Search for the Standard Model Higgs boson produced in association with top quarks and decaying into a $b\bar{b}$ pair in $pp$ collisions at $\sqrt{s} = 13$ TeV with the ATLAS detector

The ATLAS Collaboration

A search for the Standard Model Higgs boson produced in association with a top-quark pair, $t\bar{t}H$, is presented. The analysis uses 36.1 fb$^{-1}$ of $pp$ collision data at $\sqrt{s} = 13$ TeV collected with the ATLAS detector at the Large Hadron Collider in 2015 and 2016. The search targets the $H \rightarrow b\bar{b}$ decay mode. The selected events contain either one or two electrons or muons from the top-quark decays, and are then categorized according to the number of jets and how likely these are to contain $b$-hadrons. Multivariate techniques are used to discriminate between signal and background events, the latter being dominated by $t\bar{t} +$ jets production. For a Higgs boson mass of 125 GeV, the ratio of the measured $t\bar{t}H$ signal cross-section to the Standard Model expectation is found to be $\mu = 0.84^{+0.64}_{-0.61}$. A value of $\mu$ greater than 2.0 is excluded at 95% confidence level while the expected upper limit is $\mu < 1.2$ in the absence of a $t\bar{t}H$ signal.
1 Introduction

After the discovery of the Higgs boson [1–3] in 2012 by the ATLAS [4] and CMS [5] collaborations, attention has turned to more detailed measurements of its properties and couplings as a means of testing the predictions of the Standard Model (SM) [6–8]. In particular, the coupling to the top quark, the heaviest particle in the SM, could be very sensitive to effects of physics beyond the SM (BSM) [9]. Assuming that no BSM particle couples to the Higgs boson, the ATLAS and CMS experiments measured a value of the top-quark’s Yukawa coupling equal to $0.87 \pm 0.15$ times the SM prediction by combining [10] their respective Higgs-boson measurements from the Run 1 dataset collected at center-of-mass energies of 7 TeV and 8 TeV at the Large Hadron Collider (LHC). This measurement relies largely on the gluon–gluon fusion production mode and on the decay mode to photons, which both depend on loop contributions with a top quark. If no assumption is made about the particle content of such loop contributions, then the
The top-quark coupling is only determined through tree-level processes, and a value of $1.4 \pm 0.2$ times the SM prediction is obtained.

Higgs-boson production in association with a pair of top quarks, $t\bar{t}H$, is the most favorable production mode for a direct measurement of the top-quark’s Yukawa coupling [11–14]. Although this production mode only contributes around 1% of the total Higgs-boson production cross-section [15], the top quarks in the final state offer a distinctive signature and allow many Higgs-boson decay modes to be accessed. Of these, the decay to two $b$-quarks is predicted to have a branching fraction of about 58% [15], the largest Higgs-boson decay mode. This decay mode is sensitive to the $b$-quark’s Yukawa coupling, the second largest in the SM. In order to select events at the trigger level and reduce the backgrounds, the analysis targets events in which one or both top quarks decay semi-leptonically, producing an electron or a muon.\footnote{Throughout this document, ‘lepton’ refers to electron or muon, unless otherwise specified. Electrons and muons from the decay of a $\tau$ itself originating from a $W$ boson are included.}

The main experimental challenges for this channel are the low combined efficiency to reconstruct and identify all final-state particles, the combinatorial ambiguity from the many jets containing $b$-hadrons in the final state which makes it difficult to reconstruct the Higgs boson, and the large backgrounds from the production of $t\bar{t} + $ jets especially when the associated jets stem from $b$- or $c$-quarks. Some representative Feynman diagrams for the $t\bar{t}H$ signal are shown in Figure 1, together with the dominant $t\bar{t} + b\bar{b}$ background.

![Feynman diagrams](image)

**Figure 1:** Representative tree-level Feynman diagrams for (a) $t$-channel and (b) $s$-channel production of the Higgs boson in association with a top-quark pair ($t\bar{t}H$) and the subsequent decay of the Higgs boson to $b\bar{b}$, and (c) for the main background, $t\bar{t} + b\bar{b}$.

The ATLAS collaboration searched for $t\bar{t}H$ production with Higgs-boson decays to $b\bar{b}$ at $\sqrt{s} = 8$ TeV, using $t\bar{t}$ decays with at least one lepton [16] or no leptons [17]. A combined signal strength $\mu = \sigma / \sigma_{\text{SM}}$ of $1.4 \pm 1.0$ was measured. The CMS collaboration searched for the same process at $\sqrt{s} = 7$ TeV and $\sqrt{s} = 8$ TeV using $t\bar{t}$ decays with a single-lepton or dilepton in the final state, obtaining a signal strength of $0.7 \pm 1.9$ [18]. These results were combined with each other, and with results for Higgs boson decay to vector bosons, to $\tau$-leptons or to photons [18–20], resulting in an observed (expected) significance of 4.4 (2.0) standard deviations for $t\bar{t}H$ production [10]. The measured signal strength is $2.3^{+0.7}_{-0.6}$.

In this article, a search for $t\bar{t}H$ production with 36.1 fb$^{-1}$ of $pp$ collision data at $\sqrt{s} = 13$ TeV is presented. The analysis targets Higgs-boson decays to $b$-quarks, but all the decay modes are considered and may contribute to the signal. Events with either one or two leptons are taken into account, and exclusive
analysis categories are defined according to the number of leptons, the number of jets, and the value of a $b$-tagging discriminant which provides a measure of how likely a jet is to contain a $b$-hadron. In the single-lepton channel, a specific category, referred to as ‘boosted’ in the following, is designed to select events containing a Higgs boson and with at least one of the two top quarks produced at high transverse momentum. In the analysis categories with the largest signal contributions, multivariate discriminants are used to classify events as more or less signal-like. The signal-rich categories are analyzed together with the signal-depleted ones in a combined profile likelihood fit that simultaneously determines the event yields for the signal and for the most important background components, while constraining the overall background model within the assigned systematic uncertainties. The combination of the results presented in this article with the results from other analyses targeting $t\bar{t}H$ production with different final states is reported in Ref. [21].

The article is organized as follows. The ATLAS detector is described in Section 2. Section 3 summarizes the selection criteria applied to events and physics objects. The signal and background modeling are presented in Section 4. Section 5 describes the event categorization while Section 6 presents the multivariate analysis techniques. The systematic uncertainties are summarized in Section 7. Section 8 presents the results and Section 9 gives the conclusions.

2 ATLAS detector

The ATLAS detector [22] at the LHC covers nearly the entire solid angle$^2$ around the collision point. It consists of an inner tracking detector surrounded by a thin superconducting solenoid magnet producing a 2 T axial magnetic field, electromagnetic and hadronic calorimeters, and an external muon spectrometer (MS) incorporating three large toroid magnet assemblies. The inner detector (ID) consists of a high-granularity silicon pixel detector and a silicon microstrip tracker, together providing precision tracking in the pseudorapidity range $|\eta| < 2.5$, complemented by a straw-tube transition radiation tracker providing tracking and electron identification information for $|\eta| < 2.0$. A new innermost silicon pixel layer, the insertable B-layer [23] (IBL), was added to the detector between Run 1 and Run 2. The IBL improves the ability to identify displaced vertices and thereby significantly improves the $b$-tagging performance [24].

The electromagnetic sampling calorimeter uses lead or copper as the absorber material and liquid argon (LAr) as the active medium, and is divided into barrel ($|\eta| < 1.475$), endcap ($1.375 < |\eta| < 3.2$) and forward ($3.1 < |\eta| < 4.9$) regions. Hadron calorimetry is also based on the sampling technique and covers $|\eta| < 4.9$, with either scintillator tiles or LAr as the active medium and with steel, copper or tungsten as the absorber material. The muon spectrometer measures the deflection of muons with $|\eta| < 2.7$ using multiple layers of high-precision tracking chambers located in a toroidal field. The field integral of the toroids ranges between 2.0 and 6.0 Tm across most of the detector. The muon spectrometer is also instrumented with separate trigger chambers covering $|\eta| < 2.4$. A two-level trigger system [25], using custom hardware followed by a software-based level, is used to reduce the trigger rate to an average of around one kHz for offline storage.

$^2$ ATLAS uses a right-handed coordinate system with its origin at the nominal interaction point (IP) in the center of the detector and the z-axis coinciding with the axis of the beam pipe. The x-axis points from the IP to the center 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 $\eta = -\ln\tan(\theta/2)$. Unless stated otherwise, angular distance is measured in units of $\Delta R \equiv \sqrt{(\Delta \eta)^2 + (\Delta \phi)^2}$.
3 Event selection

Events are selected from $pp$ collisions at $\sqrt{s} = 13$ TeV recorded by the ATLAS detector in 2015 and 2016. Only events for which all relevant subsystems were operational are considered. Events are required to have at least one vertex with two or more tracks with transverse momentum $p_T > 0.4$ GeV. The vertex with the largest sum of the squared transverse momentum of associated tracks is taken as the primary vertex.

The event reconstruction is affected by multiple $pp$ collisions in a single bunch crossing and by collisions in neighboring bunch crossings, referred to as ‘pileup’. The number of interactions per bunch crossing in this dataset ranges from about 8 to 45 interactions. The dataset corresponds to an integrated luminosity of $3.2 \pm 0.1 \text{ fb}^{-1}$ recorded in 2015 and $32.9 \pm 0.7 \text{ fb}^{-1}$ recorded in 2016, for a total of $36.1 \pm 0.8 \text{ fb}^{-1}$ [26].

Events in both the single-lepton and dilepton channels were recorded using single-lepton triggers. Events are required to fire triggers with either low lepton $p_T$ thresholds and a lepton isolation requirement, or with higher thresholds but with a looser identification criterion and without any isolation requirement. The lowest $p_T$ threshold used for muons is 20 (26) GeV in 2015 (2016), while for electrons the threshold is 24 (26) GeV.

Electrons are reconstructed from energy deposits (clusters) in the electromagnetic calorimeter matched to tracks reconstructed in the ID [27, 28] and are required to have $p_T > 10$ GeV and $|\eta| < 2.47$. Candidates in the calorimeter barrel–endcap transition region ($1.37 < |\eta| < 1.52$) are excluded. Electrons must satisfy the loose identification criterion described in Ref. [28], based on a likelihood discriminant combining observables related to the shower shape in the calorimeter and to the track matching the electromagnetic cluster. Muons are reconstructed from either track segments or full tracks in the MS which are matched to tracks in the ID [29]. Tracks are then re-fitted using information from both detector systems. Muons are required to have $p_T > 10$ GeV and $|\eta| < 2.5$. To reduce the contribution of leptons from hadronic decays (non-prompt leptons), both electrons and muons must satisfy isolation criteria based on information from both the tracker and the calorimeter. The loose lepton isolation working point [28, 29] is used. Finally, lepton tracks must match the primary vertex of the event: the longitudinal impact parameter $|\text{IP}_z|$ must be less than 5 mm, while the transverse impact parameter significance, $|\text{IP}_r|/\sigma_{\text{IP}_r}$, must be less than 2.5. Quality criteria are imposed to identify jets arising from non-collision sources or detector noise, and any event containing such a jet is removed [34]. Finally, to reduce the effect of pileup, an additional requirement is made using an algorithm that matches jets with $p_T < 60$ GeV and $|\eta| < 2.4$ to tracks with $p_T > 0.4$ GeV to identify jets consistent with the primary vertex. This algorithm is known as jet vertex tagger [35], referred to as JVT in the remainder of this article.

Jets are reconstructed from three-dimensional topological energy clusters [30] in the calorimeter using the anti-$k_t$ jet algorithm [31] implemented in the FastJet package [32] with a radius parameter of 0.4. Each topological cluster is calibrated to the electromagnetic scale response prior to jet reconstruction. The reconstructed jets are then calibrated to the jet energy scale derived from simulation and in situ corrections based on 13 TeV data [33]. After energy calibration, jets are required to have $p_T > 25$ GeV and $|\eta| < 2.5$. Quality criteria are imposed to identify jets arising from non-collision sources or detector noise, and any event containing such a jet is removed [34]. Finally, to reduce the effect of pileup, an additional requirement is made using an algorithm that matches jets with $p_T < 60$ GeV and $|\eta| < 2.4$ to tracks with $p_T > 0.4$ GeV to identify jets consistent with the primary vertex. This algorithm is known as jet vertex tagger [35], referred to as JVT in the remainder of this article.

Jets are tagged as containing $b$-hadrons through a multivariate $b$-tagging algorithm (MV2c10) that combines information from an impact-parameter-based algorithm, from the explicit reconstruction of an inclusive secondary vertex and from a multi-vertex fitter that attempts to reconstruct the $b$- to $c$-hadron decay chain [36, 37]. This algorithm is optimized to efficiently select jets containing $b$-hadrons ($b$-jets) and separate them from jets containing $c$-hadrons ($c$-jets), jets containing hadronically decaying $\tau$-leptons ($\tau$-jets) and from other jets (light jets). Four working points are defined by different MV2c10 discriminant
output thresholds and are referred to in the following as *loose*, *medium*, *tight* and *very tight*. The efficiency for $b$-jets with $p_T > 20$ GeV in simulated $t\bar{t}$ events to pass the different working points are 85%, 77%, 70% and 60%, respectively, corresponding to rejection factors\(^3\) of $c$-jets in the range 3–35 and of light jets in the range 30–1500. A $b$-tagging discriminant value is assigned to each jet according to which working point it satisfies, ranging from 1 for a jet that does not satisfy any of the $b$-tagging criteria defined by the considered working points up to 5 for jets satisfying the *very tight* criteria. This $b$-tagging discriminant is used to categorize selected events as discussed in Section 5 and as an input to multivariate analysis techniques described in Section 6.

Hadronically decaying $\tau$ leptons ($\tau_{\text{had}}$) are distinguished from jets using the track multiplicity and a multivariate discriminant based on the track collimation, further jet substructure, and kinematic information \([38]\). These $\tau_{\text{had}}$ candidates are required to have $p_T > 25$ GeV, $|\eta| < 2.5$ and pass the *Medium* $\tau$-identification working point.

To avoid counting a single detector response as more than one lepton or jet, an overlap removal procedure is adopted. To prevent double-counting of electron energy deposits as jets, the closest jet within $\Delta R_y = \sqrt{(\Delta y)^2 + (\Delta \phi)^2} = 0.2$ of a selected electron is removed.\(^4\) If the nearest jet surviving that selection is within $\Delta R_y = 0.4$ of the electron, the electron is discarded. Muons are removed if they are separated from the nearest jet by $\Delta R_y < 0.4$, which reduces the background from heavy-flavor decays inside jets. However, if this jet has fewer than three associated tracks, the muon is kept and the jet is removed instead; this avoids an inefficiency for high-energy muons undergoing significant energy loss in the calorimeter. A $\tau_{\text{had}}$ candidate is rejected if it is separated by $\Delta R_y < 0.2$ from any selected electron or muon.

The missing transverse momentum in the event is defined as the negative vector sum of the $p_T$ of all the selected electrons, muons and jets described above, with an extra term added to account for energy in the event which is not associated with any of these. This extra term, referred to as the ‘soft term’ in the following, is calculated from ID tracks matched to the primary vertex to make it resilient to pileup contamination \([39, 40]\). The missing transverse momentum is not used for event selection but is included in the inputs to the multivariate discriminants that are built in the most sensitive analysis categories.

For the boosted category, the selected jets are used as inputs for further jet reclustering \([41]\) through an anti-$k_t$ algorithm with a radius parameter of $R = 1.0$, resulting in a collection of large-$R$ jets. Large-$R$ jets with a reconstructed invariant mass lower than 50 GeV are removed. The resulting large-$R$ jets are used to identify top quarks and Higgs bosons in signal events when these have high transverse momenta (boosted) and decay into collimated hadronic final states. Boosted Higgs-boson candidates are required to have $p_T > 200$ GeV and contain at least two constituent jets, among which at least two are $b$-tagged at the *loose* working point. If more than one boosted Higgs-boson candidate is identified, the one with the highest sum of constituent-jet $b$-tagging discriminants is selected. Additional large-$R$ jets are considered as potential boosted top-quark candidates. Boosted top-quark candidates are required to have $p_T > 250$ GeV, exactly one constituent jet satisfying the *loose* $b$-tagging working point plus at least one additional constituent jet not $b$-tagged. If more than one boosted top-quark candidate is identified, the one with the highest mass is selected.

Events are required to have at least one reconstructed lepton with $p_T > 27$ GeV matching a lepton with the same flavor reconstructed by the trigger algorithm within $\Delta R < 0.15$. Events in the dilepton channel must have exactly two leptons with opposite electric charge. The subleading lepton $p_T$ must be above

\(^3\) The rejection factor is defined as the inverse of the efficiency to pass a given $b$-tagging working point.

\(^4\) The rapidity is defined as $y = \frac{1}{2} \ln \frac{E + p_T}{E - p_T}$, where $E$ is the energy and $p_T$ is the longitudinal component of the momentum along the beam pipe.
15 GeV in the \( ee \) channel or above 10 GeV in the \( e\mu \) and \( \mu\mu \) channels. In the \( ee \) and \( \mu\mu \) channels, the dilepton invariant mass must be above 15 GeV and outside of the Z-boson mass window 83–99 GeV. To maintain orthogonality with other \( t\bar{t}H \) search channels [21], dilepton events are vetoed if they contain one or more \( \tau_{\text{had}} \) candidates. Events enter the single-lepton channel if they contain exactly one lepton with \( p_T > 27 \text{ GeV} \) and no other selected leptons with \( p_T > 10 \text{ GeV} \). In the single-lepton channel, events are removed if they contain two or more \( \tau_{\text{had}} \) candidates.

To improve the purity in events passing the above selection, selected leptons are further required to satisfy additional identification and isolation criteria, otherwise the corresponding events are removed. For electrons, the \textit{tight} identification criterion based on a likelihood discriminant [28] is used, while for muons the \textit{medium} identification criterion [29] is used. Both the electrons and muons are required to satisfy the Gradient isolation criteria [28, 29], which become more stringent as the \( p_T \) of the leptons considered drops.

Finally, events in the dilepton channel must have at least three jets, of which at least two must be \( b \)-tagged at the \textit{medium} working point. Single-lepton events containing at least one boosted Higgs-boson candidate, at least one boosted top-quark candidate and at least one additional jet \( b \)-tagged at the \textit{loose} working point enter the boosted category. Events that do not enter the boosted category and have at least five jets, with at least two of them \( b \)-tagged at the \textit{very tight} working point or three of them \( b \)-tagged at the \textit{medium} working point, are classified as ‘resolved’ single-lepton events. The fraction of simulated \( t\bar{t}H(H \rightarrow b\bar{b}) \) events passing the dilepton event selection is 2.5%. These fractions are 8.7% for the resolved single-lepton channel and 0.1% for the boosted category.

### 4 Signal and background modeling

This section describes the simulation and data-driven techniques used to model the \( t\bar{t}H \) signal and the background processes, to train the multivariate discriminants and to define the templates for the signal extraction fit. In this analysis, most Monte Carlo (MC) samples were produced using the full ATLAS detector simulation [42] based on Geant4 [43]. A faster simulation, where the full Geant 4 simulation of the calorimeter response is replaced by a detailed parameterization of the shower shapes [44], was adopted for some of the samples used to estimate modeling systematic uncertainties. To simulate the effects of pileup, additional interactions were generated using Pythia 8.186 [45] and overlaid onto the simulated hard-scatter event. Simulated events are reweighted to match the pileup conditions observed in the data. All simulated events are processed through the same reconstruction algorithms and analysis chain as the data. In the simulation, the top-quark mass is assumed to be \( m_t = 172.5 \text{ GeV} \). Decays of \( b \)- and \( c \)-hadrons were performed by EvtGen v1.2.0 [46], except in samples simulated by the Sherpa event generator.

#### 4.1 Signal modeling

The \( t\bar{t}H \) signal process was modeled using MadGraph5_aMC@NLO [47] (referred to in the following as MG5_aMC@NLO) version 2.3.2 for the matrix element (ME) calculation at next-to-leading-order (NLO) accuracy in quantum chromodynamics (QCD), interfaced to the Pythia 8.210 parton shower (PS) and hadronization model using the A14 set of tuned parameters [48]. The NNPDF3.0NLO parton distribution function (PDF) set [49] was used, and the factorization and renormalization scales were set to \( \mu_F = \mu_R = H_T / 2 \), with \( H_T \) defined as the scalar sum of the transverse masses \( \sqrt{p_T^2 + m^2} \) of all final-state
particles. The top quarks were decayed using MADSpin [50], preserving all spin correlations. The Higgs-boson mass was set to 125 GeV and all decay modes were considered. The $t\bar{t}H$ cross-section of $507^{+50}_{-50}$ pb was computed [15, 51–55] at NLO accuracy in QCD and includes NLO electroweak corrections. The branching fractions were calculated using HDECAY [15, 56].

4.2 $t\bar{t}$ + jets background

The nominal sample used to model the $t\bar{t}$ background was generated using the Powheg-Box v2 NLO event generator [57–60], referred to as Powheg in the remainder of this article, with the NNPDF3.0NLO PDF set. The $h_{\text{damp}}$ parameter, which controls the transverse momentum of the first gluon emission beyond the Born configuration, was set to 1.5 times the top-quark mass [61]. The parton shower and the hadronization were modeled by Pythia 8.210 with the A14 set of tuned parameters. The renormalization and factorization scales were set to the transverse mass of the top quark, defined as $m_{T,t} = \sqrt{m_t^2 + p_{T,t}^2}$, where $p_{T,t}$ is the transverse momentum of the top quark in the $t\bar{t}$ center-of-mass reference frame. The sample is normalized using the predicted cross-section of $32^{+46}_{-51}$ pb, calculated with the Top++2.0 program [62] at next-to-next-to-leading order (NNLO) in perturbative QCD including resummation of next-to-next-to-leading logarithmic (NNLL) soft gluon terms [63–66]. Alternative $t\bar{t}$ samples used to derive systematic uncertainties are described in Section 7.

The $t\bar{t}$ + jets background is categorized according to the flavor of additional jets in the event, using the same procedure as described in Ref. [16]. Generator-level particle jets are reconstructed from stable particles (mean lifetime $\tau > 3 \times 10^{-11}$ seconds) using the anti-$k_t$ algorithm with a radius parameter $R = 0.4$, and are required to have $p_T > 15$ GeV and $|\eta| < 2.5$. This categorization employs a jet flavor-labeling procedure that is more refined than the one described in Section 3. The flavor of a jet is determined by counting the number of $b$- or $c$-hadrons within $\Delta R < 0.4$ of the jet axis. Jets matched to exactly one $b$-hadron, with $p_T$ above 5 GeV, are labeled single-$b$-jets, while those matched to two or more $b$-hadrons are labeled $B$-jets (with no $p_T$ requirement on the second hadron); single-$c$- and $C$-jets are defined analogously, only considering jets not already defined as single-$b$- or $B$-jets. Events that have at least one single-$b$- or $B$-jet, not counting heavy-flavor jets from top-quark or $W$-boson decays, are labeled as $t\bar{t} + \geq 1b$; those with no single-$b$- or $B$-jet but at least one single-$c$- or $C$-jet are labeled as $t\bar{t} + \geq 1c$. Finally, events not containing any heavy-flavor jets aside from those from top-quark or $W$-boson decays are labeled as $t\bar{t} + \text{light}$. This classification is used to define the background categories in the likelihood fit. A finer classification is then used to assign correction factors and estimate uncertainties: events with exactly two single-$b$-jets are labeled as $t\bar{t} + b\bar{b}$, those with only one single-$b$-jet are labeled as $t\bar{t} + b$, and those with only one $B$-jet are labeled as $t\bar{t} + B$, the rest of the $t\bar{t} + \geq 1b$ events being labeled as $t\bar{t} + \geq 3b$. Events with additional $b$-jets entirely originating from multi-parton interactions (MPI) or $b$-jets from final-state radiation (FSR), i.e. originating from gluon radiation from the top-quark decay products, are considered separately in the $t\bar{t} + b$ (MPI/FSR) subcategory. Background events from $t\bar{t}$ containing extra $c$-jets are divided analogously.

To model the dominant $t\bar{t} + \geq 1b$ background with the highest available precision, the relative contributions of the different subcategories, $t\bar{t} + \geq 3b$, $t\bar{t} + b\bar{b}$, $t\bar{t} + B$ and $t\bar{t} + b$, in the Powheg+Pythia 8 sample described above are scaled to match the predictions of an NLO $t\bar{t}b\bar{b}$ sample including parton showering and hadronization [67], generated with Sherpa+OpenLoops [68, 69]. The sample was produced with Sherpa version 2.1.1 and the CT10 four-flavor (4F) scheme PDF set [70, 71]. The renormalization scale for this sample was set to the CMMPS value, $\mu_{\text{CMMPS}} = \prod_{i=t,t,b,b} E_{T,i}^{1/4}$ [67], while the factorization
scale was set to $H_T/2 = \frac{1}{2} \sum_{i=t,t,b} E_{T,i}$. The resummation scale $\mu_Q$, which sets an upper bound for the hardness of the parton-shower emissions, was also set to $H_T/2$. This sample, referred to as ‘SHERPA4F’ in the remainder of this article, employs a description of the kinematics of the two additional $b$-jets with NLO precision in QCD, taking into account the $b$-quark mass, and is therefore the most precise MC prediction for the $tt+\geq 1b$ process available at present. Topologies that are not included in this NLO calculation but are labeled as $tt+b$, i.e. events in the $tt + b$ (MPI/FSR) subcategory, are not scaled.

Figure 2 shows the predicted fractions for each of the $tt+\geq 1b$ subcategories, with the POWHEG+PYTHIA 8 inclusive $tt$ sample compared to the $tt + bb$ SHERPA4F sample. The $tt + b$ (MPI/FSR) subcategory is not present in the $tt + bb$ SHERPA4F sample and accounts for 10% of the events in the POWHEG+PYTHIA 8 $tt+\geq 1b$ sample.

![Figure 2: The relative predicted fractions of the $tt + b$, $tt + bb$, $tt + B$ and $tt + \geq 3b$ subcategories before any event selection. The prediction from the inclusive POWHEG+PYTHIA 8 sample is compared to the four-flavor $ttbb$ calculation from SHERPA4F, with its uncertainties (from a combination of the sources discussed in Section 7) shown as the shaded area. The fractions are normalized to the sum of the four contributions shown here, without considering the $tt + b$ (MPI/FSR) subcategory as part of the total.]

### 4.3 Other backgrounds

Samples of $t\bar{t}W$ and $t\bar{t}Z$ ($t\bar{t}V$) events were generated with an NLO matrix element using MG5_aMC@NLO interfaced to PYTHIA 8.210 with the NNPDF3.0NLO PDF and the A14 parameter set.

Samples of $Wt$ and $s$-channel single-top-quark backgrounds were generated with POWHEG-Box v1 at NLO accuracy using the CT10 PDF set. Overlap between the $tt$ and $Wt$ final states was handled using the ‘diagram removal’ scheme [72]. The $t$-channel single-top-quark events were generated using the POWHEG-Box v1 event generator at NLO accuracy with the four-flavor PDF set CT10 4F. For this process, the top quarks were decayed using MADSPIN. All single-top-quark samples were interfaced to PYTHIA 6.428 [73] with the Perugia 2012 set of tuned parameters [74]. The single-top-quark $Wt$, $t$- and $s$-channel samples are normalized using the approximate NNLO theoretical cross-sections [75–77].
Samples of $W/Z$ production in association with jets were generated using Sherpa 2.2.1. The matrix elements were calculated for up to two partons at NLO and four partons at leading order (LO) using Comix [78] and OpenLoops, and merged with the Sherpa parton shower [79] using the ME+PS@NLO prescription [80]. The NNPDF3.0NNLO PDF set was used in conjunction with dedicated parton-shower tuning. The $W/Z +$ jet events are normalized using the NNLO cross-sections [81]. For $Z +$ jet events, the normalization of the heavy-flavor component is corrected by a factor 1.3, extracted from dedicated control regions in data, defined by requiring two opposite-charge same-flavor leptons ($e^+e^-$ or $\mu^+\mu^-$) with an invariant mass, $m_{\ell\ell}$, inside the $Z$-boson mass window 83–99 GeV. The diboson + jet samples were generated using Sherpa 2.1.1 as described in Ref. [82].

Higgs-boson production in association with a single top quark is rare in the SM, but is included in the analysis and treated as background. Samples of single top quarks produced in association with a $W$ boson and with a Higgs boson, $tWH$, were produced with MG5_aMC@NLO interfaced to Herwig++ [83] with the CTEQ6L1 PDF set. Samples of single top quarks plus Higgs boson plus jets, $tHq\bar{b}$, were produced at LO with MG5_aMC@NLO interfaced to Pythia 8, using the CT10 4F scheme PDF set. The other Higgs-boson production modes were found to be negligible and are not considered. Four-top production ($ttt\bar{t}$) as well as $t\bar{t}WW$ events were generated with MG5_aMC@NLO with LO accuracy and interfaced with Pythia 8. Events from $tZ$ production were also generated with MG5_aMC@NLO with LO accuracy, but interfaced with Pythia 6. The process $tZ\bar{W}$ was also generated with MG5_aMC@NLO interfaced with Pythia 8, but with NLO accuracy.

In the single-lepton channel, the background from events with a jet or a photon misidentified as a lepton (hereafter referred to as fake lepton) or non-prompt lepton is estimated directly from data using a matrix method [84]. A data sample enhanced in fake and non-prompt leptons is selected by removing the lepton isolation requirements and, for electrons, loosening the identification criteria. Next, the efficiency for these ‘loose’ leptons to satisfy the nominal selection (‘tight’) criteria is measured in data, separately for real prompt leptons and for fake or non-prompt leptons. For real prompt leptons the efficiency is measured in $Z$-boson events, while for fake and non-prompt leptons it is estimated from events with low missing transverse momentum and low values of the reconstructed leptonic $W$-boson transverse mass.\(^5\) With this information, the number of fake or non-prompt leptons satisfying the tight criteria can be calculated by inverting the matrix defined by the two equations:

\[
N^1 = N_t^1 + N_f^1, \quad N^1 = \epsilon_t N_t^1 + \epsilon_f N_f^1,
\]

where $N^1$ ($N^1$) is the number of events observed in data passing the loose (tight) lepton selection, $N_t^1$ ($N_f^1$) is the number of events with a real prompt (fake or non-prompt) lepton in the loose lepton sample, and $\epsilon_t$ ($\epsilon_f$) is the efficiency for these events to pass the tight lepton selection. By generalizing the resulting formula to extract $\epsilon_t N_t^1$, a weight is assigned to each event selected in the loose lepton data sample, providing a prediction for both the yields and the kinematic distribution shapes for the fake and non-prompt lepton background. In the three most sensitive single-lepton signal regions, $SR_1^{\pm 6j}$, $SR_2^{\pm 6j}$ and $SR_3^{ij}$ (see Section 5), the contribution from events with a fake or non-prompt lepton is found to be very small, consistent with zero, and is neglected. In the dilepton channel, this background is estimated from simulation and is normalized to data in a control region with two same-sign leptons.

\(^5\) The reconstructed leptonic $W$-boson transverse mass is defined as $\sqrt{\rho_T^{\text{lepton}} E_T^{\text{miss}} (1 - \cos \Delta \phi)}$, where $\rho_T^{\text{lepton}}$ is the transverse momentum of the selected lepton, $E_T^{\text{miss}}$ is the magnitude of the missing transverse momentum and $\Delta \phi$ is the azimuthal angle between the lepton and the missing transverse momentum.
All background samples described in this section, apart from the \( t\bar{t}V \) samples, are referred to as ‘non-\( t\bar{t} \)’ and grouped together in the figures and tables. The contribution to the total background prediction from non-\( t\bar{t} \) varies between 4% and 15% depending on the considered signal or control region, as can be seen in Appendix A.

5 Event categorization

After the selection, the data sample is dominated by background from \( t\bar{t} \) events. In order to take advantage of the higher jet and \( b \)-jet multiplicities of the \( t\bar{t}H \) signal process, events are classified into non-overlapping analysis categories based on the total number of jets, as well as the number of \( b \)-tagged jets at the four working points. Events in the boosted single-lepton category are not further categorized due to the small number of selected events in this category. Events in the dilepton (resolved single-lepton) channel are first classified according to whether the number of jets is exactly three (five) or at least four (six). These events are then further subdivided into analysis categories, depending on the number of jets tagged at the four \( b \)-tagging working points, or, equivalently, on the values of the \( b \)-tagging discriminant for the jets. The \( b \)-tagging requirements are optimized in order to obtain categories enriched in one of the relevant sample components: \( t\bar{t}H \) plus \( t\bar{t} + b\bar{b}, \ t\bar{t} + b, \ t\bar{t} + \geq 1c, \) and \( t\bar{t} + \) light. The analysis categories where \( t\bar{t}H \) and \( t\bar{t} + b\bar{b} \) are enhanced relative to the other backgrounds are referred to as ‘signal regions’; in these, multivariate techniques are used to further separate the \( t\bar{t}H \) signal from the background events. The remaining analysis categories are referred to as ‘control regions’; no attempt is made to separate the signal from the background in these analysis categories, but they provide stringent constraints on backgrounds and systematic uncertainties in a combined fit with the signal regions.

In the dilepton channel, three signal regions are defined, with different levels of purity for the \( t\bar{t}H \) and \( t\bar{t} + b\bar{b} \) components. The signal region with the highest \( t\bar{t}H \) signal purity, referred to as \( \text{SR}_{1}^{\geq 4j} \), is defined by requiring at least four jets of which three are \( b \)-tagged at the \textit{very tight} working point and another one is \( b \)-tagged at the \textit{tight} working point. The other two signal regions, \( \text{SR}_{2}^{\geq 4j} \) and \( \text{SR}_{3}^{\geq 4j} \), are defined with looser \( b \)-tagging requirements. The remaining dilepton events with at least four jets are divided into two control regions, one enriched in \( t\bar{t} + \) light, \( \text{CR}_{t\bar{t}+\text{light}}^{\geq 4j} \), and one in \( t\bar{t} + \geq 1c, \text{CR}_{t\bar{t}+\geq 1c}^{\geq 4j} \). Dilepton events with three jets are split into two control regions, \( \text{CR}_{t\bar{t}+\text{light}}^{3j} \) and \( \text{CR}_{t\bar{t}+\geq 1b}^{3j} \), enriched in \( t\bar{t} + \) light and \( t\bar{t} + \geq 1b \), respectively. The detailed definition of the signal and control regions for the dilepton channel is presented in Figure 3.

In the single-lepton channel, five signal regions are formed from events passing the resolved selection, three requiring at least six jets, and the other two requiring exactly five jets. They are referred to as \( \text{SR}_{1}^{\geq 6j}, \ \text{SR}_{2}^{\geq 6j}, \ \text{SR}_{3}^{\geq 6j}, \ \text{SR}_{1}^{5j}, \) and \( \text{SR}_{2}^{5j} \). The two purest signal regions, \( \text{SR}_{1}^{\geq 6j} \) and \( \text{SR}_{1}^{5j} \), require four \( b \)-tagged jets at the \textit{very tight} working point, while looser requirements are applied in the other signal regions. Events passing the boosted single-lepton selection form a sixth signal region, \( \text{SR}_{\text{boosted}}^{4j} \). The remaining events with at least six jets are then categorized into three control regions enriched in \( t\bar{t} + \) light, \( t\bar{t} + \geq 1c \) and \( t\bar{t} + b \), referred to as \( \text{CR}_{t\bar{t}+\text{light}}^{\geq 6j}, \text{CR}_{t\bar{t}+\geq 1c}^{\geq 6j}, \text{CR}_{t\bar{t}+b}^{\geq 6j} \), respectively. Analogously, remaining events with exactly five jets are categorized into other three control regions, referred to as \( \text{CR}_{t\bar{t}+\text{light}}^{5j}, \text{CR}_{t\bar{t}+\geq 1c}^{5j} \) and \( \text{CR}_{t\bar{t}+b}^{5j} \). The detailed definition of the signal and control regions for the resolved single-lepton channel is presented in Figure 4.

Figures 5 and 6 show, respectively, the fraction of the different background components as well as the \( t\bar{t}H \) signal purity for each of the signal and control regions in the dilepton and single-lepton channels. The
Dilepton, $3 \geq \text{jet}$

$CR_t \bar{t} + \geq 1b$

$CR_t \bar{t} + \text{light}$

$3\text{rd jet}$

$\text{bbb-tagging discriminant}$

(1st, 2nd) jet $\text{bbb-tagging discriminant}$

(a)

Figure 3: Definition of the (a) three-jet and (b) four-jet signal and control regions in the dilepton channel, as a function of the $b$-tagging discriminant defined in Section 3. The vertical axis shows the values of the $b$-tagging discriminant for the first two jets, while the horizontal axis shows these values for (a) the third jet or (b) the third and fourth jets. The jets are ordered according to their value of the $b$-tagging discriminant in descending order.

$H \rightarrow b\bar{b}$ decay represents 89% of the $t\bar{t}H$ signal events in the signal regions of the dilepton channel, 96% in the signal regions of the resolved single-lepton channel and 86% in the boosted signal region.
Figure 4: Definition of the (a) five-jet and (b) six-jet signal and control regions in the single-lepton resolved channel, as a function of the $b$-tagging discriminant defined in Section 3. The vertical axis shows the values of the $b$-tagging discriminant for the first two jets, while the horizontal axis shows these values for the third and fourth jets. The jets are ordered according to their value of the $b$-tagging discriminant in descending order.

6 Multivariate analysis techniques

In each of the signal regions, a boosted decision tree (BDT) is exploited to discriminate between the $t\bar{t}H$ signal and the backgrounds. This BDT is referred to as the ‘classification BDT’ in the following. The distributions of the classification BDTs in the signal regions are used as the final discriminants for the profile likelihood fit described in Section 8. In the control regions, the overall event yield is used as input to the fit, except in those enriched in $t\bar{t} + \geq 1c$ in the single-lepton channel, $CR_{t\bar{t}+\geq 1c}^{3j}$ and $CR_{t\bar{t}+\geq 1c}^{\geq 6j}$; in
Figure 5: Fractional contributions of the various backgrounds to the total background prediction in each analysis category (a) in the dilepton channel and (b) in the single-lepton channel. The predictions for the various background contributions are obtained through the simulation and the data-driven estimates described in Section 4. The $t\bar{t}$ background is divided as described in Section 4.

these two control regions, the distribution of the scalar sum of the $p_T$ of the jets, $H_T^{\text{had}}$, is used to further control the $t\bar{t} + \geq 1c$ background.

The final state of the $t\bar{t}H(H \rightarrow b\bar{b})$ process is composed of many jets stemming from the Higgs-boson and top-quark decay products, as well as from additional radiation. Many combinations of these jets are possible when reconstructing the Higgs-boson and top-quark candidates to explore their properties and the signal event topology. To enhance the signal separation, three intermediate multivariate techniques are implemented prior to the classification BDT: (a) the ‘reconstruction BDT’ used to select the best combination of jet–parton assignments in each event and to build the Higgs-boson and top-quark candidates, (b) a likelihood discriminant (LHD) method that combines the signal and background probabilities of all possible combinations in each event, (c) a matrix element method (MEM) that exploits the full matrix element calculation to separate the signal from the background. The outputs of the three intermediate multivariate methods are used as input variables to the classification BDT in one or more of the signal regions. The properties of the Higgs-boson and top-quark candidates from the reconstruction BDT are used to define additional input variables to the classification BDT. Although the intermediate techniques exploit similar information, they make use of this information from different perspectives and based on
Figure 6: The ratios $S/B$ (black solid line, referring to the vertical axis on the left) and $S/\sqrt{B}$ (red dashed line, referring to the vertical axis on the right) for each of the analysis categories (a) in the dilepton channel and (b) in the single-lepton channel, where $S$ ($B$) is the number of selected signal (background) events predicted by the simulation or through the data-driven estimates as described in Section 4.

different assumptions, so that their combination further improves the separation power of the classification BDT. Details of the implementation of these multivariate techniques are described in Sections 6.1–6.4.

6.1 Classification BDT

The classification BDT is trained to separate the signal from the $t\bar{t}$ background on a sample that is statistically independent of the sample used for the evaluation. The toolkit for multivariate analysis (TMVA) [85] is used to train both this and the reconstruction BDT. The classification BDT is built by combining several input variables that exploit the different kinematics of signal and background events, as well as the $b$-tagging information. General kinematic variables, such as invariant masses and angular separations of pairs of reconstructed jets and leptons, are combined with outputs of the intermediate multivariate discriminants and the $b$-tagging discriminants of the selected jets. In the case of the boosted single-lepton signal region, kinematic variables are built from the properties of the large-$R$ jets and their jet constituents. The input variables to the classification BDT in each of the signal regions are listed in Appendix B. The input variables are selected to maximize the performance of the classification BDT; however, only variables with good modeling of data by simulation are considered. The output of the reconstruction BDT, the LHD and the MEM represent the most powerful variables in the classification BDT.

6.2 Reconstruction BDT

The reconstruction BDT is employed in all dilepton and resolved single-lepton signal regions. It is trained to match reconstructed jets to the partons emitted from top-quark and Higgs-boson decays. For this purpose, $W$-boson, top-quark and Higgs-boson candidates are built from combinations of jets and leptons. The $b$-tagging information is used to discard combinations containing jet–parton assignments inconsistent with the correct parton candidate flavor.
In the single-lepton channel, leptonically decaying $W$-boson candidates are assembled from the lepton four-momentum ($p_\ell$) and the neutrino four-momentum ($p_\nu$); the latter is built from the missing transverse momentum, its $z$ component being inferred by solving the equation $m_W^2 = (p_\ell + p_\nu)^2$, where $m_W$ represents the $W$-boson mass. Both solutions of this quadratic equation are used in separate combinations. If no real solutions exist, the discriminant of the quadratic equation is set to zero, giving a unique solution. The hadronically decaying $W$-boson and the Higgs-boson candidates are each formed from a pair of jets. The top-quark candidates are formed from one $W$-boson candidate and one jet. The top-quark candidate containing the hadronically (leptonically) decaying $W$ boson is referred to as the hadronically (leptonically) decaying top-quark candidate. In the single-lepton signal regions with exactly five selected jets, more than 70% of the events do not contain both jets from the hadronically decaying $W$ boson. Therefore, the hadronically decaying top-quark candidate is assembled from two jets, one of which is $b$-tagged. In the dilepton channel, no attempt to build leptonically decaying $W$-boson candidates is made and the top-quark candidates are formed by one lepton and one jet.

Simulated $t\bar{t}H$ events are used to iterate over all allowed combinations. The reconstruction BDT is trained to distinguish between correct and incorrect jet assignments, using invariant masses and angular separations in addition to other kinematic variables as inputs. In each event a specific combination of jet–parton assignments, corresponding to the best BDT output, is chosen in order to compute kinematic and topological information of the top-quark and Higgs-boson candidates to be input to the classification BDT. However, although the best possible reconstruction performance can be obtained by including information related to the Higgs boson, such as the candidate Higgs-boson invariant mass, in the reconstruction BDT, this biases the background distributions of these Higgs-boson-related observables in the chosen jet–parton assignment towards the signal expectation, reducing their ability to separate signal from background. For this reason, two versions of the reconstruction BDT are used, one with and one without the Higgs-boson information and the resulting jet–parton assignments from one, the other or both are considered when computing input variables for the classification BDT, as detailed in Appendix B.

The Higgs boson is correctly reconstructed in 48% (32%) of the selected $t\bar{t}H$ events in the single-lepton channel $\text{SR}_{\geq 6j}$ using the reconstruction BDT with (without) information about the Higgs-boson kinematics included. For the dilepton channel, the corresponding reconstruction efficiencies are 49% (32%) in $\text{SR}_{\geq 4j}$. The reconstruction techniques are not needed in the signal region $\text{SR}_{\text{boosted}}$, as the Higgs-boson and the top-quark candidates are chosen as the selected large-$R$ jets described in Section 3. The large-$R$ jet selected as a Higgs-boson candidate contains two $b$-tagged jets stemming from the decay of a Higgs boson in 47% of the selected $t\bar{t}H$ events.

### 6.3 Likelihood discriminant

In the resolved single-lepton signal regions, the output from a likelihood discriminant is included as an additional input variable for the classification BDT. The LHD is computed analogously to Ref. [86] as a product of one-dimensional probability density functions, pdfs, for the signal and the background hypotheses. The pdfs are built for various invariant masses and angular distributions from reconstructed jets and leptons and from the missing transverse momentum, in a similar way to those used in the reconstruction BDT.

Two background hypotheses are considered, corresponding to the production of $t\bar{t} + \geq 2 \text{ } b$-jets and $t\bar{t} + \text{ exactly one } b$-jet, respectively. The likelihoods for both hypotheses are averaged, weighted by their relative fractions in simulated $t\bar{t} + \text{ jets}$ events. In a significant fraction of both the $t\bar{t}H$ and $t\bar{t}$ simulated events with
at least six selected jets, only one jet stemming from the hadronically decaying $W$ boson is selected. An additional hypothesis, for both the signal and the background, is considered to account for this topology. In events with exactly five selected jets, variables including the hadronically decaying top-quark candidate are built similarly to those for the reconstruction BDT.

The probabilities $p^{\text{sig}}$ and $p^{\text{bkg}}$, for signal and background hypotheses, respectively, are obtained as the product of the pdfs for the different kinematic distributions, averaged among all possible jet–parton matching combinations. Combinations are weighted using the $b$-tagging information to suppress the impact from parton–jet assignments that are inconsistent with the correct parton candidates flavor. For each event, the discriminant is defined as the ratio of the probability $p^{\text{sig}}$ to the sum of $p^{\text{sig}}$ and $p^{\text{bkg}}$, and added as an input variable to the classification BDT. As opposed to the reconstruction BDT method, the LHD method takes advantage of all possible combinations in the event, but it does not fully account for correlations between variables in one combination, as it uses a product of one-dimensional pdfs.

### 6.4 Matrix element method

A discriminant (MEM$_{D1}$) based on the MEM is computed following a method similar to the one described in Ref. [16] and is included as another input to the classification BDT. The MEM consumes a significant amount of computation time and thus is implemented only in the most sensitive single-lepton signal region, SR$^{6\ell}$. The degree to which each event is consistent with the signal and background hypotheses is expressed via signal and background likelihoods, referred to as $L_S$ and $L_B$, respectively. These are computed using matrix element calculations at parton level rather than using simulated MC samples as for the LHD method. The matrix element evaluation is performed with MG5_aMC@NLO at the LO accuracy. The $t\bar{t}H(H \to b\bar{b})$ process is used as a signal hypothesis, while $t\bar{t} + b\bar{b}$ is used as a background hypothesis. To reduce computation time, only diagrams representing gluon-induced processes are considered. The parton distribution functions are modeled with the CT10 PDF set, interfaced via the LHAPDF package [87]. Transfer functions, that map the detector quantities to the parton level quantities, are derived from a $t\bar{t}$ sample generated with Powheg+Pythia 6 and validated with the nominal Powheg+Pythia 8 $t\bar{t}$ sample. The directions in $\eta$ and $\phi$ of all visible final-state objects are assumed to be well measured, and their transfer functions are thus represented by $\delta$-functions. The neutrino momentum is constrained by imposing transverse momentum conservation in each event, while its $p_z$ is integrated over. The integration is performed using VEGAS [88], following the implementation described in Ref. [89]. As in the reconstruction BDT, $b$-tagging information is used to reduce the number of jet–parton assignments considered in the calculation. The discriminating variable, MEM$_{D1}$, is defined as the difference between the logarithms of the signal and background likelihoods: $\text{MEM}_{D1} = \log_{10}(L_S) - \log_{10}(L_B)$.

### 7 Systematic uncertainties

Many sources of systematic uncertainty affect the search, including those related to the luminosity, the reconstruction and identification of leptons and jets, and the theory modeling of signal and background processes. Different uncertainties may affect only the overall normalization of the samples, or also the shapes of the distributions used to categorize the events and to build the final discriminants. All the sources of experimental uncertainty considered, with the exception of the uncertainty in the luminosity, affect both the normalizations and the shapes of distributions in all the simulated samples. Uncertainties related to modeling of the signal and the backgrounds affect both the normalizations and the shapes
of the distributions for the processes involved, with the exception of cross-section and normalization uncertainties that affect only the normalization of the considered sample. Nonetheless, the normalization uncertainties modify the relative fractions of the different samples leading to a shape uncertainty in the distribution of the final discriminant for the total prediction in the different analysis categories.

A single independent nuisance parameter is assigned to each source of systematic uncertainty, as described in Section 8. Some of the systematic uncertainties, in particular most of the experimental uncertainties, are decomposed into several independent sources, as specified in the following. Each individual source then has a correlated effect across all the channels, analysis categories, signal and background samples. For modeling uncertainties, especially $t\bar{t}$ modeling, additional nuisance parameters are included to split some uncertainties into several sources independently affecting different subcomponents of a particular process.

7.1 Experimental uncertainties

The uncertainty of the combined 2015+2016 integrated luminosity is 2.1%. It is derived, following a methodology similar to that detailed in Ref. [26], from a calibration of the luminosity scale using $x$–$y$ beam-separation scans performed in August 2015 and May 2016. A variation in the pileup reweighting of MC events is included to cover the uncertainty in the ratio of the predicted and measured inelastic cross-sections in the fiducial volume defined by $M_X > 13$ GeV where $M_X$ is the mass of the hadronic system [90].

The jet energy scale and its uncertainty are derived by combining information from test-beam data, LHC collision data and simulation [33]. The uncertainties from these measurements are factorized into eight independent sources. Additional uncertainties are considered, related to jet flavor, pileup corrections, $\eta$ dependence, and high-$p_T$ jets, yielding a total of 20 independent sources. Although the uncertainties are not large, totaling 1%–6% per jet (depending on the jet $p_T$), the effects are amplified by the large number of jets in the final state. Uncertainties in the jet energy resolution and in the efficiency to pass the JVT requirement that is meant to remove jets from pileup are also considered. The jet energy resolution is divided into two independent components.

The efficiency to correctly tag $b$-jets is measured in data using dileptonic $t\bar{t}$ events. The mis-tag rate for $c$-jets is also measured in $t\bar{t}$ events, identifying hadronic decays of $W$ bosons including $c$-jets, while for light jets it is measured in multi-jet events using jets containing secondary vertices and tracks with impact parameters consistent with a negative lifetime [36]. The $b$-tagging efficiencies and mis-tag rates are first extracted for each of the four working points used in the analysis as a function of jet kinematics, and then combined into a calibration of the $b$-tagging discriminant distribution, with corresponding uncertainties that correctly describe correlations across multiple working points. The uncertainty associated with the $b$-tagging efficiency, whose size ranges between 2% and 10% depending on the working point and on the jet $p_T$, is factorized into 30 independent sources. The size of the uncertainties associated with the mis-tag rates is 5%–20% for $c$-jets depending on the working point and on the jet $p_T$, and 10%–50% for light jets depending on the working point and on the jet $p_T$ and $\eta$. These uncertainties are factorized into 15 (80) independent sources for $c$-jets (light jets). Jets from $\tau_{\text{had}}$ candidates are treated as $c$-jets for the mis-tag rate corrections and systematic uncertainties. An additional source of systematic uncertainty is considered on the extrapolation between $c$-jets and these $\tau$-jets.

Uncertainties associated with leptons arise from the trigger, reconstruction, identification, and isolation efficiencies, as well as the lepton momentum scale and resolution. These are measured in data using
leptons in $Z \to \ell^+\ell^-$, $J/\psi \to \ell^+\ell^-$ and $W \to e\nu$ events [28, 29]. Uncertainties of these measurements account for a total of 24 independent sources, but have only a small impact on the result.

All uncertainties in energy scales or resolutions are propagated to the missing transverse momentum. Additional uncertainties in the scale and resolution of the soft term are considered, for a total of three additional sources of systematic uncertainty.

### 7.2 Modeling uncertainties

The predicted $t\bar{t}H$ signal cross-section uncertainty is $^{+5.8\%}_{-9.2\%}$ (scale) $\pm 3.6\%$ (PDF), the first component representing the QCD scale uncertainty and the second the PDF+$\alpha_S$ uncertainty [15, 51–55]. These two components are treated as uncorrelated in the fit. The effect of QCD scale and PDF variations on the shape of the distributions considered in this analysis is found to be negligible. Uncertainties in the Higgs-boson branching fractions are also considered; these amount to 2.2% for the $b\bar{b}$ decay mode [15]. An additional uncertainty associated with the choice of parton shower and hadronization model is derived by comparing the nominal prediction from MG5_aMC@NLO+PYTHIA 8 to the one from MG5_aMC@NLO interfaced to Herwig++.

The systematic uncertainties affecting the modeling of the $t\bar{t}$+jets background are summarized in Table 1. An uncertainty of $\pm 6\%$ is assumed for the inclusive $t\bar{t}$ NNLO+NNLL production cross-section [62], including effects from varying the factorization and renormalization scales, the PDF, $\alpha_S$, and the top-quark mass. The $t\bar{t} + \geq 1b$, $t\bar{t} + \geq 1c$ and $t\bar{t} +$ light processes are affected by different types of uncertainties: $t\bar{t} +$ light has additional diagrams and profits from relatively precise measurements in data; $t\bar{t} + \geq 1b$ and $t\bar{t} + \geq 1c$ can have similar or different diagrams depending on the flavor scheme used for the PDF, and the mass differences between $c$- and $b$-quarks contribute to additional differences between these two processes. For these reasons, all uncertainties in $t\bar{t} +$ jets background modeling, except the uncertainty in the inclusive cross-section, are assigned independent nuisance parameters for the $t\bar{t} + \geq 1b$, $t\bar{t} + \geq 1c$ and $t\bar{t} +$ light processes. The normalizations of $t\bar{t} + \geq 1b$ and $t\bar{t} + \geq 1c$ are allowed to float freely in the fit. Systematic uncertainties in the shapes are extracted from the comparison between the nominal sample and various alternative samples. For all these uncertainties, alternative samples are reweighted in such a way that they have the same fractions of $t\bar{t} + \geq 1c$ and $t\bar{t} + \geq 1b$ as the nominal sample. In the case of the $t\bar{t} + \geq 1b$ background, separate uncertainties are applied to the relative normalization of the $t\bar{t} + \geq 1b$ subcategories as described later. Therefore, for all the alternative samples used to derive uncertainties that are not specifically associated with these fractions, the relative contributions of the $t\bar{t} + \geq 1b$ subcategories are scaled to match the predictions of Sherpa4F, in the same way as for the nominal sample. This scaling is not applied to the $t\bar{t} + b$ (MPI/FSR) subcategory, as explained in Section 4.

Uncertainties associated with the choice of $t\bar{t}$ inclusive NLO event generator as well as the choice of parton shower and hadronization model are derived by comparing the prediction from POWHEG+PYTHIA 8 with the Sherpa predictions (hence varying simultaneously the NLO event generator and the parton shower and hadronization model) and with the predictions from POWHEG interfaced with Herwig 7 [91] (varying just the parton shower and hadronization model). The former alternative sample was generated using Sherpa version 2.2.1 with the ME+PS@NLO setup, interfaced with OpenLoops, providing NLO accuracy for up to one additional parton and LO accuracy for up to four additional partons. The NNPDF3.0NNLO PDF set was used and both the renormalization and factorization scales were set to $\sqrt{0.5 \times (m_{T_{1,j}}^2 + m_{T_{1,i}}^2)}$. This sample is referred to as ‘Sherpa5F’ in the remainder of this article, which should not be confused
Table 1: Summary of the sources of systematic uncertainty for $t\bar{t}$ + jets modeling. The systematic uncertainties listed in the second section of the table are evaluated in such a way as to have no impact on the relative fractions of $t\bar{t} + 1b$, $t\bar{t} + 1c$ and $t\bar{t} + 1c$ light events, as well as on the relative fractions of the $t\bar{t} + b$, $t\bar{t} + b\bar{b}$, $t\bar{t} + B$ and $t\bar{t} + 3b$ subcategories, which are all kept at their nominal values. The systematic uncertainties listed in the third section of the table affect only the fractions of the various $t\bar{t} + 1b$ subcategories. The last column of the table indicates the $t\bar{t}$ category to which a systematic uncertainty is assigned. In the case where all three categories ($t\bar{t} + 1c$ and $t\bar{t} + 1b$) are involved (marked with ‘all’), the last column also specifies whether the uncertainty is considered as correlated or uncorrelated across them.

<table>
<thead>
<tr>
<th>Systematic source</th>
<th>Description</th>
<th>$t\bar{t}$ categories</th>
</tr>
</thead>
<tbody>
<tr>
<td>$t\bar{t}$ cross-section</td>
<td>Up or down by 6%</td>
<td>All, correlated</td>
</tr>
<tr>
<td>$k(t\bar{t} + 1c)$</td>
<td>Free-floating $t\bar{t} + 1c$ normalization</td>
<td>$t\bar{t} + 1c$</td>
</tr>
<tr>
<td>$k(t\bar{t} + 1b)$</td>
<td>Free-floating $t\bar{t} + 1b$ normalization</td>
<td>$t\bar{t} + 1b$</td>
</tr>
<tr>
<td>SHERPA4F vs. nominal</td>
<td>Related to the choice of NLO event generator</td>
<td>All, uncorrelated</td>
</tr>
<tr>
<td>PS &amp; hadronization</td>
<td>Powheg+Herwig 7 vs. Powheg+Pythia 8</td>
<td>All, uncorrelated</td>
</tr>
<tr>
<td>ISR / FSR</td>
<td>Variations of $\mu_R$, $\mu_F$, $h_{damp}$ and A14 Var3c parameters</td>
<td>All, uncorrelated</td>
</tr>
<tr>
<td>$t\bar{t} + 1c$ ME vs. inclusive</td>
<td>MG5_aMC@NLO+Herwig++: ME prediction (3F) vs. incl. (5F)</td>
<td>$t\bar{t} + 1c$</td>
</tr>
<tr>
<td>$t\bar{t} + 1b$ SHERPA4F vs. nominal</td>
<td>Comparison of $t\bar{t} + b\bar{b}$ NLO (4F) vs. Powheg+Pythia 8 (5F)</td>
<td>$t\bar{t} + 1b$</td>
</tr>
<tr>
<td>$t\bar{t} + 1b$ renorm. scale</td>
<td>Up or down by a factor of two</td>
<td>$t\bar{t} + 1b$</td>
</tr>
<tr>
<td>$t\bar{t} + 1b$ resumm. scale</td>
<td>Vary $\mu_Q$ from $H/2$ to $\mu_{CMMP}$</td>
<td>$t\bar{t} + 1b$</td>
</tr>
<tr>
<td>$t\bar{t} + 1b$ global scales</td>
<td>Set $\mu_Q$, $\mu_R$, and $\mu_F$ to $\mu_{CMMP}$</td>
<td>$t\bar{t} + 1b$</td>
</tr>
<tr>
<td>$t\bar{t} + 1b$ shower recoil scheme</td>
<td>Alternative model scheme</td>
<td>$t\bar{t} + 1b$</td>
</tr>
<tr>
<td>$t\bar{t} + 1b$ PDF (MSTW)</td>
<td>MSTW vs. CT10</td>
<td>$t\bar{t} + 1b$</td>
</tr>
<tr>
<td>$t\bar{t} + 1b$ PDF (NNPDF)</td>
<td>NNPDF vs. CT10</td>
<td>$t\bar{t} + 1b$</td>
</tr>
<tr>
<td>$t\bar{t} + 1b$ UE</td>
<td>Alternative set of tuned parameters for the underlying event</td>
<td>$t\bar{t} + 1b$</td>
</tr>
<tr>
<td>$t\bar{t} + 1b$ MPI</td>
<td>Up or down by 50%</td>
<td>$t\bar{t} + 1b$</td>
</tr>
<tr>
<td>$t\bar{t} + 1b$ normalization</td>
<td>Up or down by 50%</td>
<td>$t\bar{t} + 1b$</td>
</tr>
</tbody>
</table>

with the SHERPA4F sample defined in Section 4. The comparison with the latter alternative sample is considered as an independent source of uncertainty, related to the parton shower and hadronization model choice. This sample was generated with the same settings for Powheg as the nominal $t\bar{t}$ sample in terms of $h_{damp}$, PDF and renormalization and factorization scales, but it was interfaced with Herwig 7 version 7.0.1, with the H7-UE-MMHT set of tuned parameters for the underlying event. Additionally, the uncertainty in the modeling of initial- and final-state radiation (ISR / FSR) is assessed with two alternative Powheg+Pythia 8 samples [92]. One sample with the amount of radiation increased has the renormalization and factorization scales decreased by a factor of two, the $h_{damp}$ parameter doubled, and uses the Var3c upward variation of the A14 parameter set. A second sample with the amount of radiation decreased has the scales increased by a factor of two and uses the Var3c downward variation of the A14 set. The uncertainties described in this paragraph correspond to three independent sources for each of the $t\bar{t} +$ light, $t\bar{t} + 1c$ and $t\bar{t} + 1b$ components.

For the background from $t\bar{t} + 1c$, there is little guidance from theory or experiment to determine whether the nominal approach of using charm jets produced primarily in the parton shower is more or less accurate than a prediction with $t\bar{t} + c\bar{c}$ calculated at NLO in the matrix element. For this reason, an NLO prediction with $t\bar{t} + c\bar{c}$ in the matrix element, including massive c-quarks and therefore using the 3F scheme for the PDFs, is produced with MG5\_aMC@NLO interfaced to Herwig++, as described in Ref. [93]. The difference between this sample and an inclusive $t\bar{t}$ sample produced with the same event generator and a 5F scheme PDF set, in which the $t\bar{t} + 1c$ process originates through the parton shower only, is taken as an additional uncertainty in the $t\bar{t} + 1c$ prediction. This uncertainty is related to the choice between the
\( t\bar{t} + c\bar{c} \) ME calculation and the prediction from the inclusive \( t\bar{t} \) production with \( c \)-jets via parton shower and is applied as one additional independent source to the \( t\bar{t} + \geq 1c \) background.

For the \( t\bar{t} + \geq 1b \) process, the difference between the predictions from POWHEG+PYTHIA 8 and SHERPA4F is considered as one additional source of uncertainty. This uncertainty accounts for the difference between the description of the \( t\bar{t} + \geq 1b \) process by the NLO \( t\bar{t} \) inclusive MC sample with a 5F scheme and a description at NLO of \( t\bar{t} + b\bar{b} \) in the ME with a 4F scheme. This uncertainty is not applied to the \( t\bar{t} + b \) (MPI/FSR) subcategory since it is not included in the 4F calculation.

The uncertainties described above do not affect the relative fractions of the \( t\bar{t} + b \), \( t\bar{t} + b\bar{b} \), \( t\bar{t} + B \) and \( t\bar{t} + \geq 3b \) subcomponents as these fractions are fixed to the prediction of SHERPA4F. The uncertainties in these fractions in SHERPA4F are assessed separately and are divided into seven independent sources. Three of these sources are evaluated by varying the renormalization scale up and down by a factor of two, changing the functional form of the resummation scale to \( \mu_{\text{CMSPS}} \), and adopting a global scale choice, \( \mu_Q = \mu_R = \mu_F = \mu_{\text{CMSPS}} \). Additionally, two alternative PDF sets, MSTW2008NLO [94] and NNPDF2.3NLO, are considered, as well as an alternative shower recoil scheme and an alternative set of tuned parameters for the underlying event. These sources of uncertainty contribute to the uncertainty band shown in Figure 2 for the SHERPA4F prediction. Given the large difference between the 4F prediction and the various 5F predictions for the \( t\bar{t} + \geq 3b \) process, which is not covered by the uncertainties described above, this sub-process is given an extra 50% normalization uncertainty.

The relative fraction of the \( t\bar{t} + b \) (MPI/FSR) subcategory is not fixed in the alternative samples used to derive the systematic uncertainties related to the choice of NLO event generator, parton shower and hadronization model and to ISR/FSR. These sources already incorporate variations related to the fraction and shape of the \( t\bar{t} + b \) (MPI/FSR) subcategory. In addition, a 50% normalization uncertainty is assumed for the contribution from MPI, based on studies of different underlying event sets of tuned parameters.

In total, thirteen independent sources of modeling uncertainties are assigned to the \( t\bar{t} + \geq 1b \) component, four to the \( t\bar{t} + \geq 1c \) component and three to the \( t\bar{t} + \text{light} \) component in addition to the one source that corresponds to the inclusive \( t\bar{t} \) production cross-section uncertainty.

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

An uncertainty of \( \pm 5\% \) is considered for each of the three single-top production mode cross-sections [75–77]. For the \( Wt \) and \( t\bar{t} \)-channel production modes, uncertainties associated with the choice of parton shower and hadronization model and with initial- and final-state radiation are evaluated according to a set of alternative samples analogous to those used for the \( t\bar{t} \) process: the nominal prediction is compared with samples generated with POWHEG interfaced with HERWIG++ and with alternative POWHEG-BOX v1 + PYTHIA 6 samples with factorization and renormalization scale variations and appropriate variations of the Perugia 2012 set of tuned parameters. The uncertainty in the amount of interference between \( Wt \) and \( t\bar{t} \) production at NLO [72] is assessed by comparing the default ‘diagram removal’ scheme to the alternative ‘diagram subtraction’ scheme.
A 50% normalization uncertainty in the diboson background is assumed, which includes uncertainties in the inclusive cross-section and additional jet production [82]. The uncertainty of the $t\bar{t}V$ NLO cross-section prediction is 15% [95], split into PDF and scale uncertainties as for $t\bar{t}H$. An additional $t\bar{t}V$ modeling uncertainty, related to the choice of event generator, parton shower and hadronization model, is assessed by comparing the nominal sample with alternative ones generated with Sherpa. Uncertainties in $t\bar{t}V$ production are all treated as uncorrelated between $t\bar{t}Z$ and $t\bar{t}W$. A total 50% normalization uncertainty is considered for the $t\bar{t}t\bar{t}$ background. The small backgrounds from $t\bar{t}Z$, $t\bar{t}WW$, $tH$ and $WtH$ are each assigned two cross-section uncertainties, split into PDF and scale uncertainties, while $tWZ$ is assigned one cross-section uncertainty that accounts for both the scale and PDF effects.

Finally, a 50% uncertainty is assigned to the overall estimated yield of non-prompt lepton events in the single-lepton channel, taken as uncorrelated between electron-plus-jet and muon-plus-jet events, between boosted and resolved analysis categories, and between the resolved analysis categories with exactly five jets and those with six or more jets. In the dilepton channel, the non-prompt lepton background is assigned a 25% uncertainty, correlated across lepton flavors and all analysis categories.

### 8 Results

The distributions of the discriminants from each of the analysis categories are combined in a profile likelihood fit to test for the presence of a signal, while simultaneously determining the normalization and constraining the differential distributions of the most important background components. As described in Section 6, in the signal regions, the output of the classification BDT is used as the discriminant while only the total event yield is used in the control regions, with the exception of $CR_{t\bar{t}+\geq 1c}$ and $CR_{t\bar{t}+\geq 1c}$, where the $H_T^{had}$ distribution is used. No distinction is made in the fit between signal and control regions, other than a different choice of discriminant variables. The binning of the classification BDT is optimized to maximize the analysis sensitivity while keeping the total MC statistical uncertainty in each bin to a level adjusted to avoid biases due to fluctuations in the predicted number of events.

The likelihood function, $L(\mu, \theta)$, is constructed as a product of Poisson probability terms over all bins in each distribution. The Poisson probability depends on the predicted number of events in each bin, which in turn is a function of the signal-strength parameter $\mu = \sigma/\sigma_{SM}$ and $\theta$, where $\theta$ is the set of nuisance parameters that encode the effects of systematic uncertainties, and of the two free floating normalization factors $k(t\bar{t} + \geq 1b)$ and $k(t\bar{t} + \geq 1c)$ for the $t\bar{t} + \geq 1b$ and $t\bar{t} + \geq 1c$ backgrounds, respectively. The nuisance parameters are implemented in the likelihood function as Gaussian, log-normal or Poisson priors, with the exception of the normalization factors $k(t\bar{t} + \geq 1b)$ and $k(t\bar{t} + \geq 1c)$, for which no prior knowledge from theory or subsidiary measurements is assumed and hence which are only constrained by the profile likelihood fit to the data. The statistical uncertainty of the prediction, that incorporates the statistical uncertainty of the MC events and of the data-driven fake and non-prompt lepton estimate, is included in the likelihood in the form of additional nuisance parameters, one for each of the included bins. The test statistic $t_\mu$ is defined as the profile likelihood ratio: $t_\mu = -2 \ln(L(\mu, \hat{\theta}_\mu)/L(\hat{\mu}, \hat{\theta}))$, where $\hat{\mu}$ and $\hat{\theta}$ are the values of the parameters which maximize the likelihood function, and $\hat{\mu}$ are the values of the nuisance parameters which maximize the likelihood function for a given value of $\mu$. This test statistic is used to measure the probability that the observed data is compatible with the background-only hypothesis, and to perform statistical inferences about $\mu$, such as upper limits using the CLs method [96–98]. The uncertainty of the best-fit value of the signal strength, $\hat{\mu}$, is obtained varying $t_\mu$ by one unit.
Figure 7 shows the observed event yield compared to the prediction in each control and signal region, both before the fit to data (‘pre-fit’) and after the fit to data (‘post-fit’), performed in all the analysis categories in the two channels and with the signal-plus-background hypothesis. For the pre-fit prediction, the normalization factors for the $t\bar{t} + \geq 1b$ and $t\bar{t} + \geq 1c$ processes are set to 1, which corresponds to considering the prediction from *Powheg + Pythia* 8 for the fraction of each of these components relative to the total $t\bar{t}$ prediction. Figure 8 shows the $H_{T}^{\text{had}}$ distributions in the $t\bar{t} + \geq 1c$-enriched control regions of the single-lepton channel, while Figures 9, 10 and 11 show the distributions of the classification BDTs in the dilepton and single-lepton signal regions, both before and after the fit. All these distributions are reasonably well modeled pre-fit within the assigned uncertainties. The level of agreement is improved post-fit due to the nuisance parameters being adjusted by the fit. In particular, the best-fit values of $k(t\bar{t} + \geq 1b)$ and $k(t\bar{t} + \geq 1c)$ are $1.24 \pm 0.10$ and $1.63 \pm 0.23$, respectively. The uncertainties in these measured normalization factors do not include the theory uncertainty of the corresponding $t\bar{t} + \geq 1b$ and $t\bar{t} + \geq 1c$ cross-sections. The post-fit uncertainty is also significantly reduced, as a result of the nuisance-parameter constraints and the correlations generated by the fit.

In addition to the distributions that are given as input to the fit, all the distributions of the input variables to the classification BDTs in the signal regions are checked post-fit, and no significant deviations of the predictions from data are found. Figure 12 shows the data compared to the post-fit prediction for three of these distributions, namely the Higgs-boson candidate mass distributions in the most sensitive signal regions in the dilepton channel and the single-lepton resolved channels as well as in the single-lepton boosted signal region.

The best-fit $\mu$ value is:

$$\mu = 0.84 \pm 0.29 \text{ (stat.)} \pm 0.57 \text{ (syst.)} = 0.84^{+0.64}_{-0.61},$$

determined by the combined fit in all signal and control regions in the two channels. The expected uncertainty of the signal strength is identical to the measured one. An alternative combined fit is also performed in which the dilepton and single-lepton channels are assigned two independent signal strengths. The corresponding fitted values of $\mu$ are $-0.24^{+1.02}_{-1.05}$ in the dilepton channel and $0.95^{+0.65}_{-0.62}$ in the single-lepton channel. The probability that these two independent signal-strength parameters are compatible with each other is 19%. Figure 13 shows the comparison between the combined $\mu$ and the two independent signal-strength parameters from the combined fit, with their uncertainties split into the statistical and systematic components. The statistical uncertainty is obtained by redoing the fit to data after fixing all the nuisance parameters to their post-fit values, with the exception of the free normalization factors in the fit: $k(t\bar{t} + \geq 1c)$, $k(t\bar{t} + \geq 1b)$ and $\mu$. The total systematic uncertainty is obtained from the subtraction in quadrature of the statistical uncertainty from the total uncertainty. The statistical uncertainty contributes significantly less than the systematic component to the overall uncertainty of the measurement. When fitting the dilepton and single-lepton data separately, the observed signal strengths are $0.11^{+1.36}_{-1.41}$ and $0.67^{+0.71}_{-0.69}$, respectively. These two signal-strength values are both lower than the combined measured $\mu$ due to the large correlations in the systematic uncertainties of the background prediction between the two channels.

The contributions from the different sources of uncertainty in the combined fit to $\mu$ are reported in Table 2. The total systematic uncertainty is dominated by the uncertainties in the modeling of the $t\bar{t} + \geq 1b$ background, the second-largest source being the limited number of events in the simulated samples, followed by the uncertainties in the $b$-tagging efficiency, the jet energy scale and resolution, and the signal process modeling. The 20 nuisance parameters describing the independent sources of systematic uncertainty with the largest contribution to the total uncertainty of the measured signal strength are reported in Figure 14, ranked by decreasing contribution. For each of these nuisance parameters,
Figure 7: Comparison of predicted and observed event yields in each of the control and signal regions, in the dilepton channel (a) before and (b) after the fit to the data, and in the single-lepton channel (c) before and (d) after the fit to the data. The $t\bar{t}H$ signal is shown both as a filled red area stacked on the backgrounds and separately for visibility as a dashed red line, normalized to the SM cross-section before the fit and to the fitted uncertainty in the total prediction. The pre-fit plots do not include an uncertainty for the $t\bar{t} + \geq 1b$ or $t\bar{t} + \geq 1c$ normalization.

The best-fit value and the post-fit uncertainty are shown. The uncertainty coming from the comparison between the Sherpa5F and the nominal prediction for the $t\bar{t} + \geq 1b$ process, related to the choice of the NLO event generator for this background component, has the largest impact on the signal strength, followed by three uncertainties also related to the modeling of the $t\bar{t} + \geq 1b$ background. Systematic uncertainties related to the $t\bar{t}H$ signal modeling, the modeling of the $t\bar{t} + \geq 1c$ and $t\bar{t} + \text{light}$ backgrounds, and to experimental sources such as $b$-tagging, jet energy scale and resolution, also appear in Figure 14; however, their contributions are significantly smaller than the ones from the $t\bar{t} + \geq 1b$ background. The total uncertainty of the signal strength is reduced by 5% if the fit is performed excluding the systematic uncertainties not shown in this figure.
Figure 8: Comparison between data and prediction for the $H_{\text{had}}^{\text{med}}$ distributions in the single-lepton $t\bar{t} + \geq 1c$-enriched control regions (a, c) before, and (b, d) after the combined dilepton and single-lepton fit to the data. Despite its small contribution in these control regions, the $t\bar{t}H$ signal prediction is shown stacked at the top of the background prediction, normalized to the SM cross-section before the fit and to the fitted $\mu$ after the fit. The pre-fit plots do not include an uncertainty for the $t\bar{t} + \geq 1b$ or $t\bar{t} + \geq 1c$ normalization.
Figure 9: Comparison between data and prediction for the BDT discriminant in the dilepton signal regions (a, c, e) before, and (b, d, f) after the combined dilepton and single-lepton fit to the data. The $t\bar{t}H$ signal yield (solid red) is normalized to the SM cross-section before the fit and to the fitted $\mu$ after the fit. The dashed line shows the $t\bar{t}H$ signal distribution normalized to the total background prediction. The pre-fit plots do not include an uncertainty for the $t\bar{t} + \geq 1b$ or $t\bar{t} + \geq 1c$ normalization.
Figure 10: Comparison between data and prediction for the BDT discriminant in the single-lepton channel five-jet and boosted signal regions (a, c, e) before, and (b, d, f) after the combined dilepton and single-lepton fit to the data. The $t\bar{t}H$ signal yield (solid red) is normalized to the SM cross-section before the fit and to the fitted $\mu$ after the fit. The dashed line shows the $t\bar{t}H$ signal distribution normalized to the total background prediction. The pre-fit plots do not include an uncertainty for the $t\bar{t} + \geq 1b$ or $t\bar{t} + \geq 1c$ normalization.
Figure 11: Comparison between data and prediction for the BDT discriminant in the single-lepton channel six-jet signal regions (a, c, e) before, and (b, d, f) after the combined dilepton and single-lepton fit to the data. The $t\bar{t}H$ signal yield (solid red) is normalized to the SM cross-section before the fit and to the fitted signal regions (a, c, e) before, and (b, d, f) after the combined dilepton and single-lepton fit to the data. The dashed line shows the $t\bar{t}H$ signal distribution normalized to the total background prediction. The pre-fit plots do not include an uncertainty for the $t\bar{t} + \geq 1b$ or $t\bar{t} + \geq 1c$ normalization.
Figure 12: Comparison between data and prediction for the Higgs-boson candidate mass from the reconstruction BDT trained without variables involving the Higgs-boson candidate (a) in the dilepton SR$_{\geq 4j}$ and (b) in the single-lepton SR$_{\geq 6j}$, and (c) for the boosted Higgs-boson candidate in SR$^{\text{boosted}}$, after the combined dilepton and single-lepton fit to the data. The $t\bar{t}H$ signal yield (solid red) is normalized to the fitted $\mu$ after the fit. The dashed red line shows the $t\bar{t}H$ signal distribution normalized to the total background yield. The dashed black line shows the pre-fit total background prediction.
The theoretical predictions for the $\bar{t}t + \geq 1b$ process suffer from large uncertainties as reflected in the size of the difference between alternative simulated samples used to model this background. The corresponding systematic uncertainties are therefore large and are a crucial limiting factor for this search. The choice of nuisance parameters for systematic uncertainties related to the $\bar{t}t + \geq 1b$ background is studied carefully to ensure sufficient flexibility in the fit to correct for possible mis-modeling of this background and avoid any bias in the measured signal strength. In total, 13 independent nuisance parameters are assigned to $\bar{t}t + \geq 1b$ background modeling uncertainties. The capability of the fit to correct for mis-modeling effects, beyond the ones present in the distributions used in the fit, is confirmed by comparing the predictions of all input variables of the classification BDT obtained post-fit to data. As mentioned before, no significant deviations of the predictions from data are found and the agreement is improved post-fit. Alternative approaches to model the $\bar{t}t + \geq 1b$ background, to define the associated uncertainties and to correlate them are also tested, and the corresponding results are found to be compatible with the nominal result.

To further validate the robustness of the fit, a pseudo-data set is built from simulated events by replacing the nominal $\bar{t}t$ background by an alternative sample that is not used in the definition of any uncertainty. This alternative sample was generated with POWHEG+PYTHIA 6 and is similar to the sample used for the $\bar{t}tH(H \rightarrow b\bar{b})$ analysis [16] in Run 1 of the LHC. The fit to this pseudo-data sample did not reveal any bias in the signal extraction.

Figure 14 shows that some nuisance parameters are shifted in the fit from their nominal values. To understand the origin of these shifts, the corresponding nuisance parameters are switched to be uncorrelated between analysis categories and samples and the fit is repeated. These shifts are found to correct mainly the predictions of the $\bar{t}t$ background to the observed data in various regions. Similar shifts are observed when a background-only fit is performed after removing the bins with the most significant signal contributions. Moreover, the variations induced in the signal strength by these shifts are quantified by fixing the corresponding nuisance parameters to their pre-fit values, repeating the fit, and comparing the obtained $\mu$-value with the nominal one. These variations were found to be smaller than the uncertainty in the signal strength. Independent signal-strength values extracted from different sets of analysis categories and from the two channels are also found to be compatible.

Figure 14 also shows that the uncertainties corresponding to some nuisance parameters are reduced by the fit. When performing the profile likelihood fit, nuisance parameters associated with uncertainties affecting the discriminant distributions by variations that would result in large deviations from to data are significantly constrained. The capability of the fit to constrain systematic uncertainties is validated on the pseudo-data sample described above, and on the pseudo-data sample produced from the nominal predictions, the Asimov dataset [96].

An excess of events over the expected SM background is found with an observed (expected) significance of 1.4 (1.6) standard deviations. A signal strength larger than 2.0 is excluded at the 95% confidence level, as shown in Figure 15. The expected significance and exclusion limits are calculated using the background estimate after the fit to the data. Figure 16 shows the event yield in data compared to the post-fit prediction for all events entering the analysis selection, grouped and ordered by the signal-to-background ratio of the corresponding final-discriminant bins. The predictions are shown for both the fit with the background-only hypothesis and with the signal-plus-background hypothesis, where the signal is scaled to either the measured $\mu$ or the value of the upper limit on $\mu$. 

30
Figure 13: Summary of the signal-strength measurements in the individual channels and for the combination. All the numbers are obtained from a simultaneous fit in the two channels, but the measurements in the two channels separately are obtained keeping the signal strengths uncorrelated, while all the nuisance parameters are kept correlated across channels.
Figure 14: Ranking of the nuisance parameters included in the fit according to their impact on the measured signal strength $\mu$. Only the 20 most highly ranked parameters are shown. Nuisance parameters corresponding to MC statistical uncertainties are not included here. The empty blue rectangles correspond to the pre-fit impact on $\mu$ and the filled blue ones to the post-fit impact on $\mu$, both referring to the upper scale. The impact of each nuisance parameter, $\Delta \mu$, is computed by comparing the nominal best-fit value of $\mu$ with the result of the fit when fixing the considered nuisance parameter to its best-fit value, $\hat{\theta}$, shifted by its pre-fit (post-fit) uncertainties $\pm \Delta \theta$ ($\pm \Delta \hat{\theta}$). The black points show the pulls of the nuisance parameters relative to their nominal values, $\theta_0$. These pulls and their relative post-fit errors, $\Delta \hat{\theta} / \Delta \theta$, refer to the lower scale. The parameter $k(tt+1b)$ refers to the floating normalization of the $tt+1b$ background, for which the pre-fit impact on $\mu$ is not defined, and for which both $\theta_0$ and $\Delta \theta$ are set to 1. For experimental uncertainties that are decomposed into several independent sources, NP I and NP II correspond to the first and second nuisance parameters, ordered by their impact on $\mu$, respectively.
Table 2: Breakdown of the contributions to the uncertainties in $\mu$. The line ‘background-model stat. unc.’ refers to the statistical uncertainties in the MC events and in the data-driven determination of the non-prompt and fake lepton background component in the single-lepton channel. The contribution of the different sources of uncertainty is evaluated after the fit described in Section 8. The total statistical uncertainty is evaluated, as described in the text, by fixing all the nuisance parameters in the fit except for the free-floating normalization factors for the $t \bar{t} + \geq 1b$ and $t \bar{t} + \geq 1c$ background components. The contribution from the uncertainty in the normalization of both $t \bar{t} + \geq 1b$ and $t \bar{t} + \geq 1c$ is then included in the quoted total statistical uncertainty rather than in the systematic uncertainty component. The statistical uncertainty evaluated after also fixing the normalization of $t \bar{t} + \geq 1b$ and $t \bar{t} + \geq 1c$ is then indicated as ‘intrinsic statistical uncertainty’. The other quoted numbers are obtained by repeating the fit after having fixed a certain set of nuisance parameters corresponding to a group of systematic uncertainty sources, and subtracting in quadrature the resulting total uncertainty of $\mu$ from the uncertainty from the full fit. The same procedure is followed for quoting the individual effects of the $t \bar{t} + \geq 1b$ and the $t \bar{t} + \geq 1c$ normalization. The total uncertainty is different from the sum in quadrature of the different components due to correlations between nuisance parameters built by the fit.

<table>
<thead>
<tr>
<th>Uncertainty source</th>
<th>$\Delta \mu$</th>
</tr>
</thead>
<tbody>
<tr>
<td>$t \bar{t} + \geq 1b$ modeling</td>
<td>+0.46 −0.46</td>
</tr>
<tr>
<td>Background-model stat. unc.</td>
<td>+0.29 −0.31</td>
</tr>
<tr>
<td>$b$-tagging efficiency and mis-tag rates</td>
<td>+0.16 −0.16</td>
</tr>
<tr>
<td>Jet energy scale and resolution</td>
<td>+0.14 −0.14</td>
</tr>
<tr>
<td>$t \bar{t}H$ modeling</td>
<td>+0.22 −0.05</td>
</tr>
<tr>
<td>$t \bar{t} + \geq 1c$ modeling</td>
<td>+0.09 −0.11</td>
</tr>
<tr>
<td>JVT, pileup modeling</td>
<td>+0.03 −0.05</td>
</tr>
<tr>
<td>Other background modeling</td>
<td>+0.08 −0.08</td>
</tr>
<tr>
<td>$t \bar{t} +$ light modeling</td>
<td>+0.06 −0.03</td>
</tr>
<tr>
<td>Luminosity</td>
<td>+0.03 −0.02</td>
</tr>
<tr>
<td>Light lepton ($e, \mu$) id., isolation, trigger</td>
<td>+0.03 −0.04</td>
</tr>
<tr>
<td>Total systematic uncertainty</td>
<td>+0.57 −0.54</td>
</tr>
<tr>
<td>$t \bar{t} + \geq 1b$ normalization</td>
<td>+0.09 −0.10</td>
</tr>
<tr>
<td>$t \bar{t} + \geq 1c$ normalization</td>
<td>+0.02 −0.03</td>
</tr>
<tr>
<td>Intrinsic statistical uncertainty</td>
<td>+0.21 −0.20</td>
</tr>
<tr>
<td>Total statistical uncertainty</td>
<td>+0.29 −0.29</td>
</tr>
<tr>
<td>Total uncertainty</td>
<td>+0.64 −0.61</td>
</tr>
</tbody>
</table>
Figure 15: Summary of the 95% confidence level (CL) upper limits on $\sigma(t\bar{t}H)$ relative to the SM prediction in the individual channels and for the combination. The observed limits are shown, together with the expected limits both in the background-only hypothesis (dotted black lines) and in the SM hypothesis (dotted red lines). In the case of the expected limits in the background-only hypothesis, one- and two-standard-deviation uncertainty bands are also shown. The limits for the two individual channels are derived consistently with Figure 13, both extracted from the profile likelihood including the data in both channels, but with independent signal strengths in the two channels.
Figure 16: Post-fit yields of signal (S) and total background (B) as a function of \( \log(S/B) \), compared to data. Final-discriminant bins in all dilepton and single-lepton analysis categories are combined into bins of \( \log(S/B) \), with the signal normalized to the SM prediction used for the computation of \( \log(S/B) \). The signal is then shown normalized to the best-fit value and to the value excluded at the 95% CL, in both cases summed to the background prediction from the fit. The lower frame reports for each bin the pull (residual divided by its uncertainty) of the data relative to the background prediction from the fit. These data pulls are compared to the pulls of the signal-plus-background prediction from the fit, assuming a signal strength equal to the best-fit value (solid red line) and equal to the exclusion limit (dashed orange line). The background and its pull are also shown after the fit to data assuming zero signal contribution (dashed black line, obscured by solid line in the upper frame). The first bin includes the underflow.
9 Conclusion

A search for the associated production of the Standard Model Higgs boson with a pair of top quarks is presented, based on 36.1 fb$^{-1}$ of $pp$ collision data at $\sqrt{s} = 13$ TeV, collected with the ATLAS detector at the Large Hadron Collider in 2015 and 2016. The search focuses on decays of the Higgs boson to $b\bar{b}$ and decays of the top quark pair to a final state containing one or two leptons. Multivariate techniques are used to discriminate between signal and background events, the latter being dominated by $t\bar{t} +$ jets production. The observed data are consistent with both the background-only hypothesis and with the Standard Model $t\bar{t}H$ prediction. A 1.4 $\sigma$ excess above the expected background is observed, while an excess of 1.6 $\sigma$ is expected in the presence of a Standard Model Higgs boson. The signal strength is measured to be $0.84^{+0.64}_{-0.61}$, consistent with the expectation from the Standard Model. A value higher than 2.0 is excluded at the 95% confidence level, compared to an expected exclusion limit of 1.2 in the absence of signal. The measurement uncertainty is presently dominated by systematic uncertainties, and more specifically by the uncertainty in the theoretical knowledge of the $t\bar{t} + \geq 1b$ production process. An improved modelling of this background will be important for future efforts to observe the $t\bar{t}H(H \rightarrow b\bar{b})$ process.
Appendix

A Yield tables

The predicted event yields in each of the analysis categories, broken down into the different signal and background contributions and compared to the observed yields in data, are reported in Tables 3, 4 and 5. Both the pre-fit and post-fit predictions are shown, where post-fit refers to the combined fit to the dilepton and single-lepton channels with the signal-plus-background hypothesis, reported in Section 8. The total uncertainties of each of the signal and background components, and of the total prediction are also reported.

Table 3: Event yields in the dilepton channel (top) control regions and (bottom) signal regions. Post-fit yields are after the combined fit in dilepton and single-lepton channels to data. The uncertainties are the sum in quadrature of statistical and systematic uncertainties in the yields. In the post-fit case, these uncertainties are computed taking into account correlations among nuisance parameters and among the normalization of different processes. The uncertainty in the $t\bar{t} + \geq 1b$ and $t\bar{t} + \geq 1c$ normalization is not defined pre-fit and therefore only included in the post-fit uncertainties. For the $t\bar{t}H$ signal, the pre-fit yield values correspond to the theoretical prediction and corresponding uncertainties, while the post-fit yield and uncertainties correspond to those in the signal-strength measurement.

<table>
<thead>
<tr>
<th>Sample</th>
<th>CR$^{3j}_{t\bar{t}+\text{light}}$</th>
<th>CR$^{3j}_{t\bar{t}+\geq 1b}$</th>
<th>CR$^{2j}_{t\bar{t}+\text{light}}$</th>
<th>CR$^{2j}_{t\bar{t}+\geq 1c}$</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Pre-fit</td>
<td>Post-fit</td>
<td>Pre-fit</td>
<td>Post-fit</td>
</tr>
<tr>
<td>$t\bar{t}H$</td>
<td>32.2 ± 3.8</td>
<td>27 ± 20</td>
<td>8.7 ± 1.1</td>
<td>7.3 ± 5.4</td>
</tr>
<tr>
<td>$t\bar{t} + \text{light}$</td>
<td>63.1 ± 100</td>
<td>59.1 ± 100</td>
<td>360 ± 160</td>
<td>536 ± 89</td>
</tr>
<tr>
<td>$t\bar{t} + \geq 1c$</td>
<td>4800 ± 2100</td>
<td>7700 ± 1100</td>
<td>710 ± 140</td>
<td>848 ± 75</td>
</tr>
<tr>
<td>$t\bar{t} + \geq 1b$</td>
<td>2130 ± 230</td>
<td>2620 ± 240</td>
<td>710 ± 140</td>
<td>848 ± 75</td>
</tr>
<tr>
<td>$t\bar{t} + V$</td>
<td>113 ± 31</td>
<td>112 ± 29</td>
<td>7 ± 27</td>
<td>7 ± 30</td>
</tr>
<tr>
<td>Non-$t\bar{t}$</td>
<td>6300 ± 1500</td>
<td>6500 ± 1200</td>
<td>110 ± 29</td>
<td>112 ± 23</td>
</tr>
<tr>
<td>Total</td>
<td>76400 ± 6500</td>
<td>76010 ± 390</td>
<td>1500 ± 260</td>
<td>1765 ± 60</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Sample</th>
<th>SR$^{3j}_3$</th>
<th>SR$^{2j}_2$</th>
<th>SR$^{2j}_4$</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Pre-fit</td>
<td>Post-fit</td>
<td>Pre-fit</td>
</tr>
<tr>
<td>$t\bar{t}H$</td>
<td>21.9 ± 2.5</td>
<td>18 ± 13</td>
<td>29.1 ± 4.2</td>
</tr>
<tr>
<td>$t\bar{t} + \text{light}$</td>
<td>83 ± 41</td>
<td>95 ± 30</td>
<td>250 ± 110</td>
</tr>
<tr>
<td>$t\bar{t} + \geq 1c$</td>
<td>235 ± 61</td>
<td>313 ± 53</td>
<td>340 ± 210</td>
</tr>
<tr>
<td>$t\bar{t} + \geq 1b$</td>
<td>819 ± 85</td>
<td>917 ± 71</td>
<td>590 ± 96</td>
</tr>
<tr>
<td>$t\bar{t} + V$</td>
<td>15 ± 35</td>
<td>15 ± 34</td>
<td>22 ± 38</td>
</tr>
<tr>
<td>Non-$t\bar{t}$</td>
<td>75 ± 17</td>
<td>78 ± 16</td>
<td>115 ± 36</td>
</tr>
<tr>
<td>Total</td>
<td>1250 ± 140</td>
<td>1436 ± 55</td>
<td>1350 ± 320</td>
</tr>
</tbody>
</table>

Data: 1467 | 1444 | 319
Table 4: Event yields in the single-lepton channel five-jet (top) control regions and (bottom) signal regions, including the boosted signal region. Post-fit yields are after the combined fit in dilepton and single-lepton channels to data. The uncertainties are the sum in quadrature of statistical and systematic uncertainties in the yields. In the post-fit case, these uncertainties are computed taking into account correlations among nuisance parameters and among the normalization of different processes. The uncertainty in the $t\bar{t} + \geq 1b$ and $t\bar{t} + \geq 1c$ normalization is not defined pre-fit and therefore only included in the post-fit uncertainties. For the $t\bar{t}H$ signal, the pre-fit yield values correspond to the theoretical prediction and corresponding uncertainties, while the post-fit yield and uncertainties correspond to those in the signal-strength measurement.

<table>
<thead>
<tr>
<th>Sample</th>
<th>$CR_{t\bar{t} + \text{light}}^{5j}$</th>
<th>$CR_{t\bar{t} + \geq 1c}^{5j}$</th>
<th>$CR_{t\bar{t} + b}^{5j}$</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Pre-fit</td>
<td>Post-fit</td>
<td></td>
</tr>
<tr>
<td>$t\bar{t}H$</td>
<td>$224 \pm 22$</td>
<td>$190 \pm 140$</td>
<td>$15 \pm 12$</td>
</tr>
<tr>
<td>$t\bar{t} +\text{light}$</td>
<td>$197,000 \pm 26,000$</td>
<td>$179,900 \pm 49,000$</td>
<td>$2580 \pm 720$</td>
</tr>
<tr>
<td>$t\bar{t} + \geq 1c$</td>
<td>$27,500 \pm 4,300$</td>
<td>$44,100 \pm 5,500$</td>
<td>$1280 \pm 500$</td>
</tr>
<tr>
<td>$t\bar{t} + \geq 1b$</td>
<td>$11,300 \pm 1,100$</td>
<td>$13,500 \pm 1,300$</td>
<td>$790 \pm 130$</td>
</tr>
<tr>
<td>$t\bar{t} + V$</td>
<td>$589 \pm 55$</td>
<td>$584 \pm 54$</td>
<td>$23.2 \pm 4.1$</td>
</tr>
<tr>
<td>Non-$t\bar{t}$</td>
<td>$21,300 \pm 4,100$</td>
<td>$20,900 \pm 3,200$</td>
<td>$520 \pm 180$</td>
</tr>
<tr>
<td>Total</td>
<td>$258,000 \pm 29,000$</td>
<td>$259,320 \pm 910$</td>
<td>$3,200 \pm 1100$</td>
</tr>
<tr>
<td>Data</td>
<td>$259,320$</td>
<td>$5,465$</td>
<td>$11,095$</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Sample</th>
<th>$SR_{t\bar{t} + \text{light}}^{5j}$</th>
<th>$SR_{t\bar{t} + \geq 1c}^{5j}$</th>
<th>$SR_{t\bar{t} + b}^{5j}$</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Pre-fit</td>
<td>Post-fit</td>
<td></td>
</tr>
<tr>
<td>$t\bar{t}H$</td>
<td>$40.1 \pm 3.1$</td>
<td>$34 \pm 25$</td>
<td>$15.9 \pm 2.1$</td>
</tr>
<tr>
<td>$t\bar{t} +\text{light}$</td>
<td>$500 \pm 210$</td>
<td>$393 \pm 67$</td>
<td>$15 \pm 33$</td>
</tr>
<tr>
<td>$t\bar{t} + \geq 1c$</td>
<td>$436 \pm 92$</td>
<td>$610 \pm 100$</td>
<td>$30 \pm 17$</td>
</tr>
<tr>
<td>$t\bar{t} + \geq 1b$</td>
<td>$1230 \pm 200$</td>
<td>$1450 \pm 110$</td>
<td>$273 \pm 53$</td>
</tr>
<tr>
<td>$t\bar{t} + V$</td>
<td>$19.9 \pm 2.9$</td>
<td>$19.7 \pm 2.4$</td>
<td>$6.4 \pm 1.3$</td>
</tr>
<tr>
<td>Non-$t\bar{t}$</td>
<td>$269 \pm 64$</td>
<td>$220 \pm 52$</td>
<td>$54 \pm 11$</td>
</tr>
<tr>
<td>Total</td>
<td>$2440 \pm 390$</td>
<td>$2724 \pm 70$</td>
<td>$371 \pm 68$</td>
</tr>
<tr>
<td>Data</td>
<td>$2798$</td>
<td>$426$</td>
<td>$710 \pm 200$</td>
</tr>
</tbody>
</table>

Data

38
<table>
<thead>
<tr>
<th>Sample</th>
<th>CR ≥6j</th>
<th></th>
<th>CR ≥6j</th>
<th></th>
<th>CR ≥6j</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>Pre-fit</td>
<td>Post-fit</td>
<td>Pre-fit</td>
<td>Post-fit</td>
<td>Pre-fit</td>
</tr>
<tr>
<td>ttH</td>
<td>450 ± 48</td>
<td>370 ± 280</td>
<td>102 ± 13</td>
<td>87 ± 64</td>
<td>100 ± 12</td>
<td>83 ± 61</td>
</tr>
<tr>
<td>tt + light</td>
<td>125 000 ± 34 000</td>
<td>108 200 ± 4300</td>
<td>4300 ± 2000</td>
<td>3350 ± 430</td>
<td>2220 ± 520</td>
<td>1820 ± 170</td>
</tr>
<tr>
<td>tt + ≥1c</td>
<td>28 400 ± 7200</td>
<td>45 700 ± 5100</td>
<td>3600 ± 1300</td>
<td>5300 ± 680</td>
<td>1460 ± 330</td>
<td>2080 ± 300</td>
</tr>
<tr>
<td>tt + ≥1b</td>
<td>13 100 ± 1800</td>
<td>14 600 ± 1400</td>
<td>2660 ± 540</td>
<td>2950 ± 280</td>
<td>3670 ± 500</td>
<td>4080 ± 320</td>
</tr>
<tr>
<td>tt + V</td>
<td>1010 ± 120</td>
<td>996 ± 91</td>
<td>118 ± 21</td>
<td>118 ± 14</td>
<td>70.5 ± 8.5</td>
<td>67.9 ± 7.2</td>
</tr>
<tr>
<td>Non-tt</td>
<td>12 600 ± 3000</td>
<td>11 800 ± 2000</td>
<td>1060 ± 340</td>
<td>1000 ± 210</td>
<td>710 ± 160</td>
<td>600 ± 110</td>
</tr>
<tr>
<td>Total</td>
<td>181 000 ± 39 000</td>
<td>181 690 ± 860</td>
<td>11 800 ± 3200</td>
<td>12 810 ± 260</td>
<td>8200 ± 1100</td>
<td>8730 ± 230</td>
</tr>
<tr>
<td>Data</td>
<td>181 706</td>
<td>12 778</td>
<td></td>
<td></td>
<td></td>
<td>8576</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Sample</th>
<th>SR ≥6j</th>
<th></th>
<th>SR ≥6j</th>
<th></th>
<th>SR ≥6j</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>Pre-fit</td>
<td>Post-fit</td>
<td>Pre-fit</td>
<td>Post-fit</td>
<td>Pre-fit</td>
</tr>
<tr>
<td>ttH</td>
<td>85 ± 10</td>
<td>71 ± 52</td>
<td>81 ± 10</td>
<td>68 ± 50</td>
<td>62 ± 11</td>
<td>51 ± 38</td>
</tr>
<tr>
<td>tt + light</td>
<td>750 ± 370</td>
<td>586 ± 98</td>
<td>210 ± 210</td>
<td>96 ± 33</td>
<td>14 ± 10</td>
<td>12.1 ± 5.8</td>
</tr>
<tr>
<td>tt + ≥1c</td>
<td>880 ± 350</td>
<td>1330 ± 190</td>
<td>350 ± 100</td>
<td>473 ± 99</td>
<td>53 ± 33</td>
<td>44 ± 20</td>
</tr>
<tr>
<td>tt + ≥1b</td>
<td>2100 ± 420</td>
<td>2290 ± 170</td>
<td>1750 ± 370</td>
<td>1850 ± 130</td>
<td>1010 ± 240</td>
<td>1032 ± 59</td>
</tr>
<tr>
<td>tt + V</td>
<td>51.2 ± 7.4</td>
<td>50.8 ± 5.9</td>
<td>40.8 ± 5.7</td>
<td>40.3 ± 4.8</td>
<td>25.8 ± 3.7</td>
<td>25.3 ± 3.2</td>
</tr>
<tr>
<td>Non-tt</td>
<td>303 ± 82</td>
<td>267 ± 63</td>
<td>155 ± 52</td>
<td>134 ± 46</td>
<td>75 ± 20</td>
<td>58 ± 17</td>
</tr>
<tr>
<td>Total</td>
<td>4140 ± 850</td>
<td>4590 ± 110</td>
<td>2550 ± 510</td>
<td>2657 ± 82</td>
<td>1220 ± 250</td>
<td>1223 ± 42</td>
</tr>
<tr>
<td>Data</td>
<td>4698</td>
<td>2641</td>
<td></td>
<td></td>
<td></td>
<td>1222</td>
</tr>
</tbody>
</table>
B Input variables to the classification BDTs

In this appendix, the full list of variables used as inputs to the classification BDT, described in Section 6, in each of the signal regions is reported. Variables are listed separately in Table 6 for the dilepton channel, in Table 7 for the resolved single-lepton channel and in Table 8 for the boosted category. Variables are grouped according to the type of information that is exploited. The variables from the reconstruction BDT exploit the chosen jet–parton assignments described in Section 6.2. The $b$-tagging discriminant assigned to each jet is defined in Section 3. The most powerful variables in the classification BDT are the reconstruction BDT output, the LHD (Section 6.3) and the MEM$_{D1}$ (Section 6.4). The large-$R$ jets used to build the Higgs-boson and top-quark candidates in the boosted category are defined in Section 3.

Some kinematic and topological variables are built considering only $b$-tagged-jets in the event. The $b$-tagging requirements for these jets are optimized separately for each variable in each region to improve the classification BDT performance. In the resolved single-lepton channel, $b$-tagged-jets are defined as the four jets with the largest value of the $b$-tagging discriminant. If two jets have the same $b$-tagging discriminant value, they are ordered by decreasing jet $p_T$ value. In the dilepton channel, the $b$-tagging requirements depend on the signal region: in SR$_{4j}^{>}$ the tight working point is used, in SR$_{3j}^{≥4j}$ the very tight working point is used and in SR$_{2j}^{≥4j}$ the loose working point is used with the exception of $N_{bb}^{Higgs}$, which uses the medium working point, and Aplanarity$_{b$-jet} which uses the tight working point. The loose working point is used in the boosted signal region.
Table 6: Variables used in the classification BDTs in the dilepton signal regions. For variables from the reconstruction BDT, those with a * are from the BDT using Higgs-boson information, those with no * are from the BDT without Higgs-boson information while for those with a ** both versions are used. These two versions of the reconstruction BDT are described in Section 6.2.

<table>
<thead>
<tr>
<th>Variable</th>
<th>Definition</th>
<th>SR$_{1}^{≥4j}$</th>
<th>SR$_{2}^{≥4j}$</th>
<th>SR$_{3}^{≥4j}$</th>
</tr>
</thead>
<tbody>
<tr>
<td>$m_{bb}^{\text{min}}$</td>
<td>Minimum invariant mass of a $b$-tagged jet pair</td>
<td>✓</td>
<td>✓</td>
<td>-</td>
</tr>
<tr>
<td>$m_{bb}^{\text{max}}$</td>
<td>Maximum invariant mass of a $b$-tagged jet pair</td>
<td>-</td>
<td>-</td>
<td>✓</td>
</tr>
<tr>
<td>$m_{bb}^{\text{min}} \Delta R$</td>
<td>Invariant mass of the $b$-tagged jet pair with minimum $\Delta R$</td>
<td>✓</td>
<td>-</td>
<td>✓</td>
</tr>
<tr>
<td>$m_{jj}^{\text{max}} p_{T}$</td>
<td>Invariant mass of the jet pair with maximum $p_{T}$</td>
<td>✓</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>$m_{bb}^{\text{max}} p_{T}$</td>
<td>Invariant mass of the $b$-tagged jet pair with maximum $p_{T}$</td>
<td>✓</td>
<td>-</td>
<td>✓</td>
</tr>
<tr>
<td>$\Delta \eta_{bb}^{\text{avg}}$</td>
<td>Average $\Delta \eta$ for all $b$-tagged jet pairs</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
</tr>
<tr>
<td>$\Delta \eta_{bb}^{\text{max}}$</td>
<td>Maximum $\Delta \eta$ between a jet and a lepton</td>
<td>-</td>
<td>✓</td>
<td>-</td>
</tr>
<tr>
<td>$\Delta p_{T}^{\text{max,} bb}$</td>
<td>$\Delta R$ between the $b$-tagged jet pair with maximum $p_{T}$</td>
<td>-</td>
<td>✓</td>
<td>✓</td>
</tr>
<tr>
<td>$N_{\text{Higgs, } 30}^{bb}$</td>
<td>Number of $b$-tagged jet pairs with invariant mass within 30 GeV of the Higgs-boson mass</td>
<td>✓</td>
<td>✓</td>
<td>-</td>
</tr>
<tr>
<td>$n_{p_{T}&gt;40}$</td>
<td>Number of jets with $p_{T} &gt; 40$ GeV</td>
<td>-</td>
<td>✓</td>
<td>✓</td>
</tr>
<tr>
<td>Aplanarity$_{b\text{-jet}}$</td>
<td>$1.5 \lambda_2$, where $\lambda_2$ is the second eigenvalue of the momentum tensor [99] built with all $b$-tagged jets</td>
<td>-</td>
<td>✓</td>
<td>-</td>
</tr>
<tr>
<td>$H_{T}^{\text{all}}$</td>
<td>Scalar sum of $p_{T}$ of all jets and leptons</td>
<td>-</td>
<td>-</td>
<td>✓</td>
</tr>
</tbody>
</table>

Variables from reconstruction BDT

<table>
<thead>
<tr>
<th>Variable</th>
<th>Definition</th>
<th>SR$_{1}^{≥4j}$</th>
<th>SR$_{2}^{≥4j}$</th>
<th>SR$_{3}^{≥4j}$</th>
</tr>
</thead>
<tbody>
<tr>
<td>BDT output</td>
<td>Output of the reconstruction BDT</td>
<td>✓**</td>
<td>✓**</td>
<td>✓</td>
</tr>
<tr>
<td>$m_{bb}^{\text{Higgs}}$</td>
<td>Higgs candidate mass</td>
<td>✓</td>
<td>-</td>
<td>✓</td>
</tr>
<tr>
<td>$\Delta R_{H,t\bar{t}}$</td>
<td>$\Delta R$ between Higgs candidate and $t\bar{t}$ candidate system</td>
<td>✓</td>
<td>-</td>
<td>✓</td>
</tr>
<tr>
<td>$\Delta \eta_{H,\ell}^{\text{min}}$</td>
<td>Minimum $\Delta \eta$ between Higgs candidate and lepton</td>
<td>✓</td>
<td>✓</td>
<td>-</td>
</tr>
<tr>
<td>$\Delta \eta_{H,b}^{\text{min}}$</td>
<td>Minimum $\Delta \eta$ between Higgs candidate and $b$-jet from top</td>
<td>✓</td>
<td>✓</td>
<td>-</td>
</tr>
<tr>
<td>$\Delta \eta_{H,b}^{\text{max}}$</td>
<td>Maximum $\Delta \eta$ between Higgs candidate and $b$-jet from top</td>
<td>✓</td>
<td>✓</td>
<td>-</td>
</tr>
<tr>
<td>$\Delta R_{bb}^{\text{Higgs}}$</td>
<td>$\Delta R$ between the two jets matched to the Higgs candidate</td>
<td>-</td>
<td>✓</td>
<td>-</td>
</tr>
</tbody>
</table>

Variables from $b$-tagging

<table>
<thead>
<tr>
<th>Variable</th>
<th>Definition</th>
<th>SR$_{1}^{≥4j}$</th>
<th>SR$_{2}^{≥4j}$</th>
<th>SR$_{3}^{≥4j}$</th>
</tr>
</thead>
<tbody>
<tr>
<td>$w_{Higgs}^{b\text{-tag}}$</td>
<td>Sum of $b$-tagging discriminants of jets from best Higgs candidate from the reconstruction BDT</td>
<td>-</td>
<td>✓</td>
<td>-</td>
</tr>
</tbody>
</table>
Table 7: Input variables to the classification BDTs in the single-lepton signal regions. For variables from the reconstruction BDT, those with an asterisk are from the BDT using Higgs-boson information, those with no asterisk are from the BDT without Higgs-boson information. These two versions of the reconstruction BDT are described in Section 6.2. The MEM\(_{D1}\) variable is only used in SR\(_{1}^{\geq 6j}\), while variables based on the b-tagging discriminant are not used in this region.

<table>
<thead>
<tr>
<th>Variable</th>
<th>Definition</th>
<th>SR(_{1}^{\geq 6j})</th>
<th>SR(_{1}^{5j})</th>
</tr>
</thead>
<tbody>
<tr>
<td>General kinematic variables</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>(\Delta R^{\text{avg}}_{bb})</td>
<td>Average (\Delta R) for all (b)-tagged jet pairs</td>
<td>✓</td>
<td>✓</td>
</tr>
<tr>
<td>(\Delta R^{\text{max}}_{bb})</td>
<td>(\Delta R) between the two (b)-tagged jets with the largest vector sum (p_T)</td>
<td>✓</td>
<td>–</td>
</tr>
<tr>
<td>(\Delta \eta^{\text{max}}_{jj})</td>
<td>Maximum (\Delta \eta) between any two jets</td>
<td>✓</td>
<td>✓</td>
</tr>
<tr>
<td>(m^{\text{min}}_{\Delta R bb})</td>
<td>Mass of the combination of two (b)-tagged jets with the smallest (\Delta R)</td>
<td>✓</td>
<td>–</td>
</tr>
<tr>
<td>(m^{\text{min}}_{\Delta R jj})</td>
<td>Mass of the combination of any two jets with the smallest (\Delta R)</td>
<td>–</td>
<td>✓</td>
</tr>
<tr>
<td>(N^{\text{Higgs}}_{bb})</td>
<td>Number of (b)-tagged jet pairs with invariant mass within 30 GeV of the Higgs-boson mass</td>
<td>✓</td>
<td>✓</td>
</tr>
<tr>
<td>(H^{\text{had}}_\ell)</td>
<td>Scalar sum of jet (p_T)</td>
<td>–</td>
<td>✓</td>
</tr>
<tr>
<td>(\Delta \eta^{\text{min}}_{Higgs bb})</td>
<td>(\Delta \eta) between the lepton and the combination of the two (b)-tagged jets with the smallest (\Delta R)</td>
<td>–</td>
<td>✓</td>
</tr>
<tr>
<td>Aplanarity</td>
<td>1.5(\lambda_2), where (\lambda_2) is the second eigenvalue of the momentum tensor built with all jets</td>
<td>✓</td>
<td>✓</td>
</tr>
<tr>
<td>(H_1)</td>
<td>Second Fox–Wolfram moment computed using all jets and the lepton</td>
<td>✓</td>
<td>✓</td>
</tr>
<tr>
<td>Variables from reconstruction BDT</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>BDT output</td>
<td>Output of the reconstruction BDT</td>
<td>✓*</td>
<td>✓*</td>
</tr>
<tr>
<td>(m_{\text{Higgs}}^{bb})</td>
<td>Higgs candidate mass</td>
<td>✓</td>
<td>✓</td>
</tr>
<tr>
<td>(m_{H,\text{lep top}})</td>
<td>Mass of Higgs candidate and (b)-jet from leptonic top candidate</td>
<td>✓</td>
<td>–</td>
</tr>
<tr>
<td>(\Delta R^{Higgs}_{bb})</td>
<td>(\Delta R) between (b)-jets from the Higgs candidate</td>
<td>✓</td>
<td>✓</td>
</tr>
<tr>
<td>(\Delta R_{H,\ell\ell})</td>
<td>(\Delta R) between Higgs candidate and (t\bar{t}) candidate system</td>
<td>✓*</td>
<td>✓*</td>
</tr>
<tr>
<td>(\Delta R_{H,\text{lep top}})</td>
<td>(\Delta R) between Higgs candidate and leptonic top candidate</td>
<td>✓</td>
<td>–</td>
</tr>
<tr>
<td>(\Delta R_{H,\text{had top}})</td>
<td>(\Delta R) between Higgs candidate and (b)-jet from hadronic top candidate</td>
<td>–</td>
<td>✓*</td>
</tr>
<tr>
<td>Variables from likelihood and matrix element method calculations</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>LHD</td>
<td>Likelihood discriminant</td>
<td>✓</td>
<td>✓</td>
</tr>
<tr>
<td>MEM(_{D1})</td>
<td>Matrix element discriminant (in SR(_{1}^{\geq 6j}) only)</td>
<td>✓</td>
<td>–</td>
</tr>
<tr>
<td>Variables from (b)-tagging (not in SR(_{1}^{\geq 6j}))</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>(w^{\text{Higgs}}_{b\text{-tag}})</td>
<td>Sum of (b)-tagging discriminants of jets from best Higgs candidate from the reconstruction BDT</td>
<td>✓</td>
<td>✓</td>
</tr>
<tr>
<td>(B^3_{\text{jet}})</td>
<td>3rd largest jet (b)-tagging discriminant</td>
<td>✓</td>
<td>✓</td>
</tr>
<tr>
<td>(B^4_{\text{jet}})</td>
<td>4th largest jet (b)-tagging discriminant</td>
<td>✓</td>
<td>✓</td>
</tr>
<tr>
<td>(B^5_{\text{jet}})</td>
<td>5th largest jet (b)-tagging discriminant</td>
<td>✓</td>
<td>✓</td>
</tr>
</tbody>
</table>
Table 8: Input variables to the classification BDT in the boosted single-lepton signal region. Additional $b$-jets are $b$-jets not contained in the Higgs-boson and top-quark candidates.

<table>
<thead>
<tr>
<th>Variable</th>
<th>Definition</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Variables from jet reclustering</strong></td>
<td></td>
</tr>
<tr>
<td>$\Delta R_{H,t}$</td>
<td>$\Delta R$ between the Higgs-boson and top-quark candidates</td>
</tr>
<tr>
<td>$\Delta R_{t,b_{\text{add}}}$</td>
<td>$\Delta R$ between the top-quark candidate and additional $b$-jet</td>
</tr>
<tr>
<td>$\Delta R_{H,b_{\text{add}}}$</td>
<td>$\Delta R$ between the Higgs-boson candidate and additional $b$-jet</td>
</tr>
<tr>
<td>$\Delta R_{H,\ell}$</td>
<td>$\Delta R$ between the Higgs-boson candidate and lepton</td>
</tr>
<tr>
<td>$m_{\text{Higgs candidate}}$</td>
<td>Higgs-boson candidate mass</td>
</tr>
<tr>
<td>$\sqrt{d_{12}}$</td>
<td>Top-quark candidate first splitting scale [100]</td>
</tr>
<tr>
<td><strong>Variables from $b$-tagging</strong></td>
<td></td>
</tr>
<tr>
<td>$w_{b_{\text{tag}}}$</td>
<td>Sum of $b$-tagging discriminants of all $b$-jets</td>
</tr>
<tr>
<td>$w_{\text{add}}^{b_{\text{tag}}}/w_{b_{\text{tag}}}$</td>
<td>Ratio of sum of $b$-tagging discriminants of additional $b$-jets to all $b$-jets</td>
</tr>
</tbody>
</table>
Acknowledgments

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

We acknowledge the support of ANPCyT, Argentina; YerPhI, Armenia; ARC, Australia; 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; MSMT CR, MPO CR and VSC CR, Czech Republic; DNRF and DNSRC, Denmark; IN2P3-CNRS, CEA-DRF/IRFU, France; SRNSF, Georgia; BMWF, HGF, and MPG, Germany; GSRT, Greece; RGC, Hong Kong SAR, China; ISF, I-CORE and Benoziyo Center, Israel; INFN, Italy; MEXT and JSPS, Japan; CNRST, Morocco; NWO, Netherlands; RCN, Norway; MNIST and NCN, Poland; FCT, Portugal; MNE/IFA, Romania; MES of Russia and NRC KI, Russian Federation; JINR; MESTD, Serbia; MSSR, Slovakia; ARRS and MIZŠ, Slovenia; DST/NRF, South Africa; MINECO, Spain; SRC and Wallenberg Foundation, Sweden; SERI, SNSF and Cantons of Bern and Geneva, Switzerland; MOST, Taiwan; TAEK, Turkey; STFC, United Kingdom; DOE and NSF, United States of America. In addition, individual groups and members have received support from BCKDF, the Canada Council, CANARIE, CRC, Compute Canada, FQRNT, and the Ontario Innovation Trust, Canada; EPLANET, ERC, ERDF, FP7, Horizon 2020 and Marie Skłodowska-Curie Actions, European Union; Investissements d’Avenir Labex and Idex, ANR, Région Auvergne and Fondation Partager le Savoir, France; DFG and AvH Foundation, Germany; Herakleitos, Thales and Aristeia programmes co-financed by EU-ESF and the Greek NSRF; BSF, GIF and Minerva, Israel; BRF, Norway; CERCA Programme Generalitat de Catalunya, Generalitat Valenciana, Spain; the Royal Society and Leverhulme Trust, United Kingdom.

The crucial computing support from all WLCG partners is acknowledged gratefully, in particular from CERN, the ATLAS Tier-1 facilities at TRIUMF (Canada), NDGF (Denmark, Norway, Sweden), CC-IN2P3 (France), KIT/GridKA (Germany), INFN-CNAF (Italy), NL-T1 (Netherlands), PIC (Spain), ASGC (Taiwan), RAL (UK) and BNL (USA), the Tier-2 facilities worldwide and large non-WLCG resource providers. Major contributors of computing resources are listed in Ref. [101].
References


[17] ATLAS Collaboration, 

[18] CMS Collaboration, 


[93] ATLAS Collaboration, Studies of tinnitus production with MadGraph5_aMC@NLO and Herwig++ for the ATLAS experiment, ATL-PHYS-PUB-2016-011, 2016, url: https://cds.cern.ch/record/2153876.


The ATLAS Collaboration

6 High Energy Physics Division, Argonne National Laboratory, Argonne IL, United States of America
7 Department of Physics, University of Arizona, Tucson AZ, United States of America
8 Department of Physics, The University of Texas at Arlington, Arlington TX, United States of America
9 Physics Department, National and Kapodistrian University of Athens, Athens, Greece
10 Physics Department, National Technical University of Athens, Zografou, Greece
11 Department of Physics, The University of Texas at Austin, Austin TX, United States of America
12 Institute of Physics, Azerbaijan Academy of Sciences, Baku, Azerbaijan
13 Institut de Física d’Altes Energies (IAFE), The Barcelona Institute of Science and Technology, Barcelona, Spain
14 Institute of Physics, University of Belgrade, Belgrade, Serbia
15 Department for Physics and Technology, University of Bergen, Bergen, Norway
16 Physics Division, Lawrence Berkeley National Laboratory and University of California, Berkeley CA, United States of America
17 Department of Physics, Humboldt University, Berlin, Germany
18 Albert Einstein Center for Fundamental Physics and Laboratory for High Energy Physics, University of Bern, Bern, Switzerland
19 School of Physics and Astronomy, University of Birmingham, Birmingham, United Kingdom
20 (a) Department of Physics, Bogazici University, Istanbul; (b) Department of Physics Engineering, Gaziantep University, Gaziantep; (c) Istanbul Bilgi University, Faculty of Engineering and Natural Sciences, Istanbul; (d) Bahcesehir University, Faculty of Engineering and Natural Sciences, Istanbul, Turkey
21 Centro de Investigaciones, Universidad Antonio Narino, Bogota, Colombia
22 (a) INFN Sezione di Bologna; (b) Dipartimento di Fisica e Astronomia, Università di Bologna, Bologna, Italy
23 Physikalisches Institut, University of Bonn, Bonn, Germany
24 Department of Physics, Boston University, Boston MA, United States of America
25 Department of Physics, Brandeis University, Waltham MA, United States of America
26 (a) Universidade Federal do Rio De Janeiro COPPE/EE/IF, Rio de Janeiro; (b) Electrical Circuits Department, Federal University of Juiz de Fora (UFJF), Juiz de Fora; (c) Federal University of Sao Joao del Rei (UFSJ), Sao Joao del Rei; (d) Instituto de Fisica, Universidade de Sao Paulo, Sao Paulo, Brazil
27 Physics Department, Brookhaven National Laboratory, Upton NY, United States of America
28 (a) Transilvania University of Brasov, Brasov; (b) Horia Hulubei National Institute of Physics and Nuclear Engineering, Bucharest; (c) Department of Physics, Alexandru Ioan Cuza University of Iasi, Iasi; (d) National Institute for Research and Development of Isotopic and Molecular Technologies, Physics Department, Cluj Napoca; (e) University Politehnica Bucharest, Bucharest; (f) West University in Timisoara, Timisoara, Romania
29 Departamento de Física, Universidad de Buenos Aires, Buenos Aires, Argentina
30 Cavendish Laboratory, University of Cambridge, Cambridge, United Kingdom
31 Department of Physics, Carleton University, Ottawa ON, Canada
32 CERN, Geneva, Switzerland
33 Enrico Fermi Institute, University of Chicago, Chicago IL, United States of America
34 (a) Departamento de Física, Pontificia Universidad Católica de Chile, Santiago; (b) Departamento de Física, Universidad Técnica Federico Santa María, Valparaíso, Chile
35 (a) Institute of High Energy Physics, Chinese Academy of Sciences, Beijing; (b) Department of Physics, Nanjing University, Jiangsu; (c) Physics Department, Tsinghua University, Beijing 100084; (d) University of Chinese Academy of Science (UCAS), Beijing, China
36 (a) School of Physics, Shandong University, Shandong; (b) School of Physics and Astronomy, Key
71 Faculty of Science, Kyoto University, Kyoto, Japan
72 Kyoto University of Education, Kyoto, Japan
73 Research Center for Advanced Particle Physics and Department of Physics, Kyushu University, Fukuoka, Japan
74 Instituto de Física La Plata, Universidad Nacional de La Plata and CONICET, La Plata, Argentina
75 Physics Department, Lancaster University, Lancaster, United Kingdom
76 (a) INFN Sezione di Lecce; (b) Dipartimento di Matematica e Fisica, Università del Salento, Lecce, Italy
77 Oliver Lodge Laboratory, University of Liverpool, Liverpool, United Kingdom
78 Department of Experimental Particle Physics, Jožef Stefan Institute and Department of Physics, University of Ljubljana, Ljubljana, Slovenia
79 School of Physics and Astronomy, Queen Mary University of London, London, United Kingdom
80 Department of Physics, Royal Holloway University of London, Surrey, United Kingdom
81 Department of Physics and Astronomy, University College London, London, United Kingdom
82 Louisiana Tech University, Ruston LA, United States of America
83 Laboratoire de Physique Nucléaire et de Hautes Energies, UPMC and Université Paris-Diderot and CNRS/IN2P3, Paris, France
84 Fysiska institutionen, Lunds universitet, Lund, Sweden
85 Departamento de Física Teórica C-15, Universidad Autonoma de Madrid, Madrid, Spain
86 Institut für Physik, Universität Mainz, Mainz, Germany
87 School of Physics and Astronomy, University of Manchester, Manchester, United Kingdom
88 CPPM, Aix-Marseille Université and CNRS/IN2P3, Marseille, France
89 Department of Physics, University of Massachusetts, Amherst MA, United States of America
90 Department of Physics, McGill University, Montreal QC, Canada
91 School of Physics, University of Melbourne, Victoria, Australia
92 Department of Physics, The University of Michigan, Ann Arbor MI, United States of America
93 Department of Physics and Astronomy, Michigan State University, East Lansing MI, United States of America
94 (a) INFN Sezione di Milano; (b) Dipartimento di Fisica, Università di Milano, Milano, Italy
95 B.I. Stepanov Institute of Physics, National Academy of Sciences of Belarus, Minsk, Republic of Belarus
96 Research Institute for Nuclear Problems of Byelorussian State University, Minsk, Republic of Belarus
97 Group of Particle Physics, University of Montreal, Montreal QC, Canada
98 P.N. Lebedev Physical Institute of the Russian Academy of Sciences, Moscow, Russia
99 Institute for Theoretical and Experimental Physics (ITEP), Moscow, Russia
100 National Research Nuclear University MEPhI, Moscow, Russia
101 D.V. Skobeltsyn Institute of Nuclear Physics, M.V. Lomonosov Moscow State University, Moscow, Russia
102 Fakultät für Physik, Ludwig-Maximilians-Universität München, München, Germany
103 Max-Planck-Institut für Physik (Werner-Heisenberg-Institut), München, Germany
104 Nagasaki Institute of Applied Science, Nagasaki, Japan
105 Graduate School of Science and Kobayashi-Maskawa Institute, Nagoya University, Nagoya, Japan
106 (a) INFN Sezione di Napoli; (b) Dipartimento di Fisica, Università di Napoli, Napoli, Italy
107 Department of Physics and Astronomy, University of New Mexico, Albuquerque NM, United States of America
108 Institute for Mathematics, Astrophysics and Particle Physics, Radboud University Nijmegen/Nikhef, Nijmegen, Netherlands
Nikhef National Institute for Subatomic Physics and University of Amsterdam, Amsterdam, Netherlands
Department of Physics, Northern Illinois University, DeKalb IL, United States of America
Budker Institute of Nuclear Physics, SB RAS, Novosibirsk, Russia
Department of Physics, New York University, New York NY, United States of America
Ohio State University, Columbus OH, United States of America
Faculty of Science, Okayama University, Okayama, Japan
Homer L. Dodge Department of Physics and Astronomy, University of Oklahoma, Norman OK, United States of America
Department of Physics, Oklahoma State University, Stillwater OK, United States of America
Palacký University, RCPTM, Olomouc, Czech Republic
Center for High Energy Physics, University of Oregon, Eugene OR, United States of America
LAL, Univ. Paris-Sud, CNRS/IN2P3, Université Paris-Saclay, Orsay, France
Graduate School of Science, Osaka University, Osaka, Japan
Department of Physics, University of Oslo, Oslo, Norway
Department of Physics, Oxford University, Oxford, United Kingdom

(a) INFN Sezione di Pavia; (b) Dipartimento di Fisica, Università di Pavia, Pavia, Italy
Department of Physics, University of Pennsylvania, Philadelphia PA, United States of America
National Research Centre "Kurchatov Institute" B.P.Konstantinov Petersburg Nuclear Physics Institute, St. Petersburg, Russia

(a) INFN Sezione di Pisa; (b) Dipartimento di Fisica E. Fermi, Università di Pisa, Pisa, Italy
Department of Physics and Astronomy, University of Pittsburgh, Pittsburgh PA, United States of America

(a) Laboratório de Instrumentação e Física Experimental de Partículas - LIP, Lisboa; (b) Faculdade de Ciências, Universidade de Lisboa, Lisboa; (c) Department of Physics, University of Coimbra, Coimbra; (d) Centro de Física Nuclear da Universidade de Lisboa, Lisboa; (e) Departamento de Física, Universidade do Minho, Braga; (f) Departamento de Física Teorica y del Cosmos, Universidad de Granada, Granada; (g) Dep Física and CEFITEC of Faculdade de Ciencias e Tecnologia, Universidade Nova de Lisboa, Caparica, Portugal
Institute of Physics, Academy of Sciences of the Czech Republic, Praha, Czech Republic
Czech Technical University in Prague, Praha, Czech Republic
Charles University, Faculty of Mathematics and Physics, Prague, Czech Republic
State Research Center Institute for High Energy Physics (Protvino), NRC KI, Russia
Particle Physics Department, Rutherford Appleton Laboratory, Didcot, United Kingdom

(a) INFN Sezione di Roma; (b) Dipartimento di Fisica, Sapienza Università di Roma, Roma, Italy

(a) INFN Sezione di Roma Tor Vergata; (b) Dipartimento di Fisica, Università di Roma Tor Vergata, Roma, Italy

(a) INFN Sezione di Roma Tre; (b) Dipartimento di Matematica e Fisica, Università Roma Tre, Roma, Italy

(a) Faculté des Sciences Ain Chock, Réseau Universitaire de Physique des Hautes Energies - Université Hassan II, Casablanca; (b) Centre National de l’Energie des Sciences Techniques Nucléaires, Rabat; (c) Faculté des Sciences Semlalia, Université Cadi Ayyad, LPHEA-Marrakech; (d) Faculté des Sciences, Université Mohamed Premier and LPTPM, Oujda; (e) Faculté des sciences, Université Mohammed V, Rabat, Morocco

DSM/IRFU (Institut de Recherches sur les Lois Fondamentales de l’Univers), CEA Saclay (Commissariat à l’Energie Atomique et aux Energies Alternatives), Gif-sur-Yvette, France
Santa Cruz Institute for Particle Physics, University of California Santa Cruz, Santa Cruz CA, United States of America
States of America
140 Department of Physics, University of Washington, Seattle WA, United States of America
141 Department of Physics and Astronomy, University of Sheffield, Sheffield, United Kingdom
142 Department of Physics, Shinshu University, Nagano, Japan
143 Department Physik, Universität Siegen, Siegen, Germany
144 Department of Physics, Simon Fraser University, Burnaby BC, Canada
145 SLAC National Accelerator Laboratory, Stanford CA, United States of America
146 (a) Faculty of Mathematics, Physics & Informatics, Comenius University, Bratislava; (b) Department of Subnuclear Physics, Institute of Experimental Physics of the Slovak Academy of Sciences, Kosice, Slovak Republic
147 (a) Department of Physics, University of Cape Town, Cape Town; (b) Department of Physics, University of Johannesburg, Johannesburg; (c) School of Physics, University of the Witwatersrand, Johannesburg, South Africa
148 (a) Department of Physics, Stockholm University; (b) The Oskar Klein Centre, Stockholm, Sweden
149 Physics Department, Royal Institute of Technology, Stockholm, Sweden
150 Departments of Physics & Astronomy and Chemistry, Stony Brook University, Stony Brook NY, United States of America
151 Department of Physics and Astronomy, University of Sussex, Brighton, United Kingdom
152 School of Physics, University of Sydney, Sydney, Australia
153 Institute of Physics, Academia Sinica, Taipei, Taiwan
154 Department of Physics, Technion: Israel Institute of Technology, Haifa, Israel
155 Raymond and Beverly Sackler School of Physics and Astronomy, Tel Aviv University, Tel Aviv, Israel
156 Department of Physics, Aristotle University of Thessaloniki, Thessaloniki, Greece
157 International Center for Elementary Particle Physics and Department of Physics, The University of Tokyo, Tokyo, Japan
158 Graduate School of Science and Technology, Tokyo Metropolitan University, Tokyo, Japan
159 Department of Physics, Tokyo Institute of Technology, Tokyo, Japan
160 Tomsk State University, Tomsk, Russia
161 Department of Physics, University of Toronto, Toronto ON, Canada
162 (a) INFN-TIFPA; (b) University of Trento, Trento, Italy
163 (a) TRIUMF, Vancouver BC; (b) Department of Physics and Astronomy, York University, Toronto ON, Canada
164 Faculty of Pure and Applied Sciences, and Center for Integrated Research in Fundamental Science and Engineering, University of Tsukuba, Tsukuba, Japan
165 Department of Physics and Astronomy, Tufts University, Medford MA, United States of America
166 Department of Physics and Astronomy, University of California Irvine, Irvine CA, United States of America
167 (a) INFN Gruppo Collegato di Udine, Sezione di Trieste, Udine; (b) ICTP, Trieste; (c) Dipartimento di Chimica, Fisica e Ambiente, Università di Udine, Udine, Italy
168 Department of Physics and Astronomy, University of Uppsala, Uppsala, Sweden
169 Department of Physics, University of Illinois, Urbana IL, United States of America
170 Instituto de Fisica Corpuscular (IFIC), Centro Mixto Universidad de Valencia - CSIC, Spain
171 Department of Physics, University of British Columbia, Vancouver BC, Canada
172 Department of Physics and Astronomy, University of Victoria, Victoria BC, Canada
173 Department of Physics, University of Warwick, Coventry, United Kingdom
174 Waseda University, Tokyo, Japan
175 Department of Particle Physics, The Weizmann Institute of Science, Rehovot, Israel
Department of Physics, University of Wisconsin, Madison WI, United States of America
Fakultät für Physik und Astronomie, Julius-Maximilians-Universität, Würzburg, Germany
Fakultät für Mathematik und Naturwissenschaften, Fachgruppe Physik, Bergische Universität Wuppertal, Wuppertal, Germany
Department of Physics, Yale University, New Haven CT, United States of America
Yerevan Physics Institute, Yerevan, Armenia
Centre de Calcul de l’Institut National de Physique Nucléaire et de Physique des Particules (IN2P3), Villeurbanne, France
Academia Sinica Grid Computing, Institute of Physics, Academia Sinica, Taipei, Taiwan
a Also at Department of Physics, King’s College London, London, United Kingdom
b Also at Institute of Physics, Azerbaijan Academy of Sciences, Baku, Azerbaijan
c Also at Novosibirsk State University, Novosibirsk, Russia
d Also at TRIUMF, Vancouver BC, Canada
e Also at Department of Physics & Astronomy, University of Louisville, Louisville, KY, United States of America
f Also at Physics Department, An-Najah National University, Nablus, Palestine
g Also at Department of Physics, California State University, Fresno CA, United States of America
h Also at Department of Physics, University of Fribourg, Fribourg, Switzerland
i Also at II Physikalisches Institut, Georg-August-Universität, Göttingen, Germany
j Also at Departament de Fisica de la Universitat Autonoma de Barcelona, Barcelona, Spain
k Also at Departamento de Fisica e Astronomia, Faculdade de Ciencias, Universidade do Porto, Portugal
l Also at Tomsk State University, Tomsk, and Moscow Institute of Physics and Technology State University, Dolgoprudny, Russia
m Also at The Collaborative Innovation Center of Quantum Matter (CICQM), Beijing, China
n Also at Universita di Napoli Parthenope, Napoli, Italy
o Also at Institute of Particle Physics (IPP), Canada
p Also at Horia Hulubei National Institute of Physics and Nuclear Engineering, Bucharest, Romania
q Also at Department of Physics, St. Petersburg State Polytechnical University, St. Petersburg, Russia
r Also at Borough of Manhattan Community College, City University of New York, New York City, United States of America
s Also at Department of Financial and Management Engineering, University of the Aegean, Chios, Greece
t Also at Centre for High Performance Computing, CSIR Campus, Rosebank, Cape Town, South Africa
u Also at Louisiana Tech University, Ruston LA, United States of America
v Also at Instituto de Fisica Energetica e Nucleica, Lisboa, Portugal
w Also at Instituto de Fisica Energetica e Nucleica, Lisboa, Portugal
x Also at Department of Physics, The University of Michigan, Ann Arbor MI, United States of America
y Also at LAL, Univ. Paris-Sud, CNRS/IN2P3, Université Paris-Saclay, Orsay, France
z Also at Graduate School of Science, Osaka University, Osaka, Japan
aa Also at Institute for Mathematics, Astrophysics and Particle Physics, Radboud University Nijmegen/Nikhef, Nijmegen, Netherlands
ab Also at Institute of Theoretical Physics, Ilia State University, Tbilisi, Georgia
ac Also at CERN, Geneva, Switzerland
ad Also at Georgian Technical University (GTU), Tbilisi, Georgia
ae Also at Ochadai Academic Production, Ochanomizu University, Tokyo, Japan
af Also at Manhattan College, New York NY, United States of America
ag Also at Hellenic Open University, Patras, Greece
Also at The City College of New York, New York NY, United States of America

Also at Departamento de Física Teórica y del Cosmos, Universidad de Granada, Granada, Portugal

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

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

Also at Departement de Physique Nucleaire et Corpusculaire, Université de Genève, Geneva, Switzerland

Also at Department of Physics, The University of Texas at Austin, Austin TX, United States of America

Also at Departement de Physique Nucleaire et Corpusculaire, Université de Genève, Geneva, Switzerland

Also at Department of Physics, The University of Texas at Austin, Austin TX, United States of America

Also at Institut de Física d’Altes Energies (IFAE), The Barcelona Institute of Science and Technology, Barcelona, Spain

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

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

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

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

Also at National Research Nuclear University MEPhI, Moscow, Russia

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

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

Also at Giresun University, Faculty of Engineering, Turkey

Also at Institute of Physics, Academia Sinica, Taipeh, Taiwan

Also at University of Malaya, Department of Physics, Kuala Lumpur, Malaysia

* Deceased