Search for non-resonant Higgs boson pair production in the $b\bar{b}\ell\nu\ell\nu$ final state with the ATLAS detector in $pp$ collisions at $\sqrt{s} = 13$ TeV

The ATLAS Collaboration*

**Abstract**

A search for non-resonant Higgs boson pair production, as predicted by the Standard Model, is presented, where one of the Higgs bosons decays via the $H \rightarrow bb$ channel and the other via one of the $H \rightarrow WW^*/ZZ^*/\tau\tau$ channels. The analysis selection requires events to have at least two $b$-tagged jets and exactly two leptons (electrons or muons) with opposite electric charge in the final state. Candidate events consistent with Higgs boson pair production are selected using a multi-class neural network discriminant. The analysis uses 139 fb$^{-1}$ of $pp$ collision data recorded at a centre-of-mass energy of 13 TeV by the ATLAS detector at the Large Hadron Collider. An observed (expected) upper limit of 1.2 (0.9±0.1) pb is set on the non-resonant Higgs boson pair production cross-section at 95% confidence level, which is equivalent to 40 (29±14) times the value predicted in the Standard Model.

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

1. Introduction

In 2012, the ATLAS and CMS Collaborations reported the observation of a new particle in the search for the Standard Model (SM) Higgs boson ($H$) [1,2]. So far, measurements of the spin and couplings of the new particle are consistent with those predicted by the Brout–Englert–Higgs (BEH) mechanism of the SM [3–12]. The SM predicts non-resonant production of Higgs boson pairs ($HH$) in proton–proton ($pp$) collisions, referred to as non-resonant $HH$ production, with the dominant production modes at the LHC proceeding via the gluon–gluon fusion ($ggF$) process. The $ggF$ process has two leading order contributions: the first corresponds to the so-called ‘triangle diagram’, which includes the Higgs boson self-coupling, and the second is the so-called ‘box diagram’, which includes a heavy-quark loop with two fermion–fermion–Higgs ($ffH$) vertices. These two amplitudes interfere destructively, resulting in a low cross-section of only $31.05 \pm 1.90$ fb for the $ggF$ $HH$ production mode, computed at next-to-next-to-leading order (NNLO) and including finite top-quark mass effects [13–20]. Feynman diagrams illustrating these two contributions are shown in Fig. 1. The measurement of non-resonant $HH$ production at the LHC stands as an important test of the BEH mechanism. In many beyond-the-SM (BSM) theories, $HH$ production can be enhanced by modifying the Higgs boson self-coupling, $\lambda_{HHH}$, or the top-quark Yukawa coupling, $y_t$, and/or by introducing new contact interactions between two top-quarks or gluons and two Higgs bosons or introducing production mechanisms via intermediate BSM particles [21–23].

The ATLAS and CMS Collaborations have performed searches for non-resonant $HH$ production in a variety of final states at 13 TeV [24–33]. No significant excess of events beyond SM expectations is observed in these searches, with the ATLAS and CMS data-analyses setting observed (expected) limits on non-resonant $HH$ production to be no larger than 6.9 (10.0) and 22.2 (12.8) times the predicted rate in the SM, respectively [34,35].

This Letter describes a search for non-resonant $HH$ production in the $bb\ell\nu\ell\nu$ final state, where $\ell$ refers to a lepton (either an electron or a muon), using 13 TeV $pp$ collision data collected with the ATLAS detector during 2015–2018 and corresponding to a total integrated luminosity of 139 fb$^{-1}$. The analysis uses machine-learning techniques based on feedforward neural network architectures [38] to construct an event-level classifier trained to distinguish between the $HH$ signal and SM backgrounds. Analyses searching for non-resonant $HH$ production via similar decay channels were performed previously in the single-lepton final state.
by ATLAS in searches for \( HH \rightarrow bbWW^* \) [28] and in the dilepton channel by CMS in searches for \( HH \rightarrow bbWW^*/bbZZ^* \) [31].

2. ATLAS detector

The ATLAS detector [37–39] is a general-purpose particle detector with forward–backward symmetric cylindrical geometry. It includes an inner tracking detector (ID), immersed in an axial magnetic field, which provides precision tracking of charged particles over the range of \(|\eta| < 2.5\). Calorimeter systems with either liquid argon or scintillator tiles as the active medium provide energy measurements over the range of \(|\eta| < 4.9\). The muon spectrometer (MS) is positioned outside the calorimeters and includes three air-core toroidal magnets. The MS is composed of several types of muon detectors which provide trigger and high-precision tracking capabilities for \(|\eta| < 2.4\) and \(|\eta| < 2.7\), respectively. A hardware-based trigger followed by a software-based trigger reduce the recorded event rate to an average of 1 kHz [40].

3. Dataset and simulated events

The data used for this search were collected in \( pp \) collisions at the LHC with a centre-of-mass energy of 13 TeV. Only those data collected during stable LHC beam conditions and with all ATLAS detector subsystems fully operational are used, and correspond to an integrated luminosity of 139 fb\(^{-1}\). The selection of candidate events with oppositely charged leptons is based on a combination of single-lepton and dilepton triggers.\(^2\) The use of a given trigger depends on the flavour and the transverse momenta (\( p_T \)) of the two (\( p_T \)-ordered) leptons in the event, and on the data-taking period. Single-lepton triggers with \( p_T \) thresholds between 22 and 28 GeV are given priority over dilepton triggers. The criteria of the dilepton triggers are checked only if no single-lepton trigger criteria are met and have \( p_T \) thresholds as low as 19 (10) GeV for the leading (subleading) lepton. At least one reconstructed lepton (or lepton pair) has to match a corresponding trigger object, in which case their offline \( p_T \) must be higher than the trigger threshold by at least 2 GeV, in order to be on the efficiency plateau of the corresponding trigger.

Monte Carlo (MC) simulation [41] is used to model the signal processes and in the estimation of SM background processes. A GEANT4 [42] simulation of the ATLAS detector was used for the background processes. The signal MC samples were processed with a fast simulation that relies on a parameterisation of the calorimeter response [41] and on GEANT4 for the tracking detectors. Simulated events are reconstructed using the same algorithms as used for data and include the effects of multiple \( pp \) interactions in the same or neighbouring bunch crossings, collectively referred to as pile-up. The simulation of pile-up collisions was performed with PYTHIA 8.186 [43] using the ATLAS A3 set of tuned parameters [44] and the NNPDF2.3LO parton distribution function (PDF) set [45]. Simulated events were reweighted to match the distribution of pile-up interactions in data. The average amount of pile-up in the data collected during 2015–2018 was 33.7.

The signal processes with ggF-initiated non-resonant \( HH \) production in the \( bb\ell\ell\nu\nu \) final state were generated with an effective Lagrangian in the infinite top-quark mass approximation. The generated signal events were reweighted with form factors that take into account the finite mass of the top-quark [46,47]. SM background processes were simulated using different MC event generators. The MC matrix element (ME) event generators and PDF sets, the parton showering (PS) and the underlying event (UE) modelling, UE tuned parameters (tune), and the accuracy of the theoretical cross-sections used to normalise the simulated processes are summarised in Table 1. Each SM background process is normalised to the best available respective theoretical cross-section. The mass of the Higgs boson was set to 125 GeV for all signal and background processes. The \( HH \) branching fractions (BF) predicted by the SM [13] are used for all Higgs boson decays. MadSpin [48] was used to model top-quark spin correlations and EvGen [49] was used to model properties of \( b \)- and \( c \)-hadron decays for processes using \( p_T \) and for the signal processes.

SM top-quark pair production \((t\bar{t})\) and the production of single top-quarks in association with \( W \) bosons \((Wt)\) contribute with significant background contamination in the \( bb\ell\ell\nu\nu \) final state. At next-to-leading-order (NLO) accuracy, there exists non-trivial interference between these two processes that may be enhanced in phase-space regions wherein there are high fractions of \( Wt \) events [50]. Two schemes are typically used to remove the overlap between these two processes: the so-called diagram removal (DR) and diagram subtraction (DS) schemes [51]; the former is used in the present analysis to remove the overlapping events and the latter is used to evaluate the systematic uncertainty in corresponding background event yields. Because of these effects, the sum of the simulated \( t\bar{t} \) and \( Wt \) processes is considered as a single background process and referred to as the ‘Top’ process in what follows.

4. Event selection and object definitions

Selected events are required to have at least one \( pp \) interaction vertex reconstructed from at least two ID tracks with \( p_T > 0.4 \) GeV. The primary vertex for each event is defined as the vertex with the highest \( \sum (p_T)^2 \) of associated ID tracks [102]. Events that contain at least one jet arising from non-collision sources or detector noise are rejected by a set of quality criteria [103].

Loose and signal criteria are defined in order to select reconstructed lepton and jet candidates, where the latter is a subset of the former. Compared to the loose objects, the signal objects are required to satisfy tighter identification or quality criteria that are designed to suppress background contributions. Reconstructed loose (signal) electrons are required to satisfy the ‘Loose’ (‘Tight’) likelihood identification criteria [104]. Loose electrons are required to have \( p_T > 10 \) GeV and to be within \(|\eta| < 2.47\). In addition, signal electrons are required to be outside the range \( 1.37 < |\eta| < 1.52\), which corresponds to the transition regions between the barrel and endcaps of the electromagnetic calorimeters. In order to reduce background contributions from jets misidentified as electrons, signal electrons are required to be isolated according to the ‘Gradient’ selection criteria [104]. Reconstructed loose and signal muon candidates are required to have \( p_T > 10 \) GeV, to be within \(|\eta| < 2.4\), and to satisfy the ‘Medium’ identification criteria [105].

Additionally, signal muons are required to be isolated according to the ‘FixedCutLoose’ selection criteria [105]. Signal electron (muon) candidates are required to originate from the primary vertex by demanding that the significance of the transverse impact parameter, defined as the absolute value of the track transverse impact parameter, \( d_0 \), measured relative to the primary vertex, divided by its uncertainty, \( \sigma_{d_0} \), satisfy \(|d_0|/\sigma_{d_0} < 3\) (3). The difference \( \Delta z \) between the value of the \( z \) coordinate of the point on the track at which \( d_0 \) is defined and the longitudinal position of the primary

---

\(^1\) ATLAS uses a right-handed coordinate system with its origin at the nominal interaction point (IP) in the centre of the detector and the \( z \)-axis along the beam pipe. The \( x \)-axis points from the IP to the centre of the LHC ring, and the \( y \)-axis points upwards. Cylindrical coordinates \((r,\phi)\) are used in the transverse plane, \( \phi \) being the azimuthal angle around the \( z \)-axis. The pseudorapidity is defined in terms of the polar angle \( \theta \) as \( \eta = -\ln \tan(\theta/2) \). The angular distance is measured in units of \( \Delta \eta = \sqrt{(\Delta\eta)^2 + (\Delta\phi)^2} \).

\(^2\) Distinct sets of single-lepton triggers are used for electrons and muons. Dilepton triggers require either two electrons, two muons, or one electron and one muon.
vertex is required to satisfy $|\Delta z_0 \times \sin \theta| < 0.5$ mm, where $\theta$ is the polar angle of the track with respect to the $z$-axis.

Jets are reconstructed from topological clusters of energy deposits in the calorimeters [106] using the anti-$k_t$ algorithm [107, 108] with a radius parameter of $R = 0.4$ and calibrated as described in Ref. [109]. Candidate loose jets are required to have $p_T > 20$ GeV. Signal jets are required to have $|\eta| < 2.8$ and must satisfy pile-up suppression requirements based on the output of a multivariate classifier [110], which identifies jets consistent with a primary vertex in the region $|\eta| < 2.4$ and $p_T < 120$ GeV. The MV2c10 multivariate algorithm [111] is used to identify jets containing $b$-hadrons ($b$-tagged jets). A MV2c10 working point with a $b$-tagging efficiency of 70%, estimated from simulated $t\bar{t}$ events [112], is used. The $b$-tagged jets must have $p_T > 20$ GeV and $|\eta| < 2.5$. The momentum of $b$-tagged jets is adjusted using the muon-in-jet correction, as described in Ref. [6], by accounting for momentum losses due to muons originating from in-flight semileptonic $b$-hadron decays occurring within the $b$-tagged jet.

The missing transverse momentum $P_T^{\text{miss}}$, the magnitude of which is denoted by $E_T^{\text{miss}}$, is constructed from the negative vector sum of the transverse momenta of calibrated loose objects in the event. An additional term is included to account for the energy of ID tracks that are matched to the primary vertex in the event but not to any of the selected loose objects [113].

To avoid double-counting, loose objects are subject to the overlap removal procedure defined as follows. If a reconstructed electron and muon share a track in the ID, the electron is removed. However, if the muon sharing the track with the electron is calorimeter-tagged, then the muon is removed instead of the electron. If a jet and an electron are reconstructed within $\Delta R = 0.2$ of each other, then the jet is removed. If a jet and a muon are within $\Delta R = 0.2$ of each other, and the jet has less than three tracks or carries less than 50% of the muon $p_T$, then the jet is removed; otherwise, the muon is removed. Electrons or muons separated from the remaining jets by $\Delta R < 0.4$ are removed.

The analysis selects candidate events with exactly two oppositely charged signal leptons, electrons or muons, and at least two signal $b$-tagged jets. To enhance sensitivity to the signal process and to maximise rejection of the expected SM backgrounds, the analysis uses a multivariate approach to select signal events.

### 5. Analysis strategy

The analysis relies on the use of a multivariate discriminant designed to select candidate events consistent with non-resonant $HH$ production. Section 5.1 describes the architecture and the training of the deep neural network (DNN) classifier from which the discriminant is constructed. Section 5.2 describes the signal region selection criteria. Section 5.3 describes the final background estimation procedure.

#### 5.1. Deep learning approach to target $HH$

The discriminant uses the outputs of a DNN classifier that is built using the KERAS library with TENSORFLOW as a backend [114, 115] and uses the Lightning library [116] to interface with the analysis software infrastructure of the ATLAS experiment. The sample of events used for training is composed of equal numbers of events from the signal and each of the dominant background processes: Top (as defined in Section 3), $Z/\gamma^* \rightarrow \ell\ell$ ($Z\ell\ell$), and $Z/\gamma^* \rightarrow \tau\tau$ ($\tau\tau\ell\ell$) production. The signal sample used in the training of the classifier contains only the $HH \rightarrow bbWW$ component due to its larger BF relative to the $HH \rightarrow bb\tau\tau$ and $HH \rightarrow bbZZ$ components. However, the sum of all three signal components is evaluated as the signal when performing the statistical analysis. Additionally, all processes that make up the training sample ($HH \rightarrow bbWW$, Top, $Z\ell\ell$, and $\tau\tau\ell\ell$) have the same weight during the training of the classifier. The training sample is composed of simulated candidate events with $m_{WW} > 20$ GeV and having one or more $b$-tagged jets, where events with exactly one $b$-tagged jet are included to increase the number of events available for training. For the training events with exactly one $b$-tagged jet, each observable that requires at least two $b$-tagged jets is set to its mean value as computed with the full set of training events that contain at least two $b$-tagged jets. Observables that require two $b$-tagged jets are defined using the leading two $b$-tagged jets. The classifier contains two fully connected hidden layers each with 250 nodes. Rectified linear unit (ReLU) activations are used for each layer [117]. In order to improve the robustness of the training and to reduce effects due to overtraining, there is a dropout layer that randomly drops 50% of the nodes between the two fully connected layers during training [118]. The classifier produces four outputs that are passed through a softmax activation, constraining their sum to one [36].
Table 2

Description of the variables used as inputs to the DNN classifier.

<table>
<thead>
<tr>
<th>$p_T$, $\eta$, and $\phi$</th>
<th>$p_T$, $\eta$, and $\phi$ of the leptons, leading two jets (not necessarily b-tagged), and leading two b-tagged jets</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\Delta R_{\ell\ell}$</td>
<td>$\Delta R$ and magnitude of the $\Delta \phi$ between the two leptons</td>
</tr>
<tr>
<td>$m_{\ell\ell}$</td>
<td>Invariant mass and the transverse momentum of the dilepton system</td>
</tr>
<tr>
<td>$E_{\text{miss}}$, $\mathbf{p}_T^{\text{miss}}$</td>
<td>Magnitude of the missing transverse momentum vector and its $\phi$ component</td>
</tr>
<tr>
<td>$</td>
<td>\Delta \phi_{\ell\ell}</td>
</tr>
<tr>
<td>$</td>
<td>\mathbf{p}_T^{\text{miss}} + \mathbf{p}_T^1</td>
</tr>
<tr>
<td>$m_{\ell\ell}$</td>
<td>Numbers of $b$-tagged and non-$b$-tagged jets</td>
</tr>
<tr>
<td>$H_{\ell\ell}$</td>
<td>Scalar sum of the magnitudes of the momenta of the $H \rightarrow \ell \ell \nu \nu$ and $H \rightarrow bb$ systems,</td>
</tr>
<tr>
<td>$H_{\ell\ell}^b$</td>
<td>$H_{\ell\ell} = (</td>
</tr>
</tbody>
</table>

The resulting four outputs, each constrained to values between 0 and 1, are referred to as $p_1$ ($i \in \{HH, \text{Top}, Z-\ell, Z-\tau\}$). Values of $p_1$ near to 1 indicate that the event likely belongs to class $i$ and values nearer to 0 indicate otherwise. The main discriminant in the analysis, $d_{HH}$, is constructed from the four $p_1$ and is defined as

$d_{HH} = \ln [P_{HH}/(P_{\text{Top}} + P_{Z-\ell} + P_{Z-\tau})]$. 

The $HH \rightarrow bb\ell\ell\nu\nu$ signal events are characterised by two distinct ‘Higgs hemispheres’. One hemisphere contains the two $b$-tagged jets from the $H \rightarrow bb$ decay and it is typically opposite in the transverse plane to the second hemisphere that contains the two leptons and $E_{\text{miss}}$ from the $H \rightarrow WW^*/ZZ^*/\tau\tau$ decay. The final-state objects in the SM backgrounds, the Top process in particular, are distributed more uniformly within the event and they typically do not exhibit the same opposite hemispheres topology as the $HH$ signal. These Higgs hemispheres thus provide a topological criterion that distinguishes the signal from the background and motivates the choice of input observables that are provided to the classifier. Thirty-five such variables are provided as inputs to the classifier, ranging from momentum components of the visible final-state objects to observables using event-wide information, and are constructed using only calibrated final state objects that have well-understood uncertainties (Section 6). A complete list is provided in Table 2. The event-wide input observables are sensitive to the presence of Higgs hemispheres in the signal and are largely angular in nature or take advantage of the fact that the final state objects from each of the Higgs bosons in the signal tend to be near to each other. The observables $H_{\ell\ell}^b$ and $m_{\ell\ell}^{bb}$ are non-standard high-level observables that are not straightforward functions of the momenta of the final-state objects. By construction, $H_{\ell\ell}^b$ can take values between zero and one; it peaks near one for signal and is more broadly distributed for background. The $m_{\ell\ell}^{bb}$ observable is defined similarly to the $m_{\ell\ell}^{bb}$ observable in Ref. [119] but does not include the final-state leptons. As discussed in Ref. [119], for the Top backgrounds $m_{\ell\ell}^{bb}$ generally has values below the mass of the top-quark due to kinematic constraints while for the $Z/\gamma^*$ processes, which have little-to-no $E_{\text{miss}}$, $m_{\ell\ell}^{bb}$ is typically below 45 GeV. The use of dropout regularisation during the training of the classifier allows it to more effectively use the information contained in the full set of inputs presented in Table 2 by reducing its susceptibility to overtraining effects that may otherwise appear as a result of using such an extended input feature space in the case where no such regularisation is performed. To verify this, the performance of the classifier was checked using an independent sample of events not used in the training of the classifier and was found to be comparable to its performance when presented with those of the training sample.

5.2. Signal selection criteria

To define signal selection criteria, the analysis relies on the invariant mass of the two leptons, $m_{\ell\ell}$, and the invariant mass of the two leading ($p_T$-ordered) $b$-tagged jets, $m_{bb}$. Due to spin-correlation effects present in the $H \rightarrow WW^*$ decay within the dominant $HH \rightarrow bbWW^*$ signal process, the signal events exhibit values of $m_{\ell\ell}$ that are typically below 60 GeV. By selecting low values of $m_{\ell\ell}$, the signal purity can therefore be enhanced while rejecting a large component of the SM Z boson and Top backgrounds. Additionally, $m_{bb}$ has a peak at the mass of the Higgs boson for the signal process and therefore provides an effective means to define selections in which the $HH$ contribution is enhanced. The signal selection criteria therefore require $m_{\ell\ell} \in (20, 60)$ GeV and $m_{bb} \in (110, 140)$ GeV. The $m_{\ell\ell}$ > 20 GeV requirement is enforced in order to remove contamination from low-mass resonances and $Z/\gamma^*$ processes. The signal selection criteria are further broken down into same-flavour (SF), i.e. ee or $\mu\mu$, or different-flavour (DF), i.e. $e\mu$, regions. Separating by dilepton flavour enhances the separation power between the signal and $Z/\gamma^*$ background; the former has roughly equal probabilities for the SF and DF final states and the latter leads predominantly to SF final states.

In addition to the $m_{\ell\ell}$ and $m_{bb}$ requirements, the same-flavour and different-flavour signal regions, SR-SF and SR-DF, respectively, are defined by requiring high values of $d_{HH}$ and are presented in Table 3. The chosen threshold values of $d_{HH} > 5.45 (5.55)$ for SR-SF (SR-DF) are found to maximise the expected sensitivity to the non-resonant $HH$ process. The predicted $HH \rightarrow bb\ell\ell\nu\nu$ signal yields in SR-SF and SR-DF are shown in Table 3, and are composed of 90% $HH \rightarrow bbWW^*$, 9% $HH \rightarrow bb\tau\tau$, and 1% $HH \rightarrow bbZZ^*$. The predominance of the $HH \rightarrow bbWW^*$ process over the other two is a result of both its larger overall BF and of the classifier having been trained only on this component of the signal.

5.3. Background estimation

As mentioned in Section 5.1, the dominant backgrounds expected to contaminate the signal regions are the Top and $Z/\gamma^*$ processes, specifically $Z/\gamma^*$ production in association with jets.
originating from heavy-flavour hadrons ($bb$, $bc$, or $cc$), subsequently referred to as $Z/γ^*+HF$. Subdominant SM processes contribute via $tt$ production in association with an electroweak vector boson, single Higgs boson production (predominantly via the $t\bar{t}H$ mode), $Z/γ^*$ production in association with light-flavour jets, and electroweak diboson processes. There is additionally a minor contribution of background events from non-prompt leptons produced in semileptonic decays of heavy-flavour hadrons and from misidentified electron candidates arising from photon conversions and jets. This background is estimated using events with a same-charge lepton pair, following procedures described in Ref. [121], after subtracting the prompt-lepton contribution. The rest of the SM background processes detailed in Table 1 are estimated primarily using simulation.

Dedicated control regions are defined to derive data-driven normalisation corrections for the dominant background processes: CR-Top for Top and CR-Z+HF for $Z/γ^*+HF$. These normalisation corrections have a uniform prior and are checked in two validation regions, VR-1 and VR-2, enriched with events from the Top and $Z/γ^*$+HF processes. The control and validation regions are defined in Table 3 and are kinematically close to the signal regions. CR-Top (CR-Z+HF) and VR-1 (VR-2) are defined by inverting the $m_{bb}$ ($m_{ll}$) requirements relative to those of the signal regions but retain a selection of the high $d_{HH}$ region similar to the signal regions. The $d_{HH}$ selections were relaxed to increase statistical power, independent checks showed that this did not have a significant impact on the post-fit normalisation corrections in Table 3.

VR-1 keeps only those events with $m_{bb} > 140$ GeV, excluding the region $m_{bb} < 100$ GeV which is included in CR-Top, due to significant contamination of $Z/γ^*$+HF events. The correlations between the $m_{ll}$ and $m_{bb}$ observables and $d_{HH}$ after the preselection are observed to be small and do not prevent the use of the former two in the construction of the analysis regions defined in Table 3, as $d_{HH}$ is found to rely mainly on the information provided by the additional input observables listed in Table 2. This absence of strong correlation ensures that the measurements made in the tails of $d_{HH}$ in the control regions can be extrapolated to those in the signal regions.

The Top background in the signal regions is expected to be composed of approximately equal contributions from the $tt$ and single-top-quark $Wt$ process and therefore susceptible to the interference effects as described in Section 3. For this reason, CR-Top and the validation regions are defined so that they have predicted $tt$ and $Wt$ compositions similar to that of the signal regions. This ensures that the normalisation correction determined in the fit for the Top background results in an accurate estimate of the combined $tt$ and $Wt$ process in the signal regions, accounting for potential interference effects present in data but not necessarily modelled in MC simulation. Table 3 compares the observed and predicted event yields, where the background event yields obtained after background-only fits in the corresponding control regions are also shown. The post-fit normalisation correction factors for the Top and $Z/γ^*$+HF background processes, respectively $\mu_{Top} = 0.79 \pm 0.10$ and $\mu_{Z/γ^*+HF} = 1.36 \pm 0.07$, are also shown in Table 3. The uncertainties in $\mu_{Top}$ and $\mu_{Z/γ^*+HF}$ take into account the statistical and systematic uncertainties due to the experimental sources, as described in Section 6.

Distributions of $d_{HH}$ in the control regions after performing background-only fits to data in the control regions and applying the Top and $Z/γ^*$+HF normalisation corrections are shown in Fig. 2. In the control and validation regions, good agreement between the data and SM prediction provided by the post-fit MC simulation is observed for the observables relevant to the analysis.

6. Systematic uncertainties

The analysis evaluates several sources of systematic uncertainty for the signal and background processes, which are classified as either experimental (detector or luminosity related) or theoretical modelling uncertainties. Statistical uncertainties of the simulated event samples are also taken into account. The main uncertainty components are summarised in Table 4. MC modelling uncertainties in the Top and $Z/γ^*$+HF background estimates are dominant, followed by statistical and detector uncertainties.

The normalisation corrections of the Top and $Z/γ^*$+HF background processes are determined primarily by the data events in the control regions when performing the statistical analysis. These corrections take into account the statistical and systematic uncertainties due to the experimental sources, as described later in this section. In addition, the systematic uncertainties in the theoretical modelling of these processes are applied as uncertainties in
the corrected predictions in the signal regions using the following procedures. The uncertainties in the estimated Top background event yields due to parton shower modelling are assessed as the difference between the predictions of Powheg-Box showered with Pythia or Herwig, and those due to the choice of event generator are assessed by comparing the predictions of Powheg-Box or MadGraph 5. aMC@NLO [122], both showered with Pythia. The uncertainties due to missing higher-order corrections are estimated by changing the renormalisation and factorisation scales ($\mu_R$ and $\mu_F$, respectively) up and down by a factor of two (8-point variation). The uncertainties due to the modelling of initial- and final-state radiation (ISR and FSR, respectively) in the generators used to simulate the Top background processes are evaluated using the method described in Ref. [122]. The Top background composition is varied within the uncertainties in the theoretical predictions for the $t\bar{t}$ and single-top-quark $Wt$ cross-sections [65,68,123]. The uncertainty arising from the interference between the NLO predictions for $t\bar{t}$ and $Wt$ processes is estimated by taking the difference between the predicted Top background yields obtained with the DR and DS schemes used for the NLO $Wt$ calculation [122]. The uncertainties due to PDF variations are computed as the envelope of the central values of the nominal NNPDF3.0 PDF set and the CT14, MMHT14, and PDF4LHC15_30PDF PDF sets [124]. All uncertainties except those in the scale variations, cross-section, and interference are considered as fully correlated between the $t\bar{t}$ and $Wt$ processes. The $Z/\gamma^* +$ HF modelling uncertainties are estimated using the nominal Sherpa 2.2.1 samples by considering different merging (CKKW-L) [125] and resummation scales. The uncertainties due to PDF variations and changes in $\mu_R$ and $\mu_F$ are calculated using the same procedures as for the Top backgrounds. An additional uncertainty in the $Z/\gamma^*$ process is computed by taking the difference between the nominal Sherpa 2.2.1 samples with samples generated using MadGraph 5. aMC@NLO+Pythia8. The dominant uncertainties in the total background estimates in SR-SF are the $Z/\gamma^* +$ HF modelling uncertainties (8%), primarily due to comparison of Sherpa 2.2.1 and MadGraph 5. aMC@NLO, and the parton shower uncertainty affecting the Top background process (5%). The uncertainties in the background estimates in SR-DF are dominated by the uncertainty due to the parton shower affecting the Top background process (12%), the uncertainty in the Top normalisation correction $\mu_{\text{Top}}$ (10%), the uncertainty due to the comparison between the generators used for the Top process (7.5%), and the uncertainty due to the modelling of ISR and FSR in the Top process (5%).

Systematic uncertainties in the signal acceptance due to varying $\mu_R$ and $\mu_F$, as well as PDF-induced uncertainties, are evaluated using the same procedure as for the Top background process. The resulting scale (PDF) uncertainties are $<3\%$ ($<1\%$) in both signal regions. The uncertainty due to the parton shower modelling is computed by comparing Herwig7 with Pythia8, and is found to be 8% (9%) in SR-SF (SR-DF). The uncertainty in the $HH$ production cross-section, evaluated to be 5%, is included as an uncertainty in $\sigma_{\text{SM}}(gg \rightarrow HH)$ when computing the upper limits on the cross-section ratio in Table 5. This value is the quadrature sum of the scale, PDF, and top mass contributions as reported by the LHCSWG [20].

The uncertainties due to experimental sources arise primarily from the mismeasurement of reconstructed object momenta and from the mismodelling of reconstruction efficiencies. These uncertainties include uncertainties from the mismodelling of the jet energy scale [109] and jet energy resolution (JER) [126]. Additional uncertainties for $b$-tagged jets arise from the mismodelling of the $b$-tagging efficiency [111] and from the mismodelling of the rates at which charm- and light-flavoured jets are selected as $b$-tagged jets [127,128]. Lepton-related uncertainties arise from the mismodelling of the electron [104] (muon [105]) reconstructed energy (momentum) measurements, as well as in the mismodelling of their reconstruction and identification efficiencies [104, 105]. The $E_{\text{miss}}$ scale and resolution [113] uncertainties, as well as uncertainties from the mismodelling of pile-up, trigger efficiency and luminosity, are also taken into account. The uncertainty in the combined 2015–2018 integrated luminosity is 1.7% [129], obtained using the LUCID-2 detector [130] for the primary luminosity measurements. The combined effect of the experimental sources of systematic uncertainty in the predicted background yields is summarised in Table 4 and is dominated by the JER, with all other contributions found to be negligible.

7. Results

In order to extract information about the $HH \rightarrow b\bar{b}t\bar{t}v\bar{v}$ signal cross-section, a counting experiment is performed with a profile-likelihood fit [131] simultaneously across the CR-Top, CR-Z+HF, SR-SF, and SR-DF regions using the predicted and observed event counts in each region as inputs. The Top and $Z/\gamma^* +$ HF normal-

---

**Fig. 2.** Distributions of $d_{\text{HF}}$ in CR-Top (left) and CR-Z+HF (right). Distributions are shown after the fit to data under the background-only hypothesis in the shown control regions. The ratio of the data to the sum of the backgrounds is shown in the lower panel. The hatched bands indicate the combined statistical and systematic uncertainty.
Table 4
Breakdown of the main uncertainty components in the background estimates in the two signal regions for the Top, $Z/\gamma^*+HF$, and all other ("Other") backgrounds. The uncertainty components associated with the total background estimate in the signal regions (the sum of Top, $Z/\gamma^*+HF$, and Other) is listed under "Total Bkg.". As in the upper half of Table 3, all uncertainties are shown "post-fit". The percentages show the size of the uncertainty relative to the expected background in each column and uncertainties can be correlated, not necessarily adding in quadrature to the total uncertainty in each column or across each row. Uncertainties in the post-fit normalisation factors, $\mu_{\text{Top}}$ and $\mu_{Z/\gamma^*+HF}$, are only applicable for the Top and $Z/\gamma^*+HF$ processes.

<table>
<thead>
<tr>
<th>Uncertainty [%]</th>
<th>SR-SF</th>
<th></th>
<th>SR-DF</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Top</td>
<td>$Z/\gamma^*+HF$</td>
<td>Other</td>
</tr>
<tr>
<td>Total uncertainty</td>
<td>28</td>
<td>18</td>
<td>20</td>
</tr>
<tr>
<td>Theoretical</td>
<td>21</td>
<td>15</td>
<td>17</td>
</tr>
<tr>
<td>Experimental</td>
<td>12</td>
<td>$&lt;5$</td>
<td>8</td>
</tr>
<tr>
<td>MC statistics</td>
<td>8</td>
<td>8</td>
<td>6</td>
</tr>
<tr>
<td>$\mu_{\text{Top}}$, $\mu_{Z/\gamma^*+HF}$</td>
<td>13</td>
<td>5</td>
<td>n/a</td>
</tr>
</tbody>
</table>

Fig. 3. Distributions of $m_{bb}$ (left), $m_{ll}$ (middle), and the discriminant $d_{bH}$ (right). The distributions are shown after the fit to data in the control regions under the background-only hypothesis. Each distribution includes both the SF and DF events and imposes signal selection requirements on all quantities except the one being plotted, but the requirement on $d_{bH}$ has been relaxed to $d_{bH} > 5$ for the distributions of $m_{bb}$ and $m_{ll}$. The $HH \rightarrow bbb\ell$ signal ("HH") is overlaid and has its cross-section scaled by a factor of 20 relative to the SM prediction for visualisation purposes. The ratio of the data to the sum of the backgrounds is shown in the lower panel of each figure. The hatched bands indicate the combined statistical and systematic uncertainty.

Table 5
Observed and expected upper limits on the ggF-initiated non-resonant $HH$ production cross-section at 95% CL and their ratios to the SM prediction ($\sigma_{SM}(gg \rightarrow HH) = 31.05 \pm 1.90 \text{ fb}$ [13-20]). The $\pm 1\sigma$ and $\pm 2\sigma$ variations about the expected limit are also shown. Uncertainties in the SM cross-section are taken into account when computing the upper limits on the cross-section ratio.

<table>
<thead>
<tr>
<th>$\sigma (gg \rightarrow HH)$ [pb]</th>
<th>$\sigma_{SM} (gg \rightarrow HH)$</th>
<th>95% CL upper limit [pb]</th>
<th>95% CL upper limit [pb]</th>
</tr>
</thead>
<tbody>
<tr>
<td>0.5</td>
<td>14</td>
<td>0.9</td>
<td>13</td>
</tr>
<tr>
<td>0.6</td>
<td>20</td>
<td>0.9</td>
<td>19</td>
</tr>
<tr>
<td>0.9</td>
<td>29</td>
<td>1.3</td>
<td>1.9</td>
</tr>
<tr>
<td>1.3</td>
<td>43</td>
<td>1.9</td>
<td>1.2</td>
</tr>
<tr>
<td>1.9</td>
<td>62</td>
<td>1.2</td>
<td>40</td>
</tr>
<tr>
<td>1.2</td>
<td>40</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

8. Conclusions

A search for non-resonant Higgs boson pair production, as predicted by the SM, is presented in the final state with at least two $b$-tagged jets and exactly two leptons with opposite electric charge, where one of the Higgs bosons decays to $bb$ and the other decays to either $WW^*$, $ZZ^*$, or $\tau\tau$. The analysis uses $pp$ collision data recorded at $\sqrt{s}=13$ TeV by the ATLAS detector at the LHC, corresponding to an integrated luminosity of 139 fb$^{-1}$. The data are in agreement with the predictions for the SM background processes. An observed (expected) 95% CL upper limit is set on the cross-section for the production of Higgs boson pairs, corresponding to 40 (29) times the SM prediction. These limits are comparable to the previous leading searches for non-resonant Higgs boson pair production performed by the ATLAS and CMS experiments.

isation corrections are also extracted from this fit and are found to differ negligibly from those presented in Table 3. All sources of systematic and statistical uncertainty in the signal and background models are implemented as deviations from the nominal model, scaled by nuisance parameters that are profiled in the fit. The $p$-value corresponding to the background-only hypothesis, giving the probability that the data in the signal regions be at least as incompatible with the background-only hypothesis as that observed in SR-SF and SR-DF, is $p_0 = 0.15$ and corresponds to a $1.05\sigma$ significance. Distributions of $m_{bb}$, $m_{ll}$, and $d_{bH}$ after performing background-only fits to data in the control regions and applying the Top and $Z/\gamma^*+HF$ normalisation corrections are shown in Fig. 3. The signal selection criteria are imposed on all observables shown in Fig. 3 apart from the one being plotted, except that the $d_{bH}$ requirement for the $m_{bb}$ and $m_{ll}$ distributions is relaxed to $d_{bH} > 5$. No significant excess of events over the expected SM background is observed and upper limits are set on non-resonant Higgs boson pair production at 95% confidence level (CL) using the CLs method [132]. Table 5 presents these upper limits and comparisons with the SM prediction. The observed (expected) limit at 95% CL is 1.2 (0.9) pb, corresponding to 40 (29) times the SM prediction.
The ATLAS Collaboration / Physics Letters B 801 (2020) 135145


[119] CMS Collaboration, Measurement of the top quark mass in the dileptonic $t\bar{t}$ decay channel using the mass observables $M_{\ell\ell}$, $M_{T}$, and $M_{b\bar{b}}$, in $pp$ collisions at $\sqrt{s} = 8$ TeV, Phys. Rev. D 96 (2017) 032002, arXiv:1704.06142 [hep-ex].


