finder, Phys. Lett. B641 (2006) 57, arXiv:0512210 [hep-ph].

32. M. Cacciari, G. P. Salam, and G. Soyez, FastJet User Manual, Eur. Phys. J. CT2 (2012) 1896, arXiv:1111.6997 [hep-ph].

33. C. Cojocaru et al., Hadronic calibration of the ATLAS liquid argon end-cap calorimeter in the pseudorapidity region 1.6 < $|\eta|$ < 1.8 in beam tests, Nucl. Instrum. Meth. A 531 (2004) 481, arXiv:physics/0407009.

34. T. Barillari et al., Local hadronic calibration, ATLAS-LARG-PUB-2009-001 (2009).

35. ATLAS Collaboration, Jet energy measurement and its systematic uncertainty in proton-proton collisions at $\sqrt{s} = 7$ TeV with the ATLAS detector, Eur. Phys. J. C 75 (2015) 17, arXiv:1406.0765 [hep-ph].

36. ATLAS Collaboration, Calibration of the performance of b-tagging for c and light-flavour jets in the 2012 ATLAS data, ATLAS-CONF-2014-046 (2014).

37. ATLAS Collaboration, Calibration of b-tagging using dileptonic top pair events in a combinatorial likelihood approach with the ATLAS experiment, ATLAS-CONF-2014-004 (2014).

38. ATLAS Collaboration, Measurement of the top quark pair production cross section with ATLAS in pp collisions at $\sqrt{s} = 7$ TeV, Eur. Phys. J. C 71 (2011) 1577, arXiv:1012.1792 [hep-ex].

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

40. S. Frixione, P. Nason, and C. Oleari, Matching NLO QCD computations with Parton Shower simulations: the POWHEG method, arXiv:0709.2092 [hep-ph].

41. S. Alioli, P. Nason, C. Oleari, and E. Re, A general framework for implementing NLO calculations in shower Monte Carlo programs: the POWHEG BOX, JHEP 06 (2010) 040, arXiv:1002.2581 [hep-ph].

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

43. T. Sjostrand, S. Mrenna, and P. Skands, Python 6.4 Physics and Manual, JHEP 05 (2006) 026, arXiv:hep-ph/0603175.

44. P. M. Nadolsky et al., Implications of CTEQ global analysis for collider observables, Phys. Rev. D 78 (2008) 013004, arXiv:0802.0007 [hep-ph].

45. P. Z. Skands, Tuning Monte Carlo Generators: The Perugia Tunes, Phys.Rev. D80 (2009) 074018, arXiv:0905.3467 [hep-ph].

46. M. Czakon and A. Mitov, Top++: a program for the calculation of the top-pair cross-section at hadron colliders, arXiv:1112.5675 [hep-ph].

47. M. Cacciari, M. Czakon, M. Mangano, A. Mitov, and P. Nason, Top-pair production at hadron colliders with next-to-next-to-leading logarithmic soft-gluon resummation, Phys. Lett. B710 (2012) 612, arXiv:1111.5869 [hep-ph].

48. P. Barrethurer, M. Czakon, and A. Mitov, Percent Level Precision Physics at the Tevatron: First Genuine NNLO QCD Corrections to qq → tt, Phys. Rev. Lett. 109 (2012) 132001, arXiv:1204.5201 [hep-ph].

49. M. Czakon and A. Mitov, NNLO corrections to top-pair production at hadron colliders: the all-fermionic scattering channels, arXiv:1207.0238 [hep-ph].

50. M. Czakon and A. Mitov, NNLO corrections to top-pair production at hadron colliders: the quark-gluon reaction, JHEP 1301 (2013) 080, arXiv:1210.6832 [hep-ph].

51. M. Czakon, P. Fiedler, and A. Mitov, The total top quark pair production cross-section at hadron colliders through $O(a_s^6)$, Phys. Rev. Lett. 110 (2013) 252004, arXiv:1303.6254 [hep-ph].

52. J. Alwall, M. Herquet, F. Maltoni, O. Mattelaer, and T. Stelzer, MadGraph 5: Going Beyond, JHEP 1106 (2011) 128, arXiv:1106.0522 [hep-ph].

53. M. L. Mangano, M. Moretti, and R. Pittau, MultiJet matrix elements and shower evolution in hadronic collisions: W+b + jets as a case study, Nucl. Phys. B 632 (2002) 343, arXiv:hep-ph/0108069.

54. F. Cascioli, P. Maierhoefer, N. Moretti, S. Pozzorini, and F. Siegert, NLO matching for ttbb production with MadGraph 5, Eur. Phys. J. C 74 (2014) 210, arXiv:1309.5912 [hep-ph].

55. T. Gleisberg et al., Event generation with SHERPA 1.1, JHEP 0902 (2009) 007, arXiv:0811.4622 [hep-ph].

56. F. Cascioli, P. Maierhoefer, and S. Pozzorini, Scattering Amplitudes with Open Loops, Phys. Rev. Lett. 108 (2012) 072004, arXiv:1111.5906 [hep-ph].

57. ATLAS Collaboration, Measurements of normalized differential cross sections for tt production in pp collisions at $\sqrt{s} = 7$ TeV using the ATLAS detector, Phys.Rev. D90 (2014) 072004, arXiv:1407.0371 [hep-ex].
B Event yields prior to the fit

The event yields prior to the fit for the combined $e+j$-jets and $\mu+j$-jets samples for the different regions considered in the analysis are summarised in Table B.8.

The event yields prior to the fit for the combined $ee+j$-jets, $\mu\mu+j$-jets and $e\mu+j$-jets samples for the different regions considered in the dilepton channel are summarised in Table B.9.

C Discrimination power of input variables

Figures 21–24 and 25–27 show the discrimination between signal and background for the top four input variables in each region where NN is used in the single-lepton and dilepton channels, respectively. In Fig. 21, the NN is designed to separate $tt+\text{HF}$ from $tt+\text{light}$.

D Tables of systematic uncertainties in the signal region

Tables 8 and 9 show pre-fit and post-fit contributions of the different categories of uncertainties (expressed in %) for the $t\bar{t}H$ signal and main background processes in the $6\geq j, 6\geq b$ region of the single-lepton channel and the $4\geq j, 4\geq b$ region of the dilepton channel, respectively. The “Lepton efficiency” category includes systematic uncertainties on electrons and muons listed in Table 3. The “Jet efficiency” category includes uncertainties on the jet shape and on the theoretical predictions.

The “Lepton efficiency” category includes systematic uncertainties on electrons and muons listed in Table 3. The “Jet efficiency” category includes uncertainties on the jet vertex fraction and jet reconstruction. The “$t\bar{t}$ heavy-flavour modelling” category includes uncertainties on the $t\bar{t}+\text{bb}$ NLO shape and on the $t\bar{t}+c\bar{c}$ $p_T$ reweighting and generator. The “Theoretical cross sections” category includes uncertainties on the single-top, diboson, $V+jets$ and $t+V$ theoretical cross sections. The “$t\bar{t}H$ modelling” category includes contributions from $t\bar{t}H$ scale, generator, hadronisation model and PDF choice. The details of the evaluation of the uncertainties can be found in Sect. 8.

E Post-fit event yields

The post-fit event yields for the combined single-lepton channel for the different regions considered in the analysis are summarised in Table C.1. Similarly, the post-fit event yields for the combined dilepton channels for the different regions are summarised in Table C.2.

F Post-fit input variables

Figures 28–31 and 32–34 show a comparison of data and prediction for the top four input variables in each region with a neural network in the single-lepton channel and dilepton channel, respectively. All of the plots are made using post-fit predictions.
Fig. 20 Contribution of various Higgs boson decay modes to the analysis regions in (a) the single-lepton channel and (b) the dilepton channel.
Table 6  Single lepton channel: pre-fit event yields for signal, backgrounds and data in each of the analysis regions. The quoted uncertainties are the sum in quadrature of the statistical and systematic uncertainties on the yields.
Uncertainties are the sum in quadrature of the statistical and systematic uncertainties on the yields.

**Table 7** Dilepton channel: pre-fit event yields for signal, backgrounds and data in each of the analysis regions. The quoted uncertainties are the sum in quadrature of the statistical and systematic uncertainties on the yields.

|     | 2 j, 2 b | 3 j, 2 b | 3 j, 3 b |
|-----|----------|----------|----------|
| ttH (125) | 1.5 ± 0.2 | 5.3 ± 0.5 | 2.2 ± 0.3 |
| tt + light | 14000 ± 1800 | 8100 ± 880 | 96 ± 21 |
| tt + c+c | 270 ± 170 | 600 ± 320 | 76 ± 44 |
| tt + bb | 150 ± 87 | 260 ± 130 | 120 ± 65 |
| Z+jets | 330 ± 30 | 190 ± 49 | 8.2 ± 3.1 |
| Single top | 430 ± 71 | 270 ± 30 | 7.6 ± 3.5 |
| Diboson | 6.8 ± 2.2 | 4.2 ± 1.5 | ≤ 0.1 ± 0.1 |
| tt + V | 8.4 ± 2.7 | 21 ± 6 | 1.9 ± 0.6 |
| Lepton misID | 21 ± 10 | 33 ± 17 | 0.8 ± 0.4 |
| Total | 15000 ± 1900 | 9500 ± 1000 | 310 ± 85 |

| Data | 15296 | 9996 | 374 |

|     | ≥ 4 j, 2 b | ≥ 4 j, 3 b | ≥ 4 j, ≥ 4 b |
|-----|------------|------------|-------------|
| ttH (125) | 15 ± 1 | 8.6 ± 0.6 | 2.7 ± 0.3 |
| tt + light | 4400 ± 810 | 120 ± 31 | 1.9 ± 0.8 |
| tt + c+c | 710 ± 380 | 130 ± 74 | 5.0 ± 3.0 |
| tt + bb | 290 ± 150 | 200 ± 100 | 31 ± 17 |
| Z+jets | 100 ± 39 | 10 ± 4 | 0.6 ± 0.2 |
| Single top | 140 ± 55 | 11 ± 5 | 0.8 ± 0.2 |
| Diboson | 4.0 ± 1.3 | 0.4 ± 0.1 | ≤ 0.1 ± 0.1 |
| tt + V | 45 ± 14 | 7.8 ± 2.4 | 1.1 ± 0.4 |
| Lepton misID | 38 ± 19 | 4.3 ± 2.2 | 0.4 ± 0.2 |
| Total | 5800 ± 1000 | 490 ± 140 | 43 ± 18 |

| Data | 6006 | 561 | 46 |

|     |     |     |     |
|-----|-----|-----|-----|
| S/B | 0.003 | 0.015 | 0.059 |
| S/√B | 0.197 | 0.365 | 0.401 |

**Fig. 21** Single-lepton channel: comparison of tt+HF (dashed) and tt+light (solid) background for the four top-ranked input variables in the (5j, 3b) region where the NN is designed to separate these two backgrounds. The plots include (a) m_{bb}^{max} p_{T}, (b) m_{uu}^{min} ΔR, (c) p_{T,uu}^{min} ΔR and (d) m_{bb}^{min} ΔR.
**Fig. 22** Single-lepton channel: comparison of $t\bar{t}H$ signal (dashed) and background (solid) for the four top-ranked input variables in the $(5j, \geq 4b)$ region. The plots include (a) Centrality, (b) $H1$, (c) $N^\text{int}_{40}$ and (d) $m_{b\bar{b}}$. 

**Fig. 23** Single-lepton channel: comparison of $t\bar{t}H$ signal (dashed) and background (solid) for the four top-ranked input variables in the $(\geq 6j, 3b)$ region. The plots include (a) $N^\text{int}_{40}$, (b) Centrality, (c) $H1$, and (d) SSLL. 

**Fig. 24** Single-lepton channel: comparison of $t\bar{t}H$ signal (dashed) and background (solid) for the four top-ranked input variables in the $(\geq 6j, \geq 4b)$ region. The plots include (a) $D1$, (b) Centrality, (c) $p_T^{\text{int}}$, and (d) $H1$. 
Fig. 25  Dilepton channel: comparison of $t\bar{t}H$ signal (dashed) and background (solid) for the four top-ranked input variables in the $(3j, 3b)$ region. The plots include (a) $\Delta\eta^{\text{max}}\Delta\eta$, (b) $m_{jj}^{\text{min}}$, (c) $H_T$, and (d) $H_4$.

Fig. 26  Dilepton channel: comparison of $ttH$ signal (dashed) and background (solid) for the four top-ranked input variables in the $(\geq 4j, 3b)$ region. The plots include (a) $\Delta\eta^{\text{max}}\Delta\eta$, (b) $N_{30}$, (c) $m_{jj}^{\text{min}}$, and (d) $\Delta R_{bb}^{\text{max}}$.

Fig. 27  Dilepton channel: comparison of $t\bar{t}H$ signal (dashed) and background (solid) for the four top-ranked input variables in the $(\geq 4j, \geq 4b)$ region. The plots include (a) $\Delta\eta^{\text{max}}\Delta\eta$, (b) $m_{bb}^{\text{min}}\Delta R$, (c) $m_{bb}$, and (d) $\Delta R_{bb}$.
\[
\begin{array}{cccc|cccc}
\geq 6 \text{ } j, \geq 4 \text{ } b & \text{Pre-fit} & & \text{Post-fit} & \\
& ttH (125) & tt + light & tt + c\bar{c} & tt + b\bar{b} & ttH (125) & tt + light & tt + c\bar{c} & tt + b\bar{b} \\
\hline
\text{Luminosity} & \pm 2.8 & \pm 2.8 & \pm 2.8 & \pm 2.8 & \pm 2.6 & \pm 2.6 & \pm 2.6 & \pm 2.6 \\
\text{Lepton efficiencies} & \pm 1.4 & \pm 1.4 & \pm 1.4 & \pm 1.5 & \pm 1.3 & \pm 1.3 & \pm 1.3 & \pm 1.3 \\
\text{Jet energy scale} & \pm 6.4 & \pm 13 & \pm 11 & \pm 9.2 & \pm 2.3 & \pm 5.3 & \pm 4.7 & \pm 3.6 \\
\text{Jet efficiencies} & \pm 1.7 & \pm 5.2 & \pm 2.7 & \pm 2.5 & \pm 0.7 & \pm 2.3 & \pm 1.2 & \pm 1.1 \\
\text{Jet energy resolution} & \pm 0.1 & \pm 4.4 & \pm 2.5 & \pm 1.6 & \pm 0.1 & \pm 2.3 & \pm 1.3 & \pm 0.8 \\
\text{b-tagging efficiency} & \pm 9.2 & \pm 5.6 & \pm 5.1 & \pm 9.3 & \pm 5.0 & \pm 3.1 & \pm 2.9 & \pm 5.0 \\
\text{c-tagging efficiency} & \pm 1.7 & \pm 6.0 & \pm 12 & \pm 2.4 & \pm 1.4 & \pm 5.1 & \pm 10 & \pm 2.1 \\
\text{t-tagging efficiency} & \pm 1.0 & \pm 19 & \pm 5.2 & \pm 2.1 & \pm 0.6 & \pm 11 & \pm 3.0 & \pm 1.1 \\
\text{High } p_T \text{ tagging efficiency} & \pm 0.6 & - & \pm 0.7 & \pm 0.6 & \pm 0.3 & - & \pm 0.4 & \pm 0.3 \\
\text{tt: } p_T \text{ reweighting} & - & \pm 5.4 & \pm 6.1 & - & - & \pm 4.7 & \pm 5.4 & - \\
\text{tt: parton shower} & - & \pm 13 & \pm 16 & \pm 11 & - & \pm 3.6 & \pm 10 & \pm 6.0 \\
\text{tt+HF: normalisation} & - & - & \pm 50 & \pm 50 & - & - & \pm 28 & \pm 14 \\
\text{tt+HF: modelling} & - & \pm 11 & \pm 16 & \pm 8.3 & - & \pm 3.6 & \pm 9.1 & \pm 7.1 \\
\text{Theoretical cross sections} & - & \pm 6.3 & \pm 6.3 & \pm 6.3 & - & \pm 4.1 & \pm 4.1 & \pm 4.1 \\
\text{ttH modelling} & \pm 2.7 & - & - & - & \pm 2.6 & - & - & - \\
\hline
\text{Total} & \pm 12 & \pm 32 & \pm 59 & \pm 54 & \pm 6.9 & \pm 9.2 & \pm 23 & \pm 12 \\
\end{array}
\]

Table 8 Single lepton channel: normalisation uncertainties (expressed in % ) on signal and main background processes for the systematic uncertainties considered, before and after the fit to data in \((\geq 6j, \geq 4b)\) region of the single lepton channel. The total uncertainty can be different from the sum in quadrature of individual sources due to the anti-correlations between them.

\[
\begin{array}{cccc|cccc}
\geq 4 \text{ } j, \geq 4 \text{ } b & \text{Pre-fit} & & \text{Post-fit} & \\
& ttH (125) & tt + light & tt + c\bar{c} & tt + b\bar{b} & ttH (125) & tt + light & tt + c\bar{c} & tt + b\bar{b} \\
\hline
\text{Luminosity} & \pm 2.8 & \pm 2.8 & \pm 2.8 & \pm 2.8 & \pm 2.6 & \pm 2.6 & \pm 2.6 & \pm 2.6 \\
\text{Lepton efficiencies} & \pm 2.5 & \pm 2.5 & \pm 2.5 & \pm 2.5 & \pm 1.8 & \pm 1.8 & \pm 1.8 & \pm 1.8 \\
\text{Jet energy scale} & \pm 4.5 & \pm 12 & \pm 9.4 & \pm 7.0 & \pm 2.0 & \pm 5.5 & \pm 4.5 & \pm 3.3 \\
\text{Jet efficiencies} & - & \pm 5.9 & \pm 1.6 & \pm 0.9 & - & \pm 2.6 & \pm 0.7 & \pm 0.4 \\
\text{Jet energy resolution} & \pm 0.1 & \pm 4.5 & \pm 1.1 & - & \pm 0.1 & \pm 2.3 & \pm 0.6 & - \\
\text{b-tagging efficiency} & \pm 10 & \pm 5.5 & \pm 5.4 & \pm 11 & \pm 5.6 & \pm 3.1 & \pm 3.0 & \pm 5.8 \\
\text{c-tagging efficiency} & \pm 0.5 & - & \pm 12 & \pm 0.6 & \pm 0.3 & - & \pm 10 & \pm 0.3 \\
\text{t-tagging efficiency} & \pm 0.7 & \pm 34 & \pm 7.0 & \pm 1.6 & \pm 0.4 & \pm 21 & \pm 4.2 & \pm 0.9 \\
\text{High } p_T \text{ tagging efficiency} & - & - & \pm 0.6 & - & - & - & \pm 0.3 & - \\
\text{tt: } p_T \text{ reweighting} & - & \pm 5.8 & \pm 6.2 & - & - & \pm 5.0 & \pm 5.4 & - \\
\text{tt: parton shower} & - & \pm 14 & \pm 18 & \pm 14 & - & \pm 4.8 & \pm 11 & \pm 8.1 \\
\text{tt+HF: normalisation} & - & - & \pm 50 & \pm 50 & - & - & \pm 28 & \pm 14 \\
\text{tt+HF: modelling} & - & \pm 11 & \pm 16 & \pm 12 & - & \pm 3.8 & \pm 10 & \pm 10 \\
\text{Theoretical cross sections} & - & \pm 6.3 & \pm 6.3 & \pm 6.2 & - & \pm 4.1 & \pm 4.1 & \pm 4.1 \\
\text{ttH modelling} & \pm 1.9 & - & - & - & \pm 1.8 & - & - & - \\
\hline
\text{Total} & \pm 12 & \pm 40 & \pm 59 & \pm 55 & \pm 6.7 & \pm 22 & \pm 22 & \pm 13 \\
\end{array}
\]

Table 9 Dilepton channel: normalisation uncertainties (expressed in % ) on signal and main background processes for the systematic uncertainties considered, before and after the fit to data in \((\geq 4j, \geq 4b)\) region of the dilepton channel. The total uncertainty can be different from the sum in quadrature of individual sources due to the anti-correlations between them.
|                | 4 j, 2 b | 4 j, 3 b | 4 j, 4 b |
|----------------|----------|----------|----------|
| $t\bar{t}H$ (125) | $48 \pm 35$ | $20 \pm 15$ | $3.0 \pm 2.2$ |
| $t\bar{t}$ + light | $78000 \pm 1600$ | $6300 \pm 160$ | $56 \pm 5$ |
| $t\bar{t} + c\bar{c}$ | $6400 \pm 1800$ | $850 \pm 220$ | $26 \pm 7$ |
| $t\bar{t} + bb$ | $2500 \pm 490$ | $970 \pm 150$ | $63 \pm 8$ |
| $W$+jets | $3700 \pm 1100$ | $170 \pm 51$ | $4.0 \pm 1.2$ |
| $Z$+jets | $1100 \pm 540$ | $49 \pm 25$ | $1.1 \pm 0.6$ |
| Single top | $4700 \pm 320$ | $330 \pm 28$ | $6.8 \pm 0.7$ |
| Diboson | $220 \pm 65$ | $11 \pm 4$ | $0.3 \pm 0.1$ |
| $t\bar{t} + V$ | $120 \pm 38$ | $16 \pm 5$ | $0.9 \pm 0.3$ |
| Lepton misID | $1100 \pm 370$ | $78 \pm 26$ | $2.6 \pm 1.0$ |
| Total | $98000 \pm 340$ | $8800 \pm 82$ | $160 \pm 6$ |
| Data | $98049$ | $8752$ | $161$ |

|                | 5 j, 2 b | 5 j, 3 b | 5 j, $\geq$ 4 b |
|----------------|----------|----------|----------------|
| $t\bar{t}H$ (125) | $60 \pm 44$ | $34 \pm 25$ | $9.4 \pm 6.9$ |
| $t\bar{t}$ + light | $38000 \pm 1000$ | $3600 \pm 120$ | $65 \pm 6$ |
| $t\bar{t} + c\bar{c}$ | $4800 \pm 1200$ | $930 \pm 230$ | $51 \pm 12$ |
| $t\bar{t} + bb$ | $2400 \pm 360$ | $1300 \pm 180$ | $150 \pm 20$ |
| $W$+jets | $1200 \pm 420$ | $87 \pm 31$ | $4.0 \pm 1.5$ |
| $Z$+jets | $370 \pm 200$ | $28 \pm 16$ | $1.4 \pm 0.8$ |
| Single top | $1700 \pm 150$ | $190 \pm 18$ | $8.2 \pm 0.7$ |
| Diboson | $94 \pm 35$ | $8.0 \pm 3.1$ | $0.5 \pm 0.2$ |
| $t\bar{t} + V$ | $140 \pm 43$ | $26 \pm 8$ | $3.2 \pm 1.0$ |
| Lepton misID | $340 \pm 110$ | $44 \pm 16$ | $5.7 \pm 2.2$ |
| Total | $50000 \pm 220$ | $6200 \pm 54$ | $300 \pm 10$ |
| Data | $49699$ | $6199$ | $286$ |

|                | $\geq$ 6 j, 2 b | $\geq$ 6 j, 3 b | $\geq$ 6 j, $\geq$ 4 b |
|----------------|-----------------|-----------------|-----------------|
| $t\bar{t}H$ (125) | $89 \pm 65$ | $57 \pm 42$ | $24 \pm 17$ |
| $t\bar{t}$ + light | $19000 \pm 700$ | $2100 \pm 87$ | $58 \pm 5$ |
| $t\bar{t} + c\bar{c}$ | $3700 \pm 890$ | $890 \pm 210$ | $85 \pm 21$ |
| $t\bar{t} + bb$ | $2000 \pm 310$ | $1400 \pm 190$ | $330 \pm 37$ |
| $W$+jets | $450 \pm 170$ | $51 \pm 19$ | $4.4 \pm 1.9$ |
| $Z$+jets | $150 \pm 86$ | $16 \pm 9$ | $1.2 \pm 0.7$ |
| Single top | $730 \pm 83$ | $110 \pm 14$ | $11 \pm 2$ |
| Diboson | $45 \pm 20$ | $5.6 \pm 2.6$ | $0.5 \pm 0.2$ |
| $t\bar{t} + V$ | $170 \pm 52$ | $42 \pm 13$ | $8.2 \pm 2.5$ |
| Lepton misID | $120 \pm 41$ | $14 \pm 5$ | $1.1 \pm 0.5$ |
| Total | $26000 \pm 160$ | $4600 \pm 55$ | $520 \pm 18$ |
| Data | $26185$ | $4701$ | $516$ |

**Table 10** Single lepton channel: post-fit event yields under the signal-plus-background hypothesis for signal, backgrounds and data in each of the analysis regions. The quoted uncertainties are the sum in quadrature of statistical and systematic uncertainties on the yields, computed taking into account correlations among nuisance parameters and among processes.
Table 11  Dilepton channel: post-fit event yields under the signal-plus-background hypothesis for signal, backgrounds and data in each of the analysis regions. The quoted uncertainties are the sum in quadrature of statistical and systematic uncertainties on the yields, computed taking into account correlations among nuisance parameters and among processes.

|       | 2 j, 2 b | 3 j, 2 b | 3 j, 3 b |
|-------|----------|----------|----------|
| $ttH$ (125) | 2.4 ± 1.8 | 8.1 ± 5.9 | 3.0 ± 2.2 |
| $tt$ + light | 14000 ± 160 | 8300 ± 170 | 84 ± 9.6 |
| $tt$ + $cc$ | 400 ± 110 | 700 ± 160 | 92 ± 22 |
| $tt$ + $bb$ | 190 ± 36 | 350 ± 49 | 140 ± 19 |
| $Z$ + jets | 330 ± 22 | 200 ± 43 | 7.3 ± 2.4 |
| Single top | 430 ± 35 | 260 ± 21 | 7.6 ± 1.5 |
| Diboson | 6.8 ± 2.1 | 4.5 ± 1.4 | ≤ 0.1 ± 0.1 |
| $tt$ + $V$ | 8.7 ± 2.7 | 21 ± 6 | 1.8 ± 0.6 |
| Lepton misID | 19 ± 10 | 30 ± 15 | 0.7 ± 0.4 |
| Total | 15000 ± 120 | 9900 ± 82 | 340 ± 14 |
| Data | 15290 | 9996 | 374 |

Fig. 28  Single-lepton channel: post-fit comparison of data and prediction for the four top-ranked input variables in the (5j, 3b) region. The plots include (a) $m_{bb}^{min}$, (b) $m_{tt}^{min}$, (c) $p_{T}^{min}$, and (d) $m_{bb}^{max}$. The first and last bins in all figures contain the underflow and overflow, respectively. The bottom panel displays the ratio of data to the total prediction. The hashed area represents the uncertainty on the background.
**Fig. 29** Single-lepton channel: post-fit comparison of data and prediction for the four top-ranked input variables in the $(5j \geq 4b)$ region. The plots include (a) Centrality, (b) $H_1$, (c) $N^{\text{jet}}_b$ and (d) $m_{\text{tt}}$. The first and last bins in all figures contain the underflow and overflow, respectively. The bottom panel displays the ratio of data to the total prediction. The hashed area represents the uncertainty on the background. The dashed line shows $t\bar{t}H$ signal distribution normalised to background yield. The $t\bar{t}H$ signal yield (solid) is normalised to the fitted $\mu$.

**Fig. 30** Single-lepton channel: post-fit comparison of data and prediction for the four top-ranked input variables in $(\geq 6j, 3b)$ region. The plots include (a) $N^{\text{jet}}_b$, (b) Centrality, (c) $H_1$, and (d) SSSL. The first and last bins in all figures contain the underflow and overflow, respectively. The bottom panel displays the ratio of data to the total prediction. The hashed area represents the uncertainty on the background. The dashed line shows $t\bar{t}H$ signal distribution normalised to background yield. The $t\bar{t}H$ signal yield (solid) is normalised to the fitted $\mu$.

**Fig. 31** Single-lepton channel: post-fit comparison of data and prediction for the four top-ranked input variables in $(\geq 6j, \geq 4b)$ region. The plots include (a) $D1$, (b) Centrality, (c) $p_{T}^{\text{jet}}$, and (d) $H_1$. The first and last bins in all figures contain the underflow and overflow, respectively. The bottom panel displays the ratio of data to the total prediction. The hashed area represents the uncertainty on the background. The dashed line shows $t\bar{t}H$ signal distribution normalised to background yield. The $t\bar{t}H$ signal yield (solid) is normalised to the fitted $\mu$. 
Fig. 32 Dilepton channel: post-fit comparison of data and prediction for the four top-ranked input variables in the \((3j, 3b)\) region. The plots include (a) \(\Delta_{jj}^{\text{max}} \Delta R\), (b) \(m_{jj}^{\text{min}} \eta\), (c) \(H_{T}\), and (d) \(H_{4}\). The first and last bins in all figures contain the underflow and overflow, respectively. The bottom panel displays the ratio of data to the total prediction. The hashed area represents the uncertainty on the background. The dashed line shows \(t\bar{t}H\) signal distribution normalised to background yield. The \(t\bar{t}H\) signal yield (solid) is normalised to the fitted \(\mu\).

Fig. 33 Dilepton channel: post-fit comparison of data and prediction for the four top-ranked input variables in the \((\geq 4j, 3b)\) region. The plots include (a) \(\Delta_{jj}^{\text{max}} \Delta R\), (b) \(N_{H}^{\text{bb}}\), (c) \(m_{jj}^{\text{min}} \eta\), and (d) \(\Delta \Phi_{bb}^{\text{max}} \eta\). The first and last bins in all figures contain the underflow and overflow, respectively. The bottom panel displays the ratio of data to the total prediction. The hashed area represents the uncertainty on the background. The dashed line shows \(t\bar{t}H\) signal distribution normalised to background yield. The \(t\bar{t}H\) signal yield (solid) is normalised to the fitted \(\mu\).

Fig. 34 Dilepton channel: post-fit comparison of data and prediction for the four top-ranked input variables in the \((\geq 4j, \geq 4b)\) region. The plots include (a) \(\Delta_{jj}^{\text{max}} \Delta R\), (b) \(m_{bb}^{\text{min}} \eta\), (c) \(m_{bb}\), and (d) \(\Delta \Phi_{bb}^{\text{min}} \Delta R\). The first and last bins in all figures contain the underflow and overflow, respectively. The bottom panel displays the ratio of data to the total prediction. The hashed area represents the uncertainty on the background. The dashed line shows \(t\bar{t}H\) signal distribution normalised to background yield. The \(t\bar{t}H\) signal yield (solid) is normalised to the fitted \(\mu\).
