Search for a heavy Higgs boson decaying into a $Z$ boson and another heavy Higgs boson in the $\ell\ell bb$ final state in $pp$ collisions at $\sqrt{s} = 13$ TeV with the ATLAS detector

**The ATLAS Collaboration**

**A R T I C L E I N F O**

Article history:
Received 4 April 2018
Received in revised form 15 June 2018
Accepted 4 July 2018
Available online 11 July 2018
Editor: W.-D. Schlatter

**A B S T R A C T**

A search for a heavy neutral Higgs boson, $A$, decaying into a $Z$ boson and another heavy Higgs boson, $H$, is performed using a data sample corresponding to an integrated luminosity of 36.1 fb$^{-1}$ from proton–proton collisions at $\sqrt{s} = 13$ TeV recorded in 2015 and 2016 by the ATLAS detector at the Large Hadron Collider. The search considers the $Z$ boson decaying to electrons or muons and the $H$ boson into a pair of $b$-quarks. No evidence for the production of an $A$ boson is found. Considering each production process separately, the 95% confidence-level upper limits on the $pp \rightarrow A \rightarrow ZH$ production cross-section times the branching ratio $H \rightarrow bb$ are in the range of 14–830 fb for the gluon–gluon fusion process and 26–570 fb for the $b$-associated process for the mass ranges 130–700 GeV of the $H$ boson and 230–800 GeV of the $A$ boson. The results are interpreted in the context of two-Higgs-doublet models.

© 2018 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

After the discovery of a Higgs boson at the Large Hadron Collider (LHC) [1,2], one of the most important remaining questions is whether the recently discovered particle is part of an extended scalar sector or not. Additional Higgs bosons appear in all models with an extended scalar sector, such as the two-Higgs-doublet model (2HDM) [3,4]. Such extensions are motivated by, and included in, several new physics scenarios, such as supersymmetry [5], dark matter [6] and axion [7] models, electroweak baryogenesis [8] and neutrino mass models [9].

The addition of a second Higgs doublet leads to five Higgs bosons after electroweak symmetry breaking. The phenomenology of such a model is very rich and depends on many parameters, such as the ratio of the vacuum expectation values of the two Higgs doublets ($\tan\beta$), and the Yukawa couplings of the scalar sector [4]. When CP conservation is assumed, the model contains two CP-even Higgs bosons, $h$ and $H$ with $m_h < m_H$, one CP-odd, $A$, and two charged scalars, $H^\pm$. There have been many searches for the heavy neutral Higgs bosons of the 2HDM at the LHC, including $H \rightarrow WW/ZZ$ [10–13], $A/H \rightarrow \tau\tau/bb$ [14–16], $A \rightarrow Zh$ [17, 18] and $H \rightarrow hh$ [19,20]. For the interpretation of these searches it is usually assumed that the heavy Higgs bosons, $H$ and $A$, are degenerate in mass, i.e. $m_A = m_H$.

This assumption of mass degeneracy is relaxed in this Letter by assuming $m_A > m_H$. Such a case is motivated by electroweak baryogenesis scenarios in the context of the 2HDM [21–24]. For 2HDM electroweak baryogenesis to occur, the requirement $m_A > m_H$ is favoured [21] for a strong first-order phase transition to take place in the early universe. The $A$ boson mass is also bounded from above to be less than approximately 800 GeV, whereas the lighter CP-even Higgs boson, $h$, is required to have properties similar to those of a Standard Model (SM) Higgs boson and is assumed to be the Higgs boson with mass of 125 GeV that was discovered at the LHC [21]. Under such conditions and for large parts of the 2HDM parameter space, the CP-odd Higgs boson, $A$, decays into $ZH$ [25,21]. The production of the $A$ boson in the relevant 2HDM parameter space proceeds mainly through gluon–gluon fusion and $b$-associated production at the LHC.

This search for $A \rightarrow ZH$ decays uses proton–proton collision data at $\sqrt{s} = 13$ TeV corresponding to an integrated luminosity of 36.1 fb$^{-1}$ recorded by the ATLAS detector at the LHC. The search considers only $Z \rightarrow \ell\ell$, where $\ell = e, \mu$, to take advantage of the clean leptonic final state, and $H \rightarrow bb$, because of its large branching ratio. This final state allows full reconstruction of the $A$ boson’s decay kinematics. The reconstruction of the $A$ boson’s invariant mass uses the assumed value of the mass of the $H$ boson to improve its resolution. The final state is also categorised by the presence of two or three $b$-tagged jets to take advantage of the $b$-associated production mechanism. The CMS Collaboration has published a similar search at $\sqrt{s} = 8$ TeV [26]. This Letter reports the result of a search at $\sqrt{s} = 13$ TeV, which extends the
previous search by considering explicitly the gluon–gluon fusion and \( b \)-associated production processes as well as both narrow and wide widths of the \( A \) boson.

2. ATLAS detector

The ATLAS detector is a general-purpose particle detector, described in detail in Ref. [27]. It includes an inner detector surrounded by a 2 T superconducting solenoid, electromagnetic and hadronic calorimeters and a muon spectrometer with a toroidal magnetic field. The inner detector consists of a high-granularity silicon pixel detector, including the insertable B-layer [28] installed in 2014, a silicon microstrip detector, and a straw-tube tracker. It provides precision tracking of charged particles with pseudorapidity \( |\eta| < 2.5 \). The calorimeter system covers the pseudorapidity range \( |\eta| < 4.9 \). It is composed of sampling calorimeters with either liquid argon or scintillator tiles as the active medium. The muon spectrometer provides muon identification and measurement for \( |\eta| < 2.7 \). A two-level trigger system [29] is employed to select events for offline analysis, which reduced the average recorded collision rate to about 1 kHz.

3. Data and simulation

The data used in this search were collected during 2015 and 2016 from \( \sqrt{s} = 13 \) TeV proton–proton collisions and correspond to an integrated luminosity of 36.1 fb\(^{-1}\), which includes only data-taking periods where all relevant detector subsystems were operational. The data sample was collected using a set of single-muon and single-electron triggers. The lowest-\( p_T \) trigger thresholds depend on the data-taking period and are in the range of 20–26 GeV for the single-muon triggers and 24–26 GeV for the single-electron triggers.

Simulated signal events with \( A \) bosons produced by gluon–gluon fusion were generated at leading order with MADGRAPH5_aMC@NLO 2.3.3 [30,31] using PYTHIA 8.210 [32] with a set of tuned parameters called the A14 tune [33] for parton showering. For the generation of \( A \) bosons produced in association with \( b \)-quarks, MADGRAPH5_aMC@NLO 2.1.2 [31,34,35] was used following Ref. [36] together with PYTHIA 8.212 and the A14 tune for parton showering. The gluon–gluon fusion production used NNPDF2.3LO [37] as the parton distribution functions (PDF), while the \( b \)-associated production used CT10nlo_m4 [38]. The signal samples were generated for \( A \) bosons with masses in the range of 230–800 GeV and widths up to 20% of the mass and for narrow-width \( H \) bosons with masses in the range of 130–700 GeV.

Background events from the production of \( W \) and \( Z \) bosons in association with jets were simulated with Sherpa 2.2.1 [39] using the NNPDF3.0NNLO PDF set [40]. Top-quark-pair production was simulated with PowHEG-Box v2 [41–43] and the CT10nlo PDF set [38], while the electroweak single-top-quark production was simulated with PowHEG-Box v1 and the fixed four-flavour PDF set CT10nlo_f4 [38]. The parton shower was performed with PYTHIA 6.428 [44] using the Perugia 2012 set of tuned parameters [45]. The production of top-quark pairs in association with a vector boson was simulated using MADGRAPH5_aMC@NLO 2.2.3 and the NNPDF3.0NNLO PDF set, whereas PYTHIA 8.186 was used for the parton shower with the A14 tune. Production of \( WW \), \( ZZ \) and \( WZ \) pairs was simulated using SHERPA 2.2.1 and the NNPDF3.0NNLO PDF set. Finally, SM Higgs boson production in association with a \( Z \) boson was generated with PowHEG-Box v2 and the NNPDF3.0NNLO PDF set, whereas the parton shower was performed with PYTHIA 8.186 using the AZNLO tune [46].

The modelling of bottom- and charm-hadron decays was performed with the EvtGen v1.2.0 package [47] for all samples apart from those simulated with Sherpa. The simulated events were overlaid with inelastic proton–proton collisions to account for the effect of multiple interactions occurring in the same and neighbouring bunch crossings (‘pile-up’). These events were generated using PYTHIA 8 with the A2 tune [48] and the MSTW2008LO PDF set [49]. The events were reweighted so that the distribution of the average number of interactions per bunch crossing agreed with the data.

All generated background samples were passed through the GEANT4-based [50] detector simulation [51] of the ATLAS detector. The ATLAST2 simulation [51] was used for the signal samples to allow for the generation of many different \( A \) and \( H \) boson masses. The simulated events were reconstructed in the same way as the data.

4. Object reconstruction

Electrons are reconstructed from energy clusters in the electromagnetic calorimeter that are matched to tracks in the inner detector [52]. Electrons are required to have \( |\eta| < 2.47 \) and \( p_T > 7 \) GeV. To distinguish electrons from jets, isolation and quality requirements are applied [53]. The isolation requirements (the ‘LooseTrackOnly’ working point) are defined by the \( p_T \) of tracks within cones around the electron with a size that decreases as a function of the transverse energy. The quality requirements (the ‘Loose’ working point) refer to both the inner detector track and the calorimeter shower shape. The efficiency for an electron to be reconstructed and meet these criteria is about 85% for electron \( p_T > 7 \) GeV and increases to about 90% for \( p_T > 27 \) GeV.

Muons are reconstructed by matching tracks reconstructed in the inner detector to tracks or track segments in the muon spectrometer [54]. Muons used for this search must have \( |\eta| < 2.5 \) and \( p_T > 7 \) GeV, and are required to satisfy ‘LooseTrackOnly’ isolation requirements, similar to those used for electrons, as well as inner detector and muon spectrometer track ‘Loose’ quality criteria, corresponding to an efficiency of about 97%.

Jets are reconstructed using the anti-\( k_t \) algorithm [55,56] with radius parameter \( R = 0.4 \) from clusters of energy deposits in the calorimeter system [57]. Candidate jets are required to have \( p_T > 20 \) GeV (\( p_T > 30 \) GeV) for \( |\eta| < 2.5 \) (2.5 < \( |\eta| < 4.5 \)). Low-\( p_T \) jets from pile-up are rejected by a multivariate algorithm that uses properties of the reconstructed tracks in the event [58].

Jets containing \( b \)-hadrons are selected using a multivariate tagging algorithm (\( b \)-tagging) [59,60]. The energy of the tagged jet (\( b \)-jet) is corrected for the average energy loss from semileptonic decays of \( b \)-hadrons and out-of-jet-cone tracks with large impact parameters [61]. The \( b \)-tagging efficiency for the jet \( p_T \) range used in this analysis is between 65% and 75%. Applying the \( b \)-tagging algorithm reduces the number of light-flavour (\( c \)-quark) jets by a factor of 250–550 (10–20), depending on the jet kinematics.

When electrons, muons and jets are spatially close, these algorithms can lead to ambiguous identifications. An overlap removal procedure [61] is therefore applied to uniquely identify these objects.

The missing transverse momentum, \( E_T^{\text{miss}} \), is computed using reconstructed and calibrated leptons, photons and jets [62].
from the primary vertex which are not associated with any identified lepton or jet are also taken into account in the \( E_T \) reconstruction \[63\].

5. Event selection

The decay \( A \rightarrow ZH \rightarrow \ell\ell b \bar{b} \) features a pair of oppositely charged, same flavour leptons and two b-jets. Three resonances can be formed by combining the selected objects: the \( Z \) boson (\( \ell\ell \)), the \( H \) boson (\( b\bar{b} \)) and the \( A \) boson (\( \ell\ell b\bar{b} \)). Moreover, additional b-jets may be present if the \( A \) boson is produced via the \( b \)-associated production mechanism. These features are used to define the event selection as summarised in Table 1.

The events recorded by the single-muon and the single-electron triggers are required to contain exactly two muons or two electrons, respectively. At least one of the leptons must have \( p_T > 27 \) GeV. Only events that contain a primary vertex with at least two associated tracks with \( p_T > 400 \) MeV \[64\] are considered. In the case of muons, they are required to have opposite electric charges. No such requirement is applied to electrons due to their non-negligible charge misidentification rates resulting from conversions of bremsstrahlung photons. The invariant mass of the lepton pair, \( m_\ell\ell \), must be in the range of 80–100 GeV to be compatible with the mass of the \( Z \) boson.

The \( H \rightarrow b\bar{b} \) decay is reconstructed by requiring at least two b-jets with the highest-\( p_T \) one having \( p_T > 45 \) GeV. When more than two b-jets are present, the two highest-\( p_T \) b-jets are considered to be from the \( H \) decay. Requiring b-jets increases the fraction of top-quark background in the signal region, including top-quark pair and single-top-quark production. This is reduced by requiring \( E_T^{\text{miss}} / \sqrt{H_T} < 3.5 \) GeV\(^{1/2} \), where \( H_T \) is the scalar sum of the \( p_T \) of all jets and leptons in the event. In addition, a requirement that reduces the Z+jets background is also applied: \( \sqrt{\sum p_T^2 / m_{\ell\ell b\bar{b}}} > 0.4 \), where \( m_{\ell\ell b\bar{b}} \) is the four-body invariant mass of the two-lepton, two-b-jet system assigned to the \( A \) boson and the summation is performed over the \( p_T \) of these objects.

Subsequently, two categories are defined: the \( n_\ell = 2 \) category, which contains events with exactly two b-jets, and the \( n_\ell = 3 \) category, which contains events with three or more b-jets. For the gluon–gluon fusion production, 94%–97% of the events passing the above selection fall into the \( n_\ell = 2 \) category, depending on the assumed \( m_A \) and \( m_H \). However for the \( A \)-associated production, 27%–36% fall into the \( n_\ell = 3 \) category. The remaining \( A \)-associated produced signal events are categorised as \( n_\ell = 2 \) events, even though more than two b-jets are expected, due to the relatively soft \( p_T \) spectrum of the associated b-jets and the geometric acceptance of the tracker.

Finally, the invariant mass of the two leading b-jets, \( m_{b\bar{b}} \), must be compatible with the assumed \( H \) boson mass by satisfying the requirement of \( 0.85 \cdot m_H - 20 \) GeV \( < m_{b\bar{b}} < m_H + 20 \) GeV for the \( n_\ell = 2 \) category, and \( 0.85 \cdot m_H - 25 \) GeV \( < m_{b\bar{b}} < m_H + 50 \) GeV for the \( n_\ell \geq 3 \) category. The wider window for \( n_\ell \geq 3 \) is motivated by a slightly degraded resolution due to potential b-jet mis-assignments (see later). The overall signal efficiency of the \( n_\ell = 2 \) category after this requirement is 5%–11% (3%–7%) for gluon–gluon fusion (b-associated production), depending on the \( m_A \) and \( m_H \) values. Similarly, the efficiency of the \( n_\ell \geq 3 \) category is 2%–4% for the b-associated production. The signal region selection is summarised in Table 1.

The \( m_{\ell\ell b\bar{b}} \) distribution after the \( m_{b\bar{b}} \) requirement is used to discriminate between signal and background. To improve the \( m_{\ell\ell b\bar{b}} \) resolution, the \( b\bar{b} \) system's four-momentum components are scaled to match the assumed \( H \) boson mass and the \( \ell\ell \) system's four-momentum components are scaled to match the \( Z \) boson mass. This procedure, performed after the event selection, improves the \( m_{\ell\ell b\bar{b}} \) resolution by a factor of two without significantly distorting the background distributions, resulting in an \( A \) boson mass resolution of 0.3%–4%.

The dominant backgrounds after these selections are from Z+jets and top-quark production. For top-quark-pair production, a very pure (\( > 99 \% \) of predicted events) control region is used to determine the normalisation of the background, whereas its shape in the signal region is taken from the simulation. This control region is defined by keeping the same selection as discussed previously, apart from an opposite-flavour lepton criterion, i.e., an opposite-charge \( e\mu \) pair is required instead of an \( ee \) or \( \mu\mu \) pair (see also Table 1). The shape of the Z+jets background distribution is obtained from simulation and the normalisation is extracted from data together with the signal (see also Section 7). This procedure is possible because of the very different shapes of the \( m_{\ell\ell b\bar{b}} \) distributions from signal and Z+jets events. The normalisation of the Z+jets production is further constrained by a control region defined by inverting the \( m_{b\bar{b}} \) window criterion for each \( H \) boson mass hypothesis (see also Table 1). The control regions are distinct for the \( n_\ell = 2 \) and the \( n_\ell \geq 3 \) categories, since the accuracy of the background simulation depends on the number of b-jets present in the event. Backgrounds from diboson, single top, and Higgs boson production, as well as top-quark-pair production in association with a vector boson, give a typical contribution of ~5% to the total background. Their shapes are taken from simulation, whereas they are normalised using precise inclusive cross-sections calculated from theory. The diboson samples are normalised using next-to-next-to-leading-order (NNLO) cross-sections \[65–68\]. Single-top-quark production and top-quark-pair production in association with vector bosons are normalised to next-to-leading-order (NLO)
cross-sections from Refs. [69–71] and Ref. [31], respectively. The normalisation of the Higgs boson production in association with a vector boson follows the recommendations of Ref. [36] using NNLO QCD and NLO electroweak corrections.

6. Signal modelling

The good $m_{t\tau b}$ mass resolution together with the fact that theory models often predict A bosons with large widths inflates the number of signal mass and width hypotheses that need to be considered. For this reason, the $m_{t\tau b}$ distributions are taken from simulation of a limited number of $(m_A, m_H)$ mass points and an interpolation using analytical functions is employed for the rest.

The $m_{t\tau b}$ distributions for A bosons produced by gluon–gluon fusion and with negligible widths compared with the experimental resolution are found to be adequately described by the \( \text{ExpGauss-Exp} \) (EGE) function [72]:

$$f_{\text{EGE}}(m; a, \sigma, k_L, k_H) = \begin{cases} e^{\frac{1}{2}k_L^2 + k_H(n_{\sigma}^2)} & \text{for } m - a \sigma \leq -k_L \\ e^{-\frac{1}{2}(\frac{m-a}{\sigma})^2} & \text{for } -k_L < m - a \sigma \leq k_H \\ e^{\frac{1}{2}k_H^2 - k_L(n_{\sigma}^2)} & \text{for } m - a \sigma > k_H \end{cases}$$

On the other hand, $m_{t\tau b}$ distributions for A bosons from $b$-associated production, also with negligible widths compared with the experimental resolution, are better described by a double-Gaussian Crystal Ball (DSCB) function [73]:

$$f_{\text{DSCB}}(m; a, \sigma, k_L, k_H, n_1, n_2) = g(m; a, \sigma, k_L, n_1) \cdot e^{-\frac{1}{2}k_L^2}$$

where $g(m; a, \sigma, k, n) = [(k/|n|)(n - |k| + (m - a)/\sigma)]^{-n}$. Both functions consist of a Gaussian core with mean $a$ and variance $\sigma^2$, whereas the rest of the parameters ($k_L$, $k_H$, $n_1$, $n_2$) describe the tails. The DSCB function describes better than the EGE function the slightly longer tails of the mass distribution of the $b$-associated production compared to gluon–gluon fusion. This is due to the few cases in which the $b$-quark produced in association with the Higgs boson is taken to be one of the $b$-quarks from the Higgs boson decay. The values of the function parameters are extracted from unbinned maximum-likelihood fits to the simulated $m_{t\tau b}$ distributions. The core mean, $a$, is parameterised using a linear function of $m_A$. The core width, $\sigma$, is observed to monotonically increase with $\Delta m = m_A - m_H$ and is parameterised with a third-degree polynomial. The rest of the parameters are largely constant and are fixed to their average values from the fits, with the exception of mass points with $\Delta m = 100$ GeV. The distributions at mass points with $\Delta m = 100$ GeV correspond to the smallest mass splitting considered in this search and are close to the kinematic cutoff. Their non-core parameters are fixed to the average fit values obtained from signal samples with this mass splitting only. As an example of the performance of this procedure, Fig. 1 shows a comparison for the $(m_A, m_H) = (500, 250)$ GeV mass point between the simulated distributions and the parametric functions described previously. The cores of the $m_{t\tau b}$ distributions are well-parameterised by the chosen functional forms. The small differences seen in the tails of some distributions between the functional forms and the simulations have only negligible effects on the final results, and moreover they are included as a source of systematic uncertainty.

The previously described parameterisation applies to signal samples generated with narrow-width $A$ bosons. In some regions of 2HDM parameter space relevant to this analysis, the $A$ boson’s width is significant compared with the detector resolution while the $H$ boson’s width remains negligible. In order to model the $m_{t\tau b}$ shape of $A$ bosons with large natural widths, a modified Breit–Wigner distribution\(^1\) is convolved with the EGE and DSCB functions. The procedure is validated by comparing the results of the convolution with those of the simulated samples of $A$ bosons with large natural widths. Widths of up to 20% of the $A$ boson mass are considered. An example of signal distributions with large natural widths is shown in Fig. 2 for the same signal points used in Fig. 1.

---

\(^1\) The modification is the multiplication of the Breit–Wigner distribution with a log-normal distribution to account for the distortion due to the event selection.
Finally, the signal efficiencies for the interpolated mass points are obtained through separate two-dimensional interpolations on the \((m_A, m_H)\) plane using thin plate splines \[74\].

7. Fit model and systematic uncertainties

The \(m_{\ell\ell\ell\ell}\) distribution is expected to exhibit a resonant structure if signal events are present, while background events result in a smooth shape. Therefore \(m_{\ell\ell\ell\ell}\) is chosen as the final signal and background discriminating variable. The shape differences in the \(m_{\ell\ell\ell\ell}\) distribution between the signal and background contributions are exploited through binned maximum-likelihood fits of the signal-plus-background hypotheses to extract potential signal contributions. The fits are based on the statistical framework described in Refs. \[75\]–\[77\]. For a given mass hypothesis \((m_A, m_H)\), the likelihood is constructed as the product of Poisson statistics in \(m_{\ell\ell\ell\ell}\) bins:

\[
L(\mu, \alpha, \beta | m_A, m_H) = \prod_{i=bins} \text{Poisson} \left( N_i \left( \mu \times S_i(m_A, m_H, \beta) + B_i(\alpha, \beta) \right) \right) \cdot G(\beta).
\]

Here \(N_i\) is the number of observed events and \(S_i(m_A, m_H, \beta)\) and \(B_i(\alpha, \beta)\) are the expected number of signal and estimated background events in bin \(i\). The vector \(\alpha\) represents free background normalisation scale factors (described later) and vector \(\beta\) denotes all non-explicitly listed parameters of the likelihood function such as nuisance parameters associated with systematic uncertainties. The function \(G(\beta)\) represents constraints on \(\beta\). The parameter of interest, \(\mu\), is a multiplicative factor to the expected signal rate and is called the signal-strength parameter. The \(m_{\ell\ell\ell\ell}\) bin widths are chosen according to the expected detector resolution and taking into account the statistical uncertainty related to the number of background Monte Carlo events. The bin centres are adjusted such that at least 68% of the test signal is contained in one bin. Only the \(n_b = 2\) category is considered for the gluon–gluon fusion production while both the \(n_b = 2\) and \(n_b \geq 3\) categories are included in the likelihood calculation for the \(b\)-associated production.

For each bin, \(S_i\) is calculated from the total integrated luminosity, the theoretical cross-section for the signal and its selection efficiency. The sum of all background contributions in the bin, \(B_i\), is estimated from simulation. However, the \(tt\) and \(Z\)-jets control regions are included in the likelihood calculation as one bin each, to help constrain their respective contributions in the signal regions. This is achieved by introducing two free normalisation scale factors, represented by \(\alpha\), for each category for the two most relevant background contributions: one for \(tt\) and the other for the heavy-flavour component \((Z+h.f.)\) of the \(Z\)-jets contribution. These scale factors are applied to their respective contributions estimated from the simulations and their values are determined from the fit. Typical values of scale factors are close to unity. Taking \((m_A, m_H) = (700, 200)\) GeV as an example, the \(Z+h.f.\) scale factor...
is 1.12 ± 0.09 for the \( n_b = 2 \) category and 1.1 ± 0.2 for the \( n_b \geq 3 \) category. Similarly, the \( \ell \ell \) scale factors are 0.96 ± 0.06 and 1.2 ± 0.2 for the two corresponding categories.

Systematic uncertainties are incorporated in the likelihood as nuisance parameters with either Gaussian or log-normal constraint terms. They include both the experimental and theoretical sources of uncertainty. Experimental uncertainties comprise those in the luminosity measurement, trigger, object identification, energy/momentum scale and resolution as well as underlying event and pile-up modelling. These uncertainties, discussed in Refs. [61, 10], impact the simulations of signal and background processes. Theoretical uncertainties include both the signal and background modelling. For the signal modelling, uncertainties due to the factorisation and renormalisation scale choice, the initial- and final-state radiation treatment and the PDF choice are considered. Additional systematic uncertainties are assigned to cover the differences in signal efficiencies and \( m_{\ell \ell} \) parameter values between the interpolations and the simulations. For the background modelling, the most important sources of systematic uncertainty are from the modelling of the \( m_{bb} \) and the \( p_T \) distributions of \( Z \)-jets. They are taken to be the difference between the data and simulation of the selected samples before the event categorisation and the \( m_{bb} \) requirement. The samples are dominated by the \( Z \)-jets contribution, and any potential signal contamination is expected to be negligible. For other background processes, they are obtained by varying the factorisation and renormalisation scales, the amount of initial- and final-state radiation, and the choices of PDF parameterisations.

The effect of these systematic uncertainties on the search is studied using the signal-strength parameter \( \mu \) for hypothesised signal production. Unainties having the largest impact depend on the choice of \( (m_A, m_H) \) signal point. Table 2 shows the relative uncertainties in the best-fit \( \mu \) value from the leading sources of systematic uncertainty for two example mass points of both gluon–gluon fusion and \( b \)-associated production of a narrow-width \( A \) boson. The leading sources of systematic uncertainty are similar for other mass points studied and for larger \( A \) boson widths. For all cases, the limited size of the simulated samples has the largest impact on the search sensitivity among all the sources of systematic uncertainty. While systematic uncertainties and the statistical uncertainty of the data have comparable impact at low masses, the search sensitivity is mostly determined at high masses by the limited size of the data sample.

8. Results

The \( m_{\ell \ell} \) distributions from different \( m_{bb} \) mass windows are scanned for potential excesses beyond the background expectations through signal-plus-background fits. The scan is performed in steps of 10 GeV for both the \( m_A \) range 230–800 GeV and the \( m_H \) range 130–700 GeV, such that \( m_A - m_H \geq 100 \) GeV. The step sizes are chosen to be compatible with the detector resolution for \( m_{\ell \ell} \) and \( m_{bb} \).

Fig. 3 shows the \( m_{\ell \ell} \) distributions in the \( n_b = 2 \) and \( n_b \geq 3 \) categories for the \( m_{bb} \) window defined for \( m_{H} = 200 \) GeV. The \( m_{bb} \) distributions before any \( m_{bb} \) window cut are also shown in this figure. The \( m_{\ell \ell} \) distributions for two other \( m_{bb} \) windows, defined for \( m_{H} = 300 \) GeV and \( m_{H} = 500 \) GeV are shown in Fig. 4. In all cases, the data are found to be well described by the background model. The most significant excess for the gluon–gluon fusion production signal assumption is at the \( (m_A, m_H) = (750, 610) \) GeV signal point, for which the local (global) significance [78] is 3.5 (2.0) standard deviations. For the \( b \)-associated production, the most significant excess is at the \( (m_A, m_H) = (510, 130) \) GeV signal point, for which the local (global) significance is 3.0 (1.2) standard deviations. The significances are calculated for each production process separately ignoring the contribution from the other.

In the absence of a statistically significant excess, constraints on the production of \( A \to ZH \) followed by the \( H \to bb \) decay are derived. The method of Ref. [79] is used to calculate 95% confidence level (CL) upper bounds on the product of cross-section and decay branching ratios, \( \sigma \times B(A \to ZH) \times B(H \to bb) \), using the asymptotic approximation [77]. The upper limits are shown in Fig. 5 for a narrow-width \( A \) boson produced via gluon–gluon fusion and \( b \)-associated production. As for the significance calculations above, these limits are derived separately for each production process. For the gluon–gluon fusion limits, only the \( n_b = 2 \) category is used. For the \( b \)-associated production, both the \( n_b = 2 \) and \( n_b \geq 3 \) categories are used. The upper limit for gluon–gluon fusion varies from 14 fb for the \( (m_A, m_H) = (800, 140) \) GeV signal point to 830 fb for the \( (m_A, m_H) = (240, 130) \) GeV signal point. This is to be compared with the corresponding expected limits of 24.1 fb and 469 fb for these two signal points. For the \( b \)-associated production the upper limits vary from 26 fb for the \( (m_A, m_H) = (780, 680) \) GeV signal point to 830 fb for the \( (m_A, m_H) = (240, 130) \) GeV signal point with expected limits of 46 fb and 360 fb, respectively.

The results of the search are interpreted in the context of the 2HDM. For this interpretation, several assumptions are made to reduce the number of free parameters in the model. The charged Higgs boson is assumed to have the same mass as the \( A \) boson. The 2HDM parameter \( m_{\ell \ell}^{\pm} \) is fixed to \( m_{\ell \ell}^{\pm} \tan \beta / (1 + \tan^2 \beta) \). The lightest Higgs boson of the model, \( h \), is assumed to have a mass of 125 GeV and its couplings are set to be the same as those of the SM Higgs boson, by choosing \( \cos \beta - \alpha = 0 \). The widths of the \( A \) and \( H \) bosons are taken from the predictions of the 2HDM. The cross-sections for \( A \) boson production in the 2HDM are calculated using up to NNLO QCD corrections for gluon–gluon fusion and \( b \)-associated production in the five-flavour scheme as implemented in SUSHI [80–83]. For \( b \)-associated production a cross-section in the four-flavour scheme is also calculated as described in Refs. [84, 85] and the results are combined with the five-flavour scheme calculation following Ref. [86]. The Higgs boson widths and branching ratios are calculated using 2HDMC [87]. The procedure for the calculation of the cross-sections and branching ratios, as well as for the choice of 2HDM parameters, follows Ref. [36].

Since both gluon–gluon fusion and \( b \)-associated production are expected, a new signal model weighted by the predicted cross-sections of the two processes is built for every point tested in the 2HDM parameter space. Upper limits on \( \sigma \times B(A \to ZH) \times B(H \to bb) \) with \( \sigma \) here including contributions from both processes are recalculated and compared with the 2HDM predictions to derive the limits in the 2HDM parameter space. Fig. 6 shows the observed and expected limits for Type I, Type II, ‘lepton specific’ and ‘flipped’ 2HDMs in the \( (m_A, m_H) \) plane for various tan \( \beta \) values. Type-II and flipped show similar constraints because in these models the Yukawa couplings are the same for all fermions apart from leptons. The same holds when comparing Type-I and lepton-specific 2HDM, where the main reason for the difference in sensitivity is the increased significance of the \( H \to \tau \tau \) decay in the lepton-specific model. The gluon–gluon fusion production cross section decreases with increasing tan \( \beta \), which explains the loss of sensitivity in Type-I and lepton specific for large tan \( \beta \) values. In the case of Type-II and flipped, at large tan \( \beta \) values it is the \( b \)-associated production that dominates instead of the gluon–gluon fusion in this region. For instance the exclusion can reach up to \( m_H \approx 400 \) GeV at lower tan \( \beta \) (less than 10) and \( m_H \approx 600 \) GeV at higher tan \( \beta \) (more than 20). At low tan \( \beta \) values the Higgs boson branching fraction to \( \ell \ell \) becomes sizable, and this is what limits the sensitivity to below \( m_H \approx 350 \) GeV in all models examined here.
9. Conclusion

Data recorded by the ATLAS experiment at the LHC, corresponding to an integrated luminosity of 36.1 fb$^{-1}$ from proton-proton collisions at a centre-of-mass energy 13 TeV, are used to search for a heavy Higgs boson, $A$, decaying into $ZH$, where $H$ denotes a heavy Higgs boson with mass $m_H > 125$ GeV. The $A$ boson is assumed to be produced via either gluon–gluon fusion or $b$-associated production. No significant deviation from the SM background predictions are observed in the $ZH \rightarrow \ell\ell b\bar{b}$ final state that is considered in this search. Considering each production process separately, upper limits are set at the 95% confidence level for $\sigma \times B(A \rightarrow ZH) \times B(H \rightarrow b\bar{b})$ of 14–830 fb for gluon–gluon fusion and 26–570 fb for $b$-associated production of a narrow $A$ boson for the mass ranges 130–700 GeV of the $H$ boson and 230–800 GeV of the $A$ boson. Taking into account both production processes, this search tightens the constraints on the 2HDM in the case of large mass splittings between its heavier neutral Higgs bosons.

Fig. 4. The $m_{\ell\ell}$ mass distribution for the $m_H = 300 \text{ GeV}$ and $m_H = 500 \text{ GeV}$ for (a, c) the $n_b = 2$ and (b, d) the $n_b \geq 3$ category, respectively. Signal distributions are also shown for gluon–gluon fusion production in (a, c) and $b$-associated production in (b, d) assuming production cross-sections times the branching ratios $B(A \to ZH)$ and $B(H \to bb)$ of $1 \text{ pb}$. The same conventions as in Fig. 3 are used.

Acknowledgements

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

We acknowledge the support of ANPCyT, Argentina; YerPhI, Armenia; ARC, Australia; BMWFW and FWF, Austria; ANAS, 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; SRNSFG, Georgia; BMBF, HGF, and MPG, Germany; CSRT, 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; MNiSW and NCN, Poland; FCT, Portugal; MINE/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
Fig. 5. Upper bounds at 95% CL on the production cross-section times the branching ratio \(\sigma \times \mathcal{B}(A \rightarrow ZH) \times \mathcal{B}(H \rightarrow bb)\) in pb for (a, b) gluon–gluon fusion and (c, d) \(b\)-associated production. The expected upper limits are shown in (a) and (c) and the observed upper limits are shown in (b) and (d).

Fig. 6. Observed and expected 95% CL exclusion regions in the \((m_A, m_H)\) plane for various \(\tan\beta\) values for (a) Type I, (b) Type II, (c) lepton specific and (d) flipped 2HDM.
by EU-ESF and the Greek NSF; 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. [88].

References


