Search for heavy Higgs bosons with flavour-violating couplings in multi-lepton plus b-jets final states in pp collisions at 13 TeV with the ATLAS detector

Aad, G.; Abbott, B.; Abeling, K.; Abicht, N.J.; Abidi, S.H.; Aboulhorma, A; Abramowicz, H.; Abreu, H.; Abulaiti, Y.; Abusleme Hoffman, A. C.; Dam, Mogens; Camplani, Alessandra; Hansen, Peter Henrik; Hansen, Jørn Dines; Hansen, Jørgen Beck; Ignazzi, Rosanna; Petersen, Troels Christian; Wiglesworth, Graig; Xella, Stefania; ATLAS Collaboration

Published in:
Journal of High Energy Physics

DOI:
10.1007/JHEP12(2023)081

Publication date:
2023

Document version
Publisher's PDF, also known as Version of record

Document license:
CC BY

Citation for published version (APA):

Download date: 05. apr. 2024
Search for heavy Higgs bosons with flavour-violating couplings in multi-lepton plus $b$-jets final states in $pp$ collisions at 13 TeV with the ATLAS detector

The ATLAS collaboration

E-mail: atlas.publications@cern.ch

ABSTRACT: A search for new heavy scalars with flavour-violating decays in final states with multiple leptons and $b$-tagged jets is presented. The results are interpreted in terms of a general two-Higgs-doublet model involving an additional scalar with couplings to the top-quark and the three up-type quarks ($\rho_{tt}$, $\rho_{tc}$, and $\rho_{tu}$). The targeted signals lead to final states with either a same-sign top-quark pair, three top-quarks, or four top-quarks. The search is based on a data sample of proton-proton collisions at $\sqrt{s} = 13$ TeV recorded with the ATLAS detector during Run 2 of the Large Hadron Collider, corresponding to an integrated luminosity of 139 fb$^{-1}$. Events are categorised depending on the multiplicity of light charged leptons (electrons or muons), total lepton charge, and a deep-neural-network output to enhance the purity of each of the signals. Masses of an additional scalar boson $m_H$ between $200 - 630$ GeV with couplings $\rho_{tt} = 0.4$, $\rho_{tc} = 0.2$, and $\rho_{tu} = 0.2$ are excluded at 95% confidence level. Additional interpretations are provided in models of $R$-parity violating supersymmetry, motivated by the recent flavour and $(g - 2)_\mu$ anomalies.

KEYWORDS: Beyond Standard Model, Exotics, Hadron-Hadron Scattering

ArXiv ePrint: 2307.14759
1 Introduction

Several extensions of the Standard Model (SM) propose the augmentation of the Higgs sector by the addition of a second complex Higgs doublet \[1, 2\] (2HDM), giving rise to five Higgs bosons: two CP-even scalar fields \(h\) and \(H\), one CP-odd pseudo-scalar \(A\), and two charged fields \(H^\pm\). The two CP-even scalars are expected to mix; however, the measurement of Higgs boson properties has revealed no deviations from the expectations of the Standard Model \[3, 4\]. This implies that extra scalars from 2HDMs have to be either very heavy (decoupling limit) or have a vanishingly small mixing with the SM Higgs (alignment...
limit). To avoid flavour changing neutral Higgs (FCNH) couplings mediated by the SM Higgs, a discrete $Z_2$ symmetry is usually imposed [1, 2]. A large set of searches for heavy scalars or pseudo-scalars with flavour-conserving decays has been performed in ATLAS [5–12] and CMS [13–25]. However, if the $Z_2$ symmetry is dropped, alignment automatically emerges when all heavy Higgs quartic couplings are $O(1)$ [26]. Therefore, models without $Z_2$ symmetry can lead naturally to the alignment limit and predict FCNH couplings in the heavy Higgs sector, while respecting the SM-like nature of the $h(125)$ discovered at the LHC.

The search presented here targets a general two Higgs doublet model (g2HDM) without $Z_2$ symmetry [27], where the heavy Higgs bosons feature FCNH couplings. Only couplings involving top-quarks are considered: $\rho_{tt}$, $\rho_{tc}$, and $\rho_{tu}$. The $\rho_{AB}$ parameters indicate the coupling of the heavy Higgs boson to particles $A$ and $B$. The notation $\hat{\rho}_{aq}$ is used to refer to both the $\rho_{tc}$ and $\rho_{tu}$ couplings. The g2HDM model with its additional top Yukawa couplings is phenomenologically interesting since it can explain the generation of the baryon asymmetry through the couplings $\rho_{tt}$ or $\rho_{tc}$ [28].

No distinction is performed between the different chiralities in the coupling, and an effective coupling $\rho_{tq} = \sqrt{\hat{\rho}_{tLcR}^2 + \hat{\rho}_{qLtR}^2} / \sqrt{2}$ is used, where the hat symbol is used to denote the original couplings in the g2HDM Lagrangian.

The production and decay modes at tree level considered in the analysis are shown in figure 1. The presence of the $\rho_{tq}$ coupling opens the possibility of same-sign top production, as shown in figures 1a (sstt) and 1b (ttq), and also three-top production, as shown in figures 1c (ttt) and 1d (tttq). The three-top signature is a sensitive probe of beyond-the-SM (BSM) physics [27, 29–31]. Additionally, four-top quarks can be produced, as shown in figure 1e (tttt). The targeted final state is characterised by multiple leptons (electrons and muons) and multiple jets containing $b$-flavoured hadrons ($b$-jets). Many of the production modes are expected to be charge-asymmetric (with preference to positively charged), and this feature is exploited in the search. The relevance of each production mode depends on the chosen coupling. A benchmark of $\rho_{tt} = 0.4$ and $\rho_{tq} = 0.2$ is chosen to guide the analysis design and optimisation. The values are chosen so that the signal could account for the higher $t\bar{t}W$ and $t\bar{t}t\bar{t}$ yields observed in ATLAS analyses [32–36]. Significant kinematic differences and a much stronger charge asymmetry are expected from the targeted signals, which allow them to be differentiated from simple rescalings of both processes. For the chosen couplings, the $tH$ production (figures 1b and 1c) cross section is two orders of magnitude larger than $ttH$ production (figures 1d and 1e), and three orders of magnitude larger than same-sign tops production via $t$-channel $H$ (figures 1a).

This analysis is the first to target BSM production leading to three-top final states and the first to probe the g2HDM. The production of four-tops in the SM or via heavy scalars was explored previously by ATLAS [35–39] and CMS [40–43]. Limits on the g2HDM model couplings can be derived from LHC Higgs measurements, $B$ physics, and assuming the couplings stay perturbative [44], leading to $\rho_{tt} < 2$, $\hat{\rho}_{LcR} < 1.5$, and $\hat{\rho}_{Ltr} < 0.1$, implying

\[\text{The g2HDM signal processes sstt, ttq, ttt, tttq, and tttt include tt and } \bar{t}\bar{t}, \text{ttq and } \bar{t} \bar{q}, \text{tt} \bar{t} \text{ and } \bar{t} \bar{t}, \text{ttq and } \bar{t} \bar{t} \bar{q}, \text{and } \bar{t} \bar{t} \bar{t}, \text{respectively, where the } q \text{ can be an up/charm or anti-up/anti-charm quark.} \]
Figure 1. Signal diagrams for the dominant production and decay modes of the heavy scalar considered in the analysis. The subsequent decay can lead to a final state with high multiplicity of leptons and $b$-jets that is targeted by the search. Single $H$ production through gluon-gluon fusion, $gg \rightarrow H \rightarrow t\bar{t}/tq$, is not considered since the decay does not lead to the relevant multi-lepton final state.

\[ \rho_{tc} < 1.06. \] In addition, $K - \bar{K}$ mixing provides the constraint $\hat{\rho}_{cL} t_R < 0.14$ [45], and $D - \bar{D}$ mixing provides the constraint $|\hat{\rho}_{L\ell R}^* \hat{\rho}_{L\ell R}| < 0.02$ [45], which translates to $|\rho_{tc} \rho_{tu}| < 0.01$ assuming a negligible $\hat{\rho}_{uL} t_R$. These constraints are derived assuming $m_H \approx m_{H^+_L} = 500$ GeV, and become weaker for higher masses.

The event selections optimised for the heavy scalar signal models are also sensitive to models based on $R$-parity-violating (RPV) supersymmetry (SUSY). These models are motivated by the recent flavour anomalies [46–50] and $(g - 2)_{\mu}$ anomaly [51], and can provide a successful explanation with different choices of particles, masses, and couplings [52–61]. The RPV SUSY models discussed below are also used to interpret the results of this search.

The first model features production of electroweakinos (wino or Higgsino) that decay via a lepton-number-violating RPV coupling of the $LQ \bar{D}$ type to a lepton and third-generation quarks. The corresponding term in the superpotential has the form $\lambda'_{33} L_i Q_3 \bar{D}_3$, where $i \in 2, 3$ is a generation index, and $L, Q, \bar{D}$ are the lepton doublet, quark doublet, and down-type quark singlet superfields, respectively. Relevant diagrams for the production and decay are shown in figures 2a and 2b. The second model features direct smuon production and decay to a bino-like neutralino, which in turn decays via the same RPV coupling ($\lambda'_{33}$), as shown in figure 2c.

2 ATLAS detector

The ATLAS detector [62] at the LHC covers nearly the entire solid angle around the collision point.\(^2\) It consists of an inner tracking detector surrounded by a thin superconducting
solenoid, electromagnetic and hadron calorimeters, and a muon spectrometer incorporating three large superconducting air-core toroidal magnets.

The inner-detector system (ID) is immersed in a 2 T axial magnetic field and provides charged-particle tracking in the region $|\eta| < 2.5$. The high-granularity silicon pixel detector covers the vertex region and typically provides four measurements per track, the first hit normally being in the insertable B-layer (IBL) installed before Run 2 [63, 64]. It is followed by the silicon microstrip tracker (SCT), which usually provides eight measurements per track. These silicon detectors are complemented by the transition radiation tracker (TRT), which enables radially extended track reconstruction up to $|\eta| = 2.0$. The TRT also provides electron identification information based on the fraction of hits (typically 30 in total) above a higher energy-deposit threshold corresponding to transition radiation.

The calorimeter system covers the pseudorapidity range $|\eta| < 4.9$. Within the region $|\eta| < 3.2$, electromagnetic calorimetry is provided by barrel and endcap high-granularity lead/liquid-argon (LAr) calorimeters, with an additional thin LAr presampler covering $|\eta| < 1.8$ to correct for energy loss in material upstream of the calorimeters. Hadron calorimetry is provided by the steel/scintillator-tile calorimeter, segmented into three barrel structures within $|\eta| < 1.7$, and two copper/LAr hadron endcap calorimeters. The solid angle coverage is completed with forward copper/LAr and tungsten/LAr calorimeter modules optimised for electromagnetic and hadronic energy measurements, respectively.

The muon spectrometer (MS) comprises separate trigger and high-precision tracking chambers measuring the deflection of muons in a magnetic field generated by the superconducting air-core toroidal magnets. The field integral of the toroids ranges between 2.0 and 6.0 Tm across most of the detector. Three layers of precision chambers, each consisting of layers of monitored drift tubes, cover the region $|\eta| < 2.7$, complemented by cathode-strip chambers in the forward region, where the background is highest. The muon trigger system covers the range $|\eta| < 2.4$ with resistive-plate chambers in the barrel, and thin-gap chambers in the endcap regions.

The 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 $\eta = -\ln \tan(\theta/2)$. Angular distance is measured in units of $\Delta R \equiv \sqrt{\Delta \eta^2 + \Delta \phi^2}$.

\begin{figure}[h]
\centering
\includegraphics[width=\textwidth]{figure2}
\caption{Signal diagrams for the RPV SUSY signals used as additional interpretation in the analysis. The subsequent decay can lead to a final state with high multiplicity of leptons and $b$-jets that is targeted by the search.}
\end{figure}
Interesting events are selected by the first-level trigger system implemented in custom hardware, followed by selections made by algorithms implemented in software in the high-level trigger [65]. The first-level trigger accepts events from the 40 MHz bunch crossings at a rate below 100 kHz, which the high-level trigger reduces further to record events to disk at about 1 kHz.

An extensive software suite [66] is used in the reconstruction and analysis of real and simulated data, in detector operations, and in the trigger and data acquisition systems of the experiment.

3 Data and simulated event samples

This analysis uses data from $pp$ collisions at $\sqrt{s} = 13$ TeV collected by the ATLAS experiment during 2015–2018. After the application of data-quality requirements [67], the data sample corresponds to an integrated luminosity of 139 fb$^{-1}$ [68]. The number of additional $pp$ interactions per bunch crossing (pile-up) in this sample ranges from about 8 to 70, with an average of 34. Only events recorded under stable beam conditions and for which all detector subsystems were known to be in a good operating condition are used. The trigger requirements are discussed in section 5.

Monte Carlo (MC) simulation samples were produced for the different signal and background processes. Table 1 shows the configurations used in this analysis, with the samples in parentheses and in grey indicating those used to estimate the systematic uncertainties. All simulated samples, except those produced with the SHERPA [69] event generator, utilised EvtGen 1.2.0 [70] to model the decays of heavy-flavour hadrons. All samples showered with PYTHIA use the A14 set of tuned parameters [71] (referred to as ‘tune’), whereas those showered with HERWIG use the H7-UE tune [72]. Pile-up was modelled using events from minimum-bias interactions generated with PYTHIA 8.186 [73] with the A3 tune [74], and overlaid onto the simulated hard-scatter events according to the luminosity profile of the recorded data. The generated events were processed through either a full simulation of the ATLAS detector geometry and response using GEANT4 [75], or a faster simulation where the full GEANT4 simulation of the calorimeter response is replaced by a detailed parameterisation of the shower shapes [76]. Both types of simulated events were processed through the same reconstruction software used for the $pp$ collision data. Corrections were applied to the simulated events so that the particle candidates’ selection efficiencies, energy scales and energy resolutions match those determined from data control samples. The simulated samples are normalised to their cross sections, and generated to the highest order available in perturbation theory.

Samples used to model the g2HDM signal were generated at leading-order (LO) in QCD with MADGRAPH v2.9.3 [77] with the NNPDF3.1nlo [78] parton distribution function (PDF) set. Samples were generated for masses in the range of 200 GeV to 1.5 TeV with a 100 GeV step, and processed with the ATLAS Fast Simulation [76]. All signals were produced with the set of couplings $\rho_{tt} = \rho_{tc} = \rho_{tu} = 0.1$. Each signal process described in section 1 was generated as a separate MC sample. The LO cross section obtained from Madgraph is used for the normalisation of the signals. Simulated events for different
coupling values are obtained by rescaling the samples to match the target cross section and branching ratio of each subprocess. For a given choice of couplings all the processes are taken into account and rescaled. The RPV SUSY signal samples were generated with Madgraph v2.9.3, with up to two extra jets at LO in QCD. The matching scale is set at 1/4 of the mass of the SUSY particle being produced. Supersymmetric particle decays via the RPV coupling are simulated with 25% branching ratio to $\mu/\tau/\nu$ each, a $b$-quark, and a $b$- or $t$-quark depending on the lepton charge. The identical branching ratio to second- and third-generation leptons follows from the choice of $\lambda'_{233} = \lambda'_{333}$, while the balance in charged and neutral leptons is an assumption. This assumption originates naturally from the presence of a left-handed lepton superfield in the $LQ\bar{D}$ coupling, but is distorted by the large mass difference between top and bottom quarks and also affected by the choice of $\tan(\beta)$. Signal cross sections are calculated to next-to-leading order in the strong coupling constant, adding the resummation of soft gluon emission at next-to-leading-logarithmic accuracy (NLO+NLL) [79–83]. The nominal cross section and the uncertainty are taken from an envelope of cross-section predictions using different PDF sets and factorisation and renormalisation scales, as described in ref. [84]. All signal events were showered with Pythia 8.245 [73] using the NNPDF2.3lo [85] PDF set.

The samples used to model the $t\bar{t}W$ and the $t\bar{t}(Z/\gamma^* \rightarrow \ell^+\ell^-)$ backgrounds were generated using Sherpa-2.2.10 [86] and Sherpa-2.2.11, where the matrix element (ME) were calculated for up to one and zero additional partons at next-to-leading-order (NLO) in QCD, respectively, and up to two partons at LO in QCD using Comix [87] and OpenLoops [88]. The ME were merged with the Sherpa parton shower (PS) [89] using the MePs@Nlo prescription [90], with a CKKW merging scale of 30 GeV for the $t\bar{t}W$ sample. These samples are generated using the NNPDF3.0nnlo [91] PDF set, along with the dedicated set of tuned parton-shower parameters developed by the Sherpa authors. The invariant mass of the lepton pair ($m_{\ell^+\ell^-}$) in the $t\bar{t}(Z/\gamma^* \rightarrow \ell^+\ell^-)$ sample is set to be greater than 1 GeV. Both the factorisation and renormalisation scales are set to $\mu_f = \mu_r = m_T/2$ in the $t\bar{t}W$ sample, where $m_T$ is defined as the scalar sum of the transverse masses $\sqrt{m^2 + p_T^2}$ of the particles generated from the ME calculation. In addition to this $t\bar{t}W$ prediction at NLO in QCD, higher-order corrections relating to electroweak (EW) contributions are also included. First, event-by-event correction factors are applied that provide virtual NLO EW corrections of the order $\alpha^2\alpha_s^2$ derived using the formalism described in ref. [92] along with LO corrections of order $\alpha^3$. Second, real emission contributions from the sub-leading EW corrections at order $\alpha^3\alpha_s$ [93] are simulated with an independent Sherpa-2.2.10 sample produced at LO in QCD. The complete $t\bar{t}W$ simulation is normalised to the total cross section of $\sigma(t\bar{t}W) = 614.7$ fb that comes from the Sherpa configuration outlined above considering NLO QCD and NLO EWK effects, based on a similar strategy as used in ref. [94]. The $t\bar{t}Z/\gamma^*$ sample is normalised to the cross section $\sigma(t\bar{t}Z/\gamma^*) = 839$ fb, calculated at NLO QCD and NLO EW accuracy using MadGraph5_aMC@NLO [95] and scaled by an off-shell correction estimated at one-loop level in $\alpha_s$.

The production of SM $t\bar{t}t\bar{t}$ events was modelled using the MadGraph5_aMC@NLO v2.6.2 generator that provides matrix elements at NLO in QCD with the NNPDF3.1nlo
PDF set. The functional form of the renormalisation and factorisation scales are set to \( \mu_r = \mu_T = m_T/4 \). Top quarks are decayed at LO using MadSpin to preserve all spin correlations. The events are interfaced with Pythia 8.230 for the parton shower and hadronisation, using the NNPDF2.3LO PDF set. The production of \( t\bar{t}t\bar{t} \) events is normalised to a cross section of 12 fb computed at NLO in QCD including EW corrections [93].

Diboson (VV) background processes were simulated with Sherpa 2.2.2 [86]. The matrix element was calculated using Comix [87] and OpenLoops [88] with NLO accuracy in QCD for up to one additional parton and at LO accuracy for up to three additional partons, and merged with the Sherpa PS using MePS@NLO prescription [90]. The NNPDF3.0NNLO set of PDFs was used, along with the dedicated parton-shower tune for Sherpa. The cross section of \( \sigma(VV) = 104 \text{ pb} \) used to normalise the sample was computed by Sherpa 2.2.2.

Samples for \( t\bar{t}h \), \( t\bar{t} \), and single top production were generated using the NLO generator Powheg-Box-v2 [96–101] and interfaced with Pythia 8 for the parton showering and fragmentation. These samples used the NNPDF3.0NLO PDF set. The \( h_{\text{damp}} \) parameter, which controls the transverse momentum of the first additional emission beyond the LO Feynman diagram in the PS and therefore regulates the high-\( p_T \) radiation, is set to \( 3/4 \times (m_t + m_\chi + m_h) \) in the \( t\bar{t}h \) sample and to \( 1.5 \times m_t \) in the \( tt \) and single top samples, where \( m_t \) (\( m_h \)) denotes the mass of the top quark (SM Higgs boson).

A dedicated \( t\bar{t} \) sample including rare \( t \rightarrow Wb\gamma^* (\rightarrow \ell^+\ell^-) \) radiative decays, \( t\bar{t} \rightarrow W^+bW^-b\ell^+\ell^- \), was generated using a ME calculated at LO in QCD and requiring \( m_{\ell^+\ell^-} > 1 \text{ GeV} \). In this sample the photon can be radiated from the top quark, the W boson, or the b-quark. Both the \( t\bar{t}(Z/\gamma^* \rightarrow \ell^+\ell^-) \) and \( t\bar{t} \rightarrow W^+bW^-b\ell^+\ell^- \) samples are combined and together form the “\( t\bar{t}Z/\gamma^* \)” sample. The contribution from internal photon conversions \( (\gamma^* \rightarrow \ell^+\ell^-) \) with \( m_{\ell^+\ell^-} < 1 \text{ GeV} \) were modelled by QED multi-photon radiation via the PS in an inclusive \( t\bar{t} \) sample and is referred to as “\( t\bar{t}\gamma^* \) (LM)”. Dedicated Z+jets samples containing electrons from material photon conversion \( (\gamma \rightarrow e^+e^-) \) or internal photon conversion were generated with Powheg-Box and interfaced with Pythia 8 for the parton showering and fragmentation. These samples are used to model the data in control regions enriched in material and internal conversion electrons, as explained in section 5.

The remaining rare background contributions listed in table 1 are normalised using their NLO theoretical cross sections, except for the \( t\bar{t}t \), \( t\bar{t}W^+W^- \), \( t\bar{t}ZZ \), \( t\bar{t}h \), and \( t\bar{t}Wh \) processes, for which a LO cross section is used.

4 Event reconstruction and object identification

Interaction vertices from the \( pp \) collisions are reconstructed from at least two tracks with transverse momentum \( (p_T) \) larger than 500 MeV that are consistent with originating from the beam collision region in the \( x-y \) plane. If more than one primary vertex candidate is found in the event, the candidate for which the associated tracks form the largest sum of squared \( p_T \) is selected as the hard-scatter primary vertex [104].

Electron candidates are reconstructed from energy clusters in the electromagnetic calorimeter matched to a track in the ID [105]. They are required to satisfy \( p_T > 10 \text{ GeV} \).
Table 1. The configurations used for event generation of signal and background processes. The samples used to estimate the systematic uncertainties are indicated in parentheses and grey. $V$ refers to production of an electroweak boson ($W$ or $Z/\gamma^*$). The matrix element order refers to the order in the strong coupling constant of the perturbative calculation. The “$t\bar{t}W$ (EW)” sample also includes next-to-leading-order electroweak corrections. Tune refers to the underlying-event tune of the parton shower generator. MG5\_aMC refers to MadGraph5\_aMC@NLO 2.2, 2.3, or 2.6; Pythia 8 refers to version 8.2 [102]; MePs@NLO refers to the method used in Sherpa to match the matrix element to the parton shower. All samples include leading-logarithm photon emission, either modelled by the parton shower generator or by Photos [103]. The mass of the top quark ($m_t$) and SM Higgs boson were set to 172.5 GeV and 125 GeV, respectively.

<table>
<thead>
<tr>
<th>Process</th>
<th>Generator</th>
<th>ME order</th>
<th>Parton shower</th>
<th>PDF</th>
<th>Tune</th>
</tr>
</thead>
<tbody>
<tr>
<td>$g\to H$ signal</td>
<td>MG5_aMC</td>
<td>LO</td>
<td>Pythia 8</td>
<td>NNPDF3.1nlo</td>
<td>A14</td>
</tr>
<tr>
<td>SUSY signal</td>
<td>MG5_aMC</td>
<td>LO</td>
<td>Pythia 8</td>
<td>NNPDF3.1nlo</td>
<td>A14</td>
</tr>
<tr>
<td>$t\bar{t}W$</td>
<td>Sherpa 2.2.10</td>
<td>MePs@NLO</td>
<td>Sherpa</td>
<td>NNPDF3.0nnlo</td>
<td>Sherpa default</td>
</tr>
<tr>
<td></td>
<td>(MG5_aMC)</td>
<td>(NLO)</td>
<td>(Pythia 8)</td>
<td>(NNPDF3.0nlo)</td>
<td>(A14)</td>
</tr>
<tr>
<td>$t\bar{t}W$ (EW)</td>
<td>Sherpa 2.2.10</td>
<td>LO</td>
<td>Sherpa 8</td>
<td>NNPDF3.0nnlo</td>
<td>Sherpa default</td>
</tr>
<tr>
<td></td>
<td>(MG5_aMC)</td>
<td>(LO)</td>
<td>(Pythia 8)</td>
<td>(NNPDF3.0nlo)</td>
<td>(A14)</td>
</tr>
<tr>
<td>$t\bar{t}t$</td>
<td>MG5_aMC</td>
<td>NLO</td>
<td>Pythia 8</td>
<td>NNPDF3.1nlo</td>
<td>A14</td>
</tr>
<tr>
<td>$t\bar{t}(Z/\gamma^* \to \ell^+\ell^-)$</td>
<td>Sherpa 2.2.11</td>
<td>MePs@NLO</td>
<td>Sherpa</td>
<td>NNPDF3.0nnlo</td>
<td>Sherpa default</td>
</tr>
<tr>
<td></td>
<td>(MG5_aMC)</td>
<td>(NLO)</td>
<td>(Pythia 8)</td>
<td>(NNPDF3.0nlo)</td>
<td>(A14)</td>
</tr>
<tr>
<td>$tt\to W^+bW^-b^+\ell^-$</td>
<td>MG5_aMC</td>
<td>LO</td>
<td>Pythia 8</td>
<td>NNPDF3.0nlo</td>
<td>A14</td>
</tr>
<tr>
<td>$t(Z/\gamma^*)$</td>
<td>MG5_aMC</td>
<td>NLO</td>
<td>Pythia 8</td>
<td>NNPDF3.0nlo</td>
<td>A14</td>
</tr>
<tr>
<td>$tW(Z/\gamma^*)$</td>
<td>MG5_aMC</td>
<td>NLO</td>
<td>Pythia 8</td>
<td>NNPDF3.0nlo</td>
<td>A14</td>
</tr>
<tr>
<td>$t\bar{t}$</td>
<td>Powheg-BOX</td>
<td>NLO</td>
<td>Pythia 8</td>
<td>NNPDF3.0nlo</td>
<td>A14</td>
</tr>
<tr>
<td>$t\bar{t}(Z/\gamma^* \to \ell^+\ell^-)$</td>
<td>Sherpa 2.2.11</td>
<td>MePs@NLO</td>
<td>Sherpa</td>
<td>NNPDF3.0nnlo</td>
<td>Sherpa default</td>
</tr>
<tr>
<td></td>
<td>(MG5_aMC)</td>
<td>(NLO)</td>
<td>(Pythia 8)</td>
<td>(NNPDF3.0nlo)</td>
<td>(A14)</td>
</tr>
<tr>
<td>$t\bar{t}t$</td>
<td>Powheg-BOX</td>
<td>NLO</td>
<td>Pythia 8</td>
<td>NNPDF3.0nlo</td>
<td>A14</td>
</tr>
<tr>
<td></td>
<td>(MG5_aMC)</td>
<td>(NLO)</td>
<td>(Pythia 8)</td>
<td>(NNPDF3.0nlo)</td>
<td>(A14)</td>
</tr>
<tr>
<td>Single top</td>
<td>Powheg-BOX</td>
<td>NLO</td>
<td>Pythia 8</td>
<td>NNPDF3.0nlo</td>
<td>A14</td>
</tr>
<tr>
<td>($t-, W-, s$-channel)</td>
<td>VV, $qqVV$, VVV</td>
<td>SHERPA 2.2.2</td>
<td>MePs@NLO</td>
<td>Sherpa</td>
<td>NNPDF3.0nnlo</td>
</tr>
<tr>
<td>$Z \to \ell^+\ell^-$</td>
<td>SHERPA 2.2.1</td>
<td>MePs@NLO</td>
<td>SHERPA</td>
<td>NNPDF3.0nlo</td>
<td>SHERPA default</td>
</tr>
<tr>
<td>$Z \to \ell^+\ell^- (\gamma \to e^+e^-)$</td>
<td>Powheg-BOX</td>
<td>NLO</td>
<td>Pythia 8</td>
<td>CTEQ6L1nlo</td>
<td>A14</td>
</tr>
<tr>
<td>$Z \to \ell^+\ell^- (\gamma* \to e^+e^-)$</td>
<td>Powheg-BOX</td>
<td>NLO</td>
<td>Pythia 8</td>
<td>CTEQ6L1nlo</td>
<td>A14</td>
</tr>
<tr>
<td>$W+$jets</td>
<td>SHERPA 2.2.1</td>
<td>MePs@NLO</td>
<td>SHERPA</td>
<td>NNPDF3.0nlo</td>
<td>SHERPA default</td>
</tr>
<tr>
<td>$Vh$</td>
<td>Powheg-BOX</td>
<td>NLO</td>
<td>Pythia 8</td>
<td>NNPDF3.0nlo</td>
<td>A14</td>
</tr>
<tr>
<td>$ttW^+W^-$</td>
<td>MG5_aMC</td>
<td>LO</td>
<td>Pythia 8</td>
<td>NNPDF2.3nlo</td>
<td>A14</td>
</tr>
<tr>
<td>$ttZZ$</td>
<td>MG5_aMC</td>
<td>LO</td>
<td>Pythia 8</td>
<td>NNPDF2.3nlo</td>
<td>A14</td>
</tr>
<tr>
<td>$ttbb$</td>
<td>MG5_aMC</td>
<td>LO</td>
<td>Pythia 8</td>
<td>NNPDF2.3nlo</td>
<td>A14</td>
</tr>
<tr>
<td>$ttWh$</td>
<td>MG5_aMC</td>
<td>LO</td>
<td>Pythia 8</td>
<td>NNPDF2.3nlo</td>
<td>A14</td>
</tr>
</tbody>
</table>
and $|\eta_{\text{cluster}}| < 2.47$, with the transition region between the endcap and barrel calorimeters ($1.37 < |\eta_{\text{cluster}}| < 1.52$) excluded. Loose and tight electron identification working points are used [105], based on a likelihood discriminant that employs calorimeter, tracking and combined variables to distinguish between electrons and jets. The associated track of an electron candidate is required to have at least two hits in the pixel detector and seven hits total in the pixel and silicon-strip detectors combined. For the tight identification working point, one of these pixel hits must be in the innermost layer, or the next-to-innermost layer if the module traversed in the innermost layer is non-operational, and there must be no association with a vertex from a reconstructed photon conversion [106] in the detector material (denoted as ‘material conversion’).

Muon candidates are reconstructed by combining tracks in the ID with tracks in the MS [107]. The resulting muon candidates are re-fit using the complete track information from both detector systems [108]. They are required to satisfy $p_T > 10$ GeV and $|\eta| < 2.5$. Loose and medium muon identification working points are used [108].

Electron (muon) candidates are matched to the primary vertex by requiring that their transverse impact parameter, $d_0$, satisfies $|d_0/\sigma(d_0)| < 5 (3)$, where $\sigma(d_0)$ is the measured uncertainty in $d_0$, and requiring that the longitudinal impact parameter, $z_0$, satisfies $|z_0\sin\theta| < 0.5$ mm, where $\theta$ is the polar angle of the track.

To further suppress leptons from heavy-flavour hadron decays, misidentified jets, or photon conversions (collectively referred to as ‘non-prompt leptons’), lepton candidates are also required to be isolated in the tracker and in the calorimeter [105]. A track-based lepton isolation criterion is defined by calculating the quantity $I_R = \sum p_T^{\text{trk}}$, where the scalar sum includes all tracks (excluding the lepton candidate itself) within the cone defined by $\Delta R < R_{\text{cut}}$ around the direction of the lepton. The value of $R_{\text{cut}}$ is the smaller of $r_{\text{min}}$ and $10$ GeV/$p_T^L$, where $r_{\text{min}}$ is set to 0.2 (0.3) for electron (muon) candidates and where $p_T^L$ is the lepton $p_T$. All lepton candidates must satisfy $I_R/p_T^L < 0.15$. Additionally, electrons (muons) are required to satisfy a calorimeter-based isolation criterion: the sum of the transverse energy within a cone of size $\Delta R = 0.2$ around the lepton, after subtracting the contributions from pile-up and the energy deposit of the lepton itself, is required to be less than 20% (30%) of $p_T^L$. Muons are required to be separated by $\Delta R > 0.2$ from any selected jets (defined below). If two electrons are closer than $\Delta R = 0.1$, only the one with the higher $p_T$ is considered. Electrons within $\Delta R = 0.1$ of a selected muon are removed.

The selection criteria described above greatly suppress the contribution from non-prompt leptons. However, several channels considered in this search have additional suppression requirements targeting the main non-prompt lepton types. Non-prompt leptons from hadron decays that contain bottom- and charm-quarks, denoted by ‘heavy-flavour (HF) non-prompt leptons’, are further rejected using a boosted decision tree (BDT) discriminant, referred to as the non-prompt lepton BDT [107, 109], which uses isolation and displacement information associated with a track jet that matches the selected light lepton. Three working points (WPs) are used: Tight, VeryTight, and Tight-not-VeryTight. The Tight WP allows to select prompt-like leptons with an efficiency for muons (barrel/endcap electrons) that satisfy the calorimeter- and track-based isolation criteria of about 60% (60/70%) for $p_T \sim 20$ GeV and reaches a plateau of 95% (95/90%) at $p_T \sim 40 (40/65)$ GeV.
The prompt lepton efficiency of the VeryTight WP for muons (barrel/endcap electrons) that satisfy the calorimeter- and track-based isolation criteria is about 55% (55/60%) for $p_T \sim 20$ GeV and reaches a plateau of 90% (85/83%) at $p_T \sim 40$ (40/65) GeV. The corresponding rejection factor$^3$ against muons (electrons) from the decay of $b$-hadrons ranges from 33 to 50 (20 to 50) for the Tight WP, and from 50 to 100 (33 to 66) for the VeryTight WP, depending on $p_T$ and $\eta$, after resolving ambiguities between overlapping reconstructed objects. The Tight-not-VeryTight WP allows to select non-prompt leptons and is part of the event selection of the control regions enriched in HF non-prompt lepton background, as described in section 6.

To further suppress electrons with incorrect charge assignment, a BDT discriminant based on calorimeter and tracking quantities [110] is used. An efficiency of approximately 96% in the barrel region and 81% in the endcaps is obtained, with rejection factors of 19 in the barrel region and 40 in the endcaps. Material and internal conversion candidates are identified based on a combination of requirements on the invariant mass of tracks and the radius from the reconstructed displaced vertex to the primary vertex. Material conversion candidates have a reconstructed displaced vertex with radius $r > 20$ mm that includes the track associated with the electron.$^4$ The invariant mass of the associated track and the closest (in $\Delta \eta$) opposite-charge track, calculated at the conversion vertex, is required to be less than 100 MeV. Internal conversion candidates, which correspond to the internal photon conversions (see section 3), must fail the requirements for material conversions, and the invariant mass of the track pair, calculated at the primary vertex, is also required to be less than 100 MeV.

The lepton working points used in this analysis are summarised in table 2. After the initial categorisation based on loose leptons (corresponding to “L”), the most optimal lepton working point to further optimise the event selection is chosen depending on the main background processes and the expected number of events in each category. The defined working points are medium inclusive (“M”), medium exclusive (“$M_{ex}$”), and tight (“$T$”). The various choices can be seen for the signal and control regions in section 5. Electron candidates from internal or material conversion are rejected from the $M$, $M_{ex}$, and $T$ electron selections. They are used to define control regions enriched in internal or material conversions, and are collectively denoted $e^*$ (see section 5).

The constituents for jet reconstruction are identified by combining measurements from both the ID and the calorimeter using a particle flow (PFlow) algorithm [111, 112]. Jet candidates are reconstructed from these PFlow objects using the anti-$k_t$ algorithm [113, 114] with a radius parameter of $R = 0.4$. They are calibrated using simulation with corrections obtained from in situ techniques in data [112]. Only jet candidates with $p_T > 25$ GeV that satisfy $|\eta| < 2.5$ are selected. To reduce the effects of pile-up, each jet with $p_T < 60$ GeV and $|\eta| < 2.4$ is required to satisfy the “Tight” working point of the Jet Vertex Tagger (JVT) [115] criteria used to identify the jets as originating from the selected primary vertex. A set of quality criteria is also applied to reject events containing at least one jet arising from non-collision sources or detector noise [116].

$^3$The rejection factor is defined as the reciprocal of the efficiency.

$^4$The beampipe and insertable B-layer inner radii are 23.5 mm and 33 mm, respectively.
<table>
<thead>
<tr>
<th>Lepton categorization</th>
<th>$\mu$</th>
<th>$e$</th>
</tr>
</thead>
<tbody>
<tr>
<td>Isolation</td>
<td>Yes</td>
<td>Yes</td>
</tr>
<tr>
<td>Non-prompt lepton BDT WP</td>
<td>No</td>
<td>Tight</td>
</tr>
<tr>
<td>Identification</td>
<td>Loose</td>
<td>Tight</td>
</tr>
<tr>
<td>Electron charge-misassignment veto</td>
<td>No</td>
<td>Yes</td>
</tr>
<tr>
<td>Electron conversion candidate veto</td>
<td>No</td>
<td>Yes (except $e^*$)</td>
</tr>
<tr>
<td>Transverse impact parameter</td>
<td>$&lt; 5$</td>
<td>$&lt; 3$</td>
</tr>
<tr>
<td>Longitudinal impact parameter</td>
<td>$&lt; 0.5$ mm</td>
<td></td>
</tr>
</tbody>
</table>

Table 2. Description of the loose inclusive (“L”), medium inclusive (“M”), medium exclusive (“$M_{ex}$”), and tight (“T”) lepton definitions. The electron $e^*$ is required to fulfill, in addition to the corresponding lepton definition requirements, those corresponding to an internal or material conversion candidate.

Jets containing $b$-hadrons are identified ($b$-tagged) via the DL1r algorithm [117, 118] that uses a deep-learning neural network based on the distinctive features of the $b$-hadrons in terms of the impact parameters of tracks and the displaced vertices reconstructed in the ID. Additional input to this network is provided by discriminant variables constructed by a recurrent neural network [119], which exploits the spatial and kinematic correlations between tracks originating from the same $b$-hadron. For each jet, a value for the multivariate $b$-tagging discriminant is calculated. A jet is $b$-tagged if the $b$-tagging score is above a certain threshold, referred to as a working point (WP). Four WPs are defined with average expected efficiencies for $b$-jets of 60%, 70%, 77% and 85%, as determined in simulated $t\bar{t}$ events. The $b$-tagging distribution obtained by ordering the resulting five exclusive bins from the four WPs from higher to lower $b$-jet efficiency is referred to as “pseudo-continuous” $b$-tagging score, and it is used as input to the multivariate analysis discriminant described in section 5. In this search, a jet is considered $b$-tagged if it passes the WP corresponding to 77% or 60% efficiency to tag a $b$-jet, with a light-jet rejection factor of about 200 or 2500, and a charm-jet ($c$-jet) rejection factor of about 6 or 40, as determined for jets with $p_T > 20$ GeV and $|\eta| < 2.5$ in simulated $t\bar{t}$ events [117]. Correction factors derived from dedicated calibration samples enriched in $b$-jets, $c$-jets, or light jets, are applied to the simulated samples [120–122]. The notation $b^{77\%}$ and $b^{60\%}$ is used to denote the number of $b$-tagged jets with the corresponding WP.

Ambiguities between independently reconstructed electrons, muons and jets can arise. A sequential procedure, referred to as ‘overlap removal’, is performed to resolve these ambiguities and, thus, avoids double counting of particle candidates. This procedure is applied to leptons satisfying the $L$ criteria. If two electrons are closer than $\Delta R = 0.1$, only the one with the higher $p_T$ is considered. If an electron and a muon overlap within $\Delta R = 0.1$, the muon is removed if it is reconstructed from a track and calorimeter deposits.

5 ‘Light jet’ refers to a jet originating from the hadronisation of a light quark ($u$, $d$, $s$) or a gluon.
consistent with a minimum ionising particle, else the electron is removed. If an electron and a selected jet are within $\Delta R < 0.2$ of each other, the jet is removed if it is not a $b$-tagged jet\(^6\) or if it has $p_T > 200$ GeV. Muons are required to be separated by $\Delta R > 0.4$ from any jet that is ghost-associated [123] to it. If the jet satisfying the $\Delta R < 0.4$ requirement is not a $b$-tagged jet and contains less than three tracks with $p_T > 500$ MeV, the overlapping jet is rejected from the event, otherwise, the muon is rejected. A lepton lying within a variable-size cone depending on the lepton $p_T$ and with a maximum radius of $R = 0.4$ around a selected jet that survived all previous overlap criteria is rejected.

The missing transverse momentum $\vec{p}_T^{\text{miss}}$ (with magnitude $E_T^{\text{miss}}$) is defined as the negative vector sum of the $p_T$ of all selected and calibrated objects in the event, including a term to account for the momentum from soft particles that are not associated with any of the selected objects [124]. This soft term is calculated from inner-detector tracks matched to the selected primary vertex, which makes it more resilient to contamination from pile-up interactions. The $E_T^{\text{miss}}$ distribution is used as an input variable to the machine learning training discussed in section 5.

5 Search strategy

Events are required to satisfy a minimal preselection and are categorised into orthogonal signal regions (SRs) based on different criteria such as number of leptons, total lepton charge (indicated by $Q$), and a multi-output deep neural network classifier (DNN\(^{\text{cat}}\)). This categorisation provides a set of regions that are sensitive to all the possible signal production and decay modes considered in this search. A deep neural network is trained in each of the signal regions to discriminate the signal from the backgrounds (DNN\(^{\text{SB}}\)). Additional orthogonal control regions (CRs) are defined in order to fit the normalisation of the main backgrounds. Dedicated kinematic selections are applied to the control regions to improve the purity of the targeted backgrounds. A maximum-likelihood fit is performed across categories to test for a possible signal and constrain in-situ the leading backgrounds simultaneously.

At trigger level, events were selected for read-out using a combination of single-lepton and dilepton triggers, requiring the electrons or muons to satisfy identification criteria similar to those used in the offline reconstruction and isolation requirements [125, 126]. Single-electron triggers require a minimum $p_T$ threshold of 24 (26) GeV in the 2015 (2016, 2017 and 2018) data-taking period(s), while single-muon triggers have a lowest $p_T$ threshold of 20 (26) GeV in 2015 (2016–2018). The dielectron triggers require two electrons with minimum $p_T$ thresholds ranging from 12 GeV in 2015 to 24 GeV in 2015–2018, whereas the dimuon triggers use asymmetric $p_T$ thresholds for leading (subleading) muons: 18 (8) GeV in 2015 and 22 (8) GeV in 2016–2018. Finally, an electron+muon trigger requires events to have an electron candidate with a 17 GeV threshold and a muon candidate with a 14 GeV threshold for all periods.

\(^6\)For the overlap removal, a jet is considered $b$-tagged if it passes the 70% working point. However, the choice of the $b$-tagging working point does not have a sizeable impact on the signal acceptance.


<table>
<thead>
<tr>
<th>Lepton category</th>
<th>2(\ell)SS</th>
<th>3(\ell)</th>
<th>4(\ell)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Lepton definition</td>
<td>((T, T)) with (\geq 1 , b^{60%})</td>
<td></td>
<td>((L, T, M)) with (\geq 1 , b^{60%})</td>
</tr>
<tr>
<td>(p_T^\ell) [GeV]</td>
<td>(20, 20)</td>
<td>(10, 20, 20)</td>
<td>(10, 10, 10, 10)</td>
</tr>
<tr>
<td>(m^{OS-SF}_{\ell^+\ell^-}) [GeV]</td>
<td>(-)</td>
<td>(&gt; 12)</td>
<td>(-)</td>
</tr>
<tr>
<td>(</td>
<td>m^{OS-SF}_{\ell^+\ell^-} - m_Z</td>
<td>) [GeV]</td>
<td>(-)</td>
</tr>
<tr>
<td>(N_{\ell})</td>
<td>(\geq 2)</td>
<td>(\geq 1 , b^{60%})</td>
<td></td>
</tr>
<tr>
<td>(N_{b-\ell})</td>
<td>(\geq 2)</td>
<td>(\geq 1 , b^{60%})</td>
<td></td>
</tr>
</tbody>
</table>

\[ \text{Region split} \quad \text{(sstt, ttq, tttq, ttttt)} \times (Q^+, Q^-) \times (Q^+, Q^-) \times (Q^+, Q^-) \times (Q^+, Q^-) \]

\[ \text{Region naming} \quad \begin{align*}
2\ell\text{SS} &++ \text{ CAT sstt} \\
2\ell\text{SS} &++ \text{ CAT ttq} \\
2\ell\text{SS} &++ \text{ CAT ttt} \\
2\ell\text{SS} &++ \text{ CAT tttq} \\
2\ell\text{SS} &-- \text{ CAT sstt} \\
2\ell\text{SS} &-- \text{ CAT ttq} \\
2\ell\text{SS} &-- \text{ CAT ttt} \\
2\ell\text{SS} &-- \text{ CAT tttq} \\
2\ell\text{SS} &-- \text{ CAT tttt} \\
3\ell &++ \text{ CAT ttq} \\
3\ell &++ \text{ CAT ttt} \\
3\ell &-- \text{ CAT ttq} \\
3\ell &-- \text{ CAT ttt} \\
3\ell &-- \text{ CAT tttq} \\
&\end{align*} \]

\[ \text{4\ell} \]

Table 3. Event selection summary in the signal regions. Leptons are ordered by \(p_T\) in the 2\(\ell\)SS and 4\(\ell\) regions. In the 3\(\ell\) regions the lepton with opposite-sign charge is taken first, followed by the two same-sign leptons in \(p_T\) order. In the lepton selection, \(T, M, L\) stand for Tight, Medium and Loose lepton definitions. In the region naming, the “CAT ttX” denotes the category based on the \(\text{DNN}^{\text{cat}}\) output enriched in the signal process “ttX”. Each of these regions is split according to the lepton charge of the same-sign lepton pair (“++” or “- -”).

For the analysis selection, at least two jets and at least two leptons are required in the event, and leptons are required to match, with \(\Delta R < 0.15\), the corresponding leptons reconstructed by the trigger and to have a \(p_T\) exceeding the trigger \(p_T\) threshold by 1 GeV. Events are required to contain at least one \(b\)-tagged jet with the 60% efficiency working point, or at least two \(b\)-tagged jets with the 77% efficiency working point. If events contain pairs of opposite-sign charge and same-flavour leptons (OS-SF), all pairs are required to satisfy a mass requirement on the dilepton system mass of \(m^{OS-SF}_{\ell^+\ell^-} > 12\) GeV and \(|m^{OS-SF}_{\ell^+\ell^-} - m_Z| > 10\) GeV. Three disjoint event categories are defined according to the number of loose leptons in the event: same-charge dilepton (2\(\ell\)SS), three-lepton (3\(\ell\)), and four-lepton (4\(\ell\)) categories. The four-lepton category is inclusive and contains events with higher lepton multiplicity, while the other two are exclusive. Leptons are ordered by \(p_T\) in the 2\(\ell\)SS and 4\(\ell\) regions. In the 3\(\ell\) regions the lepton with opposite-sign charge is taken first, followed by the two same-sign leptons in \(p_T\) order. The \(p_T\) and identification requirements of each lepton in each category are optimised based on a compromise between non-prompt lepton background suppression and signal acceptance enhancement, and are summarized in table 3.

Multiple control regions (CRs) are defined in order to fit the normalisation of the leading backgrounds. These regions are orthogonal to the signal regions and with one
In order to better target each of the possible signals, a DNN\textsuperscript{cat} is trained to identify each of the five possible production and decay modes of the g2HDM signal. Two DNN\textsuperscript{cat} are another based on different requirements on the lepton working points, dilepton invariant mass, and jet and b-jet multiplicities. Two regions enriched in diboson and ttZ are defined by requiring one OS-SF pair compatible with a Z boson, $|m^{\text{OS-SF}}_{\ell e} - m_Z| < 10$ GeV, differing in the jet multiplicity requirement. Two control regions enriched in photon conversions from $Z \rightarrow \mu\mu\gamma^* (\rightarrow ee)$ are defined, according to the identification of the electron as a material conversion or internal conversion candidate. Finally, six control regions are defined enriched in HF non-prompt leptons, making use of the exclusive lepton identification $M_{\ell e}$ to be orthogonal to the signal regions. Events with two same-sign leptons are split according to the criteria $(T, M_{\ell e})$, $(M_{\ell e}, T)$, $(M_{\ell e}, M_{\ell e})$ for the leading and subleading leptons in $p_T$, and further split according to the fake-lepton-candidate flavour. The fake-lepton candidate is assumed to be the subleading lepton in the $(T, M_{\ell e})$, $(M_{\ell e}, M_{\ell e})$ regions, and the leading lepton in the $(M_{\ell e}, T)$ region. This splitting creates six control regions sensitive to different relative composition of electron and muon non-prompt lepton backgrounds. Additionally, the transverse mass of the leading lepton and the missing transverse energy, $m_T^{\ell 0} (E_T^{\text{miss}})$, defined as $\sqrt{2E_T^{\text{miss}} p_T^{\ell 0} (1 - \cos(\phi^{\text{miss}} - \phi^{\ell 0}))}$, is required to be lower than 250 GeV in the $(T, M_{\ell e})$ and $(M_{\ell e}, T)$ regions to reduce the $tW$ contribution in these CRs. The full definition of the kinematic selection applied to each control region is given in table 4.

Figure 3 illustrates the categorisation and definition of the signal and control regions that are fit simultaneously. The signal contamination is found to be at most 3% of the total prediction in the control regions, assuming $m_H = 400$ GeV and $\rho_t = 0.4$, $\rho_c = 0.2$, and $\rho_{tu} = 0.2$.

In order to better target each of the possible signals, a DNN\textsuperscript{cat} is trained to identify each of the five possible production and decay modes of the g2HDM signal. Two DNN\textsuperscript{cat} are

<table>
<thead>
<tr>
<th>Control regions</th>
<th>WZ</th>
<th>ttZ</th>
<th>Conversions</th>
<th>HF non-prompt</th>
</tr>
</thead>
<tbody>
<tr>
<td>$N_{\ell e}$</td>
<td>2 or 3</td>
<td>$\geq 4$</td>
<td>$\geq 0$</td>
<td>$\geq 2$</td>
</tr>
<tr>
<td>$N_{\ell e}$</td>
<td>$\geq 1$</td>
<td>$\geq 2$</td>
<td>$\geq 0$</td>
<td>$\geq 2$</td>
</tr>
<tr>
<td>Lepton requirement</td>
<td>3$\ell$</td>
<td>$\mu\mu\gamma^*$</td>
<td>$\geq 2$</td>
<td></td>
</tr>
<tr>
<td>Lepton $p_T$ [GeV]</td>
<td>$(T, M_{\ell e})$</td>
<td>$(M_{\ell e}, T)$</td>
<td>$(M_{\ell e}, M_{\ell e})$</td>
<td></td>
</tr>
<tr>
<td>$m^{\text{OS-SF}}_{\ell e} - m_Z$ [GeV]</td>
<td>$&gt; 12$</td>
<td>$&gt; 12$</td>
<td>$-$</td>
<td></td>
</tr>
<tr>
<td>$m^{\text{OS-SF}}_{\ell e} - m_Z$ [GeV]</td>
<td>$&lt; 10$</td>
<td>$&gt; 10$</td>
<td>$-$</td>
<td></td>
</tr>
<tr>
<td>$m_{t\ell} - m_Z$ [GeV]</td>
<td>$-$</td>
<td>$&lt; 10$</td>
<td>$-$</td>
<td></td>
</tr>
<tr>
<td>$m_T^{\ell 0} (E_T^{\text{miss}})$ [GeV]</td>
<td>$-$</td>
<td>$-$</td>
<td>$&lt; 250$</td>
<td></td>
</tr>
</tbody>
</table>

Table 4. Event selection summary in the control regions. The notation $e^*$ is used to denote material conversion or internal conversion candidates, as described in section 4. In the HF non-prompt lepton region naming, “2SStt(e)” (“2SStt(\mu)) refers to the control region enriched in non-prompt electrons (muons) from semileptonic $b$-decays originating mostly from $t\bar{t}$ and with the lepton flavours for the leading and subleading leptons corresponding to “ee, $\mu\mu$” (“$\mu\mu, e\mu$”). The additional $(T, M_{\ell e})$, $(M_{\ell e}, T)$, and $(M_{\ell e}, M_{\ell e})$ subscripts refer to the lepton definitions required for the leading and subleading leptons in each region.
Figure 3. Illustrative sketch of the definition of the signal and control regions. The corresponding observable used in the simultaneous fit, as described in section 8, is given at the bottom of each region box. The notation N_j (N_{bj}) is used to denote selections on the jet (b-jet) multiplicity.

trained individually for the 2\ell^{SS} and 3\ell channels using the KERAS library [127] with TENSORFLOW as a backend [128] and Adam optimiser [129]. Hyperparameters are optimised with the Talos library [130]. The networks consist of nine input features, two dense fully connected layers of 33 nodes with rectified linear units as activation functions, interleaved with a drop-out layer with 20\% rate, and five (three) output nodes with a soft-max activation function for the categorisation of 2\ell^{SS} (3\ell) events. The output categories correspond to the five production modes considered, ignoring in the 3\ell category signals that cannot produce three leptons. Each event is categorised according to the highest class probability. The nine input features are the number of jets, b-tagging score of the three leading jets, sum of b-tagging score of all jets, sum of all pair-wise angular distances between leptons, scalar sum of jet p_T, scalar sum of lepton p_T, and the event E^\text{miss}_T. The network is trained with batch size of 2000 and up to 100 epochs, using all the available signal mass points. To avoid discarding signal events in the evaluation, cross-training is used with the events divided by even/odd event number.

Since several of the probed signal processes are expected to be charge-asymmetric, all the 2\ell^{SS} and 3\ell regions are further split into two categories each, corresponding to the positive and negative total lepton charge selections. Figures 4a and 4b show the normalized distributions of the targeted signals with a scalar mass of 400 GeV or 1000 GeV, compared to the expected background distribution across the various categories described in table 3. At high signal mass, a strong migration is observed from the ttt to the tttq category, due to the high probability of additional radiation. Figures 4c and 4d show the expected fractional signal contribution in each category for the benchmark coupling. The signals originating
Table 5. Input variables to the training of the DNN\textsuperscript{cat} and DNN\textsuperscript{SB} discriminants.

<table>
<thead>
<tr>
<th>Variable</th>
<th>DNN\textsuperscript{cat}</th>
<th>DNN\textsuperscript{SB}</th>
</tr>
</thead>
<tbody>
<tr>
<td>Number of jets ($N_{\text{jets}}$)</td>
<td>✓</td>
<td>✓</td>
</tr>
<tr>
<td>Sum of pseudo-continuous $b$-tagging scores of jets</td>
<td>✓</td>
<td>✓</td>
</tr>
<tr>
<td>Pseudo-continuous $b$-tagging score of 1st, 2nd, 3rd leading jet in $p_T$</td>
<td>✓</td>
<td>✓</td>
</tr>
<tr>
<td>Sum of $p_T$ of the jets and leptons ($H_{T,\text{jets}}$, $H_{T,\text{lep}}$)</td>
<td>✓</td>
<td>✓</td>
</tr>
<tr>
<td>Angular distance of leptons (sum in the case of 3$\ell$ and 4$\ell$)</td>
<td>✓</td>
<td>✓</td>
</tr>
<tr>
<td>Missing transverse energy</td>
<td>✓</td>
<td>✓</td>
</tr>
<tr>
<td>Leading transverse momentum of jet</td>
<td>✓</td>
<td>✓</td>
</tr>
<tr>
<td>Invariant mass of leading lepton and missing transverse energy</td>
<td>✓</td>
<td>✓</td>
</tr>
<tr>
<td>Di/tri/quad-lepton type variable (associated with the number of electrons/muons in event)</td>
<td>-</td>
<td>✓</td>
</tr>
</tbody>
</table>

from top-Higgs associated production (ttq and ttt) are expected to dominate across all regions, including the categories designed to target other processes, due to the much larger production cross section. This contribution is however strongly dependent on the coupling choice. For the benchmark coupling of $\rho_{tt} = 0.4$, $\rho_{tc} = 0.2$, $\rho_{tu} = 0.2$, decays to top-quark pairs dominate when not suppressed by the available phase-space.

A total of 27 analysis regions are defined, with 17 signal regions (10 with 2$\ell$SS, 6 with 3$\ell$, and one 4$\ell$) and 10 control regions. In each region, a given kinematic variable is fit to improve the sensitivity to the targeted signal process (signal regions) or to improve the modelling of a particular background process (control regions).

A DNN\textsuperscript{SB} classifier is trained in each signal region to separate the targeted signal from the sum of backgrounds. The networks consist of 12 input features, two dense fully connected layers of 36 and 48 nodes respectively with sigmoid activation functions, interleaved with a drop-out layer with 40% rate, and one output node with a sigmoid activation function. The 12 input features are the leading jet $p_T$, number of muons, transverse mass of leading lepton and $E_T^{\text{miss}}$ system, and the nine variables that are used in the DNN\textsuperscript{cat}. Table 5 summarises the input variables used for each multivariate discriminant. To achieve good sensitivity over the large range of masses that are tested, the output of the classifier is decorrelated from the signal mass introducing an additional term to the loss function via distance correlation [131, 132]. A hyperparameter $\lambda$ controls the weight of the additional penalty term, with a value of $\lambda = 0.5$. The value was optimised to achieve a minimal signal mass dependence without compromising the discrimination power. A separate training is performed in each lepton category and signal category. The same DNN\textsuperscript{SB} is used in both positive- and negative-charge regions. The same DNN\textsuperscript{cat} and DNN\textsuperscript{SB} classifiers trained under the g2HDM signal hypothesis are also used in the interpretation based on the SUSY models. Figure 5 shows the DNN\textsuperscript{SB} distribution of the targeted signal in each signal-enriched category, the total signal, and the background in the 2$\ell$SS ++ CAT stt, 2$\ell$SS ++ CAT ttq, 3$\ell$ ++ CAT ttt, 3$\ell$ ++ CAT tttq, 3$\ell$ ++ CAT tttt, and 4$\ell$ categories.

In the diboson and $t\bar{t}Z$ control regions the fitted variable is the $b$-jet multiplicity, $N_{b-\text{jets}}$, where the distribution is binned with an upper limit of $\geq 2$ $b$-jets and $\geq 3$ $b$-jets.
Figure 4. Distribution of signal processes and total background between the different DNN categories (top row), for (a) 400 GeV and (b) 1000 GeV scalar masses. All distributions are normalised to unity. The vertical dashed grey lines separate categories targeting each of the main signal processes: stt, ttq, ttt, tttq, and tttt. Signal contributions below 2% in a single bin are omitted for clarity. The expected fractional signal contribution in each category (bottom row) is shown for (c) 400 GeV and (d) 1000 GeV scalar masses for the coupling set $\rho_{tt} = 0.4$, $\rho_{tc} = 0.2$, $\rho_{tu} = 0.2$, respectively. The subleading lepton $p_T$ spectrum is used in the HF non-prompt control regions. Finally, the total event yield is fit in the control regions enriched in electrons from photon conversion.

6 Background estimation

The background processes passing the signal region selections are categorised into irreducible and reducible backgrounds. Irreducible backgrounds (section 6.1) produce prompt leptons in their decay, i.e., leptons originating from $W/Z$ boson decays, leptonic $\tau$-lepton decays, or internal conversions. Reducible backgrounds (section 6.2) have prompt leptons with misassigned charge or at least one non-prompt lepton.

Except for the background from electrons with misassigned charge (denoted as QMisID), all other backgrounds are estimated using the simulated samples described in section 3.
Figure 5. Comparison of the DNN$^{SB}$ distribution of the targeted signal (solid line), the total signal (dashed line) and the background (filled grey area) in the (a) 2$\ell$ SS CAT sstt, (b) 2$\ell$ SS CAT ttq, (c) 3$\ell$ CAT ttt, (d) 3$\ell$ CAT ttq, (e) 3$\ell$ CAT tttt, and (f) 4$\ell$ categories, for an assumed $m_{H_1} = 400$ GeV (pink) and $m_{H_1} = 800$ GeV (violet), with couplings $\rho_{tt} = 0.4$ and $\rho_{tq} = 0.2$. All distributions are normalised to unity.

In some cases, the simulation is improved using additional corrections derived from data control samples before the simultaneous fit to data. In particular, the event kinematics of the simulated $t\bar{t}$ and $VV$ backgrounds require dedicated corrections to better describe the data. In addition, the yields of some simulated backgrounds, in particular $t\bar{t}W$, $t\bar{t}Z$, $VV$ and non-prompt-lepton backgrounds, are adjusted via normalisation factors that are determined by performing a likelihood fit to data across all event categories (signal and control regions as defined in tables 3 and 4) as discussed in section 8.

6.1 Irreducible backgrounds

Background contributions with prompt leptons originate from a wide range of physics processes with the relative importance of individual processes varying by channel. The main irreducible backgrounds originate from $t\bar{t}W$, $t\bar{t}t$, and $t\bar{t}Z/\gamma^*$ production, followed by
VV (in particular WZ) and $t\bar{t}h$ production, and have final states and kinematic properties similar to the g2HDM signal. Smaller contributions originate from the following rare processes: $tZ$, $tW$, $tWZ$, $t\bar{t}WW$, $VVV$, and $t\bar{t}t$ production.

6.1.1 $t\bar{t}W$ background

The $t\bar{t}W$ background represents the leading background in several event categories. Despite the use of state-of-the-art simulations, accurate modelling of additional QCD and QED radiation in $t\bar{t}W$ production remains challenging. Given the excellent discriminating power of the DNN_{SB} in the signal regions, the events at lower values of the DNN_{SB} score are enriched in and sensitive to the $t\bar{t}W$ background. Additionally, the signal regions in the 2$\ell$ and 3$\ell$ categories are split by the sign of the total lepton charge ($Q$) to better discriminate some g2HDM signal processes and the $t\bar{t}W$ process, which have a large charge asymmetry, from other SM backgrounds that are charge symmetric. This discrimination increases the sensitivity to this background in the simultaneous fit. Finally, the DNN_{cat} categories with negative total lepton charge, which are depleted in signals with large charged lepton asymmetry, provide additional constraints on the $t\bar{t}W$ background, in particular at high values of the DNN_{SB} distribution tail.

Disagreement between the data and the prefit prediction from the simulation is observed, which is accommodated by an overall normalisation factor that is assigned to the $t\bar{t}W$ background, and that is determined during the likelihood fit. The measured normalisation factor for the background-only hypothesis is $\hat{\lambda}_{t\bar{t}W} = 1.50 \pm 0.14$, which is compatible with that determined in the SM $t\bar{t}t\bar{t}$ analysis [34], and with a previous measurement of the $t\bar{t}W$ production cross section [33].

6.1.2 VV and $t\bar{t}Z/\gamma^*$ backgrounds

A data-driven correction to the jet multiplicity spectrum is derived from an inclusive trilepton diboson-enriched region with zero $b$-jets defined with the 85% WP for $b$-jet efficiency and at least one jet (denoted as 3$\ell$VV0b region). The events are required to have three leptons that satisfy the same selection as in the 3$\ell$VV CR.

Figure 6a shows the jet multiplicity distribution in the 3$\ell$VV0b region before the correction. After the correction is applied to VV, a good modelling of the $N_{\text{jets}}$ distribution is found in a 3$\ell$ region with at least one jet and exactly one $b$-jet defined with the 60% WP, as shown in figure 6b.

The 3$\ell$VV and 3$\ell$t$tZ$ CRs are used in the likelihood fit to improve the prediction of the background contribution from the VV and $t\bar{t}Z/\gamma^*$ processes; these processes have purities of 15% and 75% in the corresponding CRs. The numbers of jets and $b$-jets provide good discrimination between these two processes and are used to build the control regions (number of jets) and as variables in the fit (number of $b$-jets). The measured normalisation factors for the background-only hypothesis are: $\hat{\lambda}_{VV} = 0.85 \pm 0.30$ and $\hat{\lambda}_{t\bar{t}Z} = 0.97 \pm 0.19$, where the uncertainty includes both statistical and systematic contributions.

Figures 7a and 7b show the $b$-jet multiplicity distribution in the 3$\ell$VV and 3$\ell$t$tZ$ CRs after the likelihood fit to data.
Figure 6. Comparison between data and the background prediction for the distribution of (a) the number of jets in the $3\ell VV0b$ region before the $VV$ jet multiplicity correction and (b) the number of jets in a $3\ell$ region with at least one jet and exactly one $b$-jet defined with the 60% WP after the $VV$ jet multiplicity correction. The ratio of the data to the background prediction (“Pred.”) is shown in the lower panel. The size of the combined statistical and systematic uncertainty in the background prediction is indicated by the blue hatched band. The last bin in each figure contains the overflow.

6.1.3 Other irreducible backgrounds

The rate of the background from internal conversions with $m(e^+e^-) < 1$ GeV is estimated using the two dedicated CRs, $3\ell$IntC and $3\ell$MatC, with a purity of 98% and 30%, respectively. The total yield in each category is used in the likelihood fit to determine the normalisation factor, which is measured for the background-only hypothesis to be $\hat{\lambda}_{\text{IntC}} = 1.06 \pm 0.23$, where the uncertainty is dominated by the statistical uncertainty.

6.2 Reducible backgrounds

6.2.1 Non-prompt leptons

Non-prompt leptons originate from material conversions, heavy-flavour hadron decays, or the improper reconstruction of other particles, with an admixture that depends strongly on the lepton quality requirements and varies across event categories. These backgrounds are small in all $2\ell$ and $3\ell$ SRs and thus are estimated from simulation, with the normalisation determined by the likelihood fit. The non-prompt lepton background contribution in the $4\ell$ SR is very small and is therefore taken from simulation without dedicated data-driven corrections. The main contribution to the non-prompt-lepton background is from $t\bar{t}$ production, with much smaller contributions from $V+$jets and single-top-quark processes. The non-prompt leptons in the simulated samples are labelled according to whether they originate from heavy-flavour (HF) or light-flavour (LF) hadron decays, or from a material
Figure 7. Comparison between data and the background prediction for the distribution of the b-jet multiplicity in the (a) 3ℓVV and (b) 3ℓttZ CRs after the VV jet multiplicity correction. The background contributions after the likelihood fit to data (“Post-Fit”) for the background-only hypothesis are shown as filled histograms. The ratio of the data to the post-fit background prediction (“Pred.”) is shown in the lower panel, separately for post-fit background (black points) and pre-fit background (dashed blue line). The size of the combined statistical and systematic uncertainty in the background prediction is indicated by the blue hatched band. The last bin in each figure contains the overflow.

conversion candidate (Mat. Conv.). The HF category includes leptons from both bottom and charm decays.

Two corrections are applied to the tt and the overall non-prompt lepton background simulation before the fit. First, the tt + ≥ 1 b-jet contribution from simulation is known to be mismodeled and is therefore corrected by a factor of 1.3 as measured by a previous ATLAS analysis sensitive to the in-situ measurement of this contribution in the single- and opposite-sign dilepton final states [133]. This correction is well motivated since the mismodelling of additional b-jets in tt is not expected to depend on the presence of additional non-prompt leptons in the event. Second, the shape of the b-jet multiplicity in the non-prompt lepton background simulation is corrected for electrons and muons separately to match data in an orthogonal 2ℓSS validation region enriched with non-prompt leptons, where one of the leptons must satisfy a looser requirement on the non-prompt lepton BDT score but fail the M lepton WP criteria.

Several of the event categories introduced in section 5 were designed to be enriched in specific processes and are used to derive normalisation factors to improve their modelling by the simulation. The 3ℓMatC CR is enriched in material conversions with a purity of 70% and only the total event yield is used. There are six 2ℓ CRs enriched in contributions from HF non-prompt leptons in tt events, i.e. 2ℓtt(e)(T,Mex), 2ℓtt(e)(Mex,T), 2ℓtt(e)(Mex,Mex),
Figure 8. Comparison between data and the background prediction for the distribution of the transverse momentum of the sub-leading lepton ($p_T$, subleading lep) in (a) the $2\ell tt(e) (T,M_{ex})$ CR and (b) the $2\ell tt(\mu) (T,M_{ex})$ CR. The background contributions after the likelihood fit to data (“Post-Fit”) for the background-only hypothesis are shown as filled histograms. The ratio of the data to the post-fit background prediction (“Pred.”) is shown in the lower panel, separately for post-fit background (black points) and pre-fit background (dashed blue line). The size of the combined statistical and systematic uncertainty in the background prediction is indicated by the blue hatched band. The last bin in each figure contains the overflow.

$2\ell tt(\mu) (T,M_{ex})$, $2\ell tt(\mu) (M_{ex},T)$, and $2\ell tt(\mu) (M_{ex},M_{ex})$. In these CRs, the transverse momentum of the fake-lepton-candidate distribution is used to be able to correct for a possible mismodelling in the $p_T$ of the non-prompt lepton. The fake-lepton candidate is assumed to be the subleading lepton in the $(T,M_{ex})$, $(M_{ex},M_{ex})$ regions, and the leading lepton in the $(M_{ex},T)$ region. The event requirement to have at least one $M_{ex}$ lepton provides separation from the irreducible backgrounds, in particular $t\bar{t}W$, and thus increases the sensitivity to the HF non-prompt electron and muon contributions. Normalisation factors for three non-prompt-lepton background contributions are estimated from the likelihood fit. The normalisation factor for HF non-prompt leptons is estimated separately for electrons and muons, $\lambda_{e}^{\text{had}}$ and $\lambda_{\mu}^{\text{had}}$, respectively. An additional normalisation factor is determined for the material conversions background, $\lambda_{e}^{\text{Mat Conv}}$. The measured normalisation factors for the background-only hypothesis are: $\hat{\lambda}_{e}^{\text{had}} = 1.05 \pm 0.31$, $\hat{\lambda}_{\mu}^{\text{had}} = 0.92 \pm 0.18$, and $\hat{\lambda}_{e}^{\text{Mat Conv}} = 1.16 \pm 0.29$, where the uncertainties include systematic effects but are dominated by the statistical uncertainty.

Figures 8a and 8b display the fake-lepton-candidate $p_T$ distribution in the $2\ell tt(e) (T,M_{ex})$ and $2\ell tt(\mu) (T,M_{ex})$ CRs after the likelihood fit to data. As shown in the figures, the purity of HF non-prompt lepton background is 45% and 55%, respectively, which was possible to achieve with the usage of the exclusive $M_{ex}$ lepton working point.
6.2.2 Charge misassignment

Backgrounds with leptons with the charge incorrectly assigned affect primarily the $2\ell$ channel and predominantly arise from $t\bar{t}$ production, where one electron undergoes a hard bremsstrahlung and an asymmetric conversion ($e^\pm \rightarrow e^\pm \gamma \rightarrow e^\pm e^\mp e^\pm$) or a mismeasured track curvature. The muon charge misassignment rate is negligible in the $p_T$ range relevant to this analysis. The electron charge misassignment rate is measured in data using samples of $Z \rightarrow e^+e^-$ events reconstructed as same-charge pairs and as opposite-charge pairs, with the background subtracted via a sideband method [105].

The charge misassignment rate is parameterised as a function of electron $p_T$ and $|\eta|$. It varies from about $10^{-5}$ for low-$p_T$ electrons ($17 \leq p_T \leq 50$ GeV) that satisfy $|\eta| \leq 1.37$, to about $4 \times 10^{-3}$ for high-$p_T$ electrons ($p_T \geq 100$ GeV) in the region $2 \leq |\eta| \leq 2.47$. To estimate the QMisID background in each of the corresponding event categories, the measured charge misassignment rate is then applied to data events satisfying the requirements of the $2\ell$ channels, except that the two leptons are required to be of opposite charge.

6.3 Background modelling

The modelling of some representative variables at preselection level is showed in figure 9. The background prediction includes all the corrections previously described as well as the normalisation factors determined through a likelihood fit to data as discussed in section 8. Good modelling is observed across all variables and lepton categories.

7 Systematic uncertainties

The signal and background yields in each signal and control region may be affected by several sources of systematic uncertainty, described in the following subsections. Given the low background yields and good signal-to-background separation provided by the final discriminating variable used in the signal-rich event categories, the search sensitivity is determined by the limited number of data events rather than by the systematic uncertainties on the background estimate. The final uncertainty in the background estimate in the SRs is dominated by the uncertainty in the fitted background normalisations, in particular $t\bar{t}W$.

A summary of all systematic uncertainties included in the analysis is given in table 6.

7.1 Experimental uncertainties

The combined 2015–2018 integrated luminosity, obtained using the LUCID-2 detector [134] for the primary luminosity measurements, has an uncertainty of 1.7% [68].

Uncertainties associated with the lepton selection arise from the trigger, reconstruction, identification and isolation efficiencies, and lepton momentum scale and resolution [105, 107, 110]. Uncertainties associated with the jet selection arise from the jet energy scale (JES), the JVT requirement and the jet energy resolution (JER) [112, 135].

The efficiency of the flavour-tagging algorithm is measured for each jet flavour using control samples in data and in simulation. From these measurements, correction factors are derived to correct the tagging rates in the simulation [120, 121, 136]. These systematic uncertainties are taken as uncorrelated between $b$-jets, $c$-jets, and light-flavour jets.
Figure 9. Comparison between data and the background prediction for the distribution of the jet multiplicity (a, d), sum of pseudo-continuous $b$-tagging scores of jets (b, e) and the sum of $p_T$ of the jets and leptons (c, f), in the $2\ell$ (a-c) and $3\ell$ (d-f) signal region selections before any categorisation based on DNN$^{\text{cat}}$. The expected signal for $m_H = 900$ GeV and couplings $\rho_{tt}=0.6$, $\rho_{tc}=0.0$, and $\rho_{tu}=1.1$, along with the background contributions, is shown after the likelihood fit to data (“Post-Fit”) for the background-only hypothesis. The ratio of the data to the prediction (“Pred.”) is shown in the lower panel, separately for post-fit signal-plus-background (black points) and pre-fit background (dashed blue line). The size of the combined statistical and systematic uncertainty in the signal-plus-background prediction is indicated by the blue hatched band. The last bin in each figure contains the overflow.
An additional uncertainty is assigned to account for the extrapolation of the $b$-tagging efficiency measurement from the $p_T$ region used to determine the correction factors to regions with higher transverse momentum [137]. This uncertainty is the leading experimental uncertainty in the analysis, with relative variations up to 25% on the signal.

The treatment of the uncertainties associated with reconstructed objects is common to all analysis channels and applies to all signal and background samples, except for the background from electrons with misidentified charge, and thus these are considered as fully correlated among different analysis regions and samples.

7.2 Theoretical uncertainties

The modelling uncertainties on the main irreducible backgrounds are assessed through comparisons with alternative MC samples, as listed in table 1. Additional uncertainties are evaluated from renormalisation and factorisation scale variations by a factor of 0.5 and 2, relative to the nominal scales, for the $t\bar{t}W$, $t\bar{t}Z$, and diboson samples. An additional 20% uncertainty is assigned to the $t\bar{t}W$ electroweak contribution [138]. An additional 50% uncertainty is assigned to $t\bar{t}W$, $t\bar{t}Z$, and $t\bar{t}$ events with additional heavy-flavour jets, following ref. [34]. This normalisation uncertainty is not applied to diboson events with heavy-flavour since its normalisation is fit to data. The statistical uncertainty on the fitted parameters for the $VV$ jet-multiplicity correction is propagated as an uncertainty on the diboson background. The leading theoretical uncertainties arise from $t\bar{t}W$ modelling and additional heavy-flavour uncertainties.

Finally, additional normalisation uncertainties are included for all processes whose normalisation is not obtained from the fit. The $t\bar{t}t\bar{t}$, $t\bar{t}h$, and $tZ$ processes are assigned uncertainties of 20% [93], 11% [95], and 5% [139], respectively. As a conservative estimate, a 50% cross section uncertainty is assigned to the $t\bar{t}t$, $tWZ$, $t\bar{t}WW$, and triboson backgrounds, which are small backgrounds with low impact on the search.

Uncertainties on the modelling of the signal samples are evaluated through independent variations of the factorisation and renormalisation scales by a factor of two. Additional uncertainties due to PDF effects are estimated through an ensemble of eigenvariations of the NNPDF set, and by taking the differences with respect to alternative PDF sets [140].

7.3 Reducible background uncertainties

The normalisation of HF non-prompt leptons is obtained from regions including at least one $M_{ex}$ lepton and extrapolated to the signal regions where the same-sign leptons fulfil the $T$ or $M$ identification requirements. An uncertainty of 20% on the extrapolation from $M_{ex}$ to $T$ leptons is applied from the comparison of the relative efficiency between nominal and alternative $t\bar{t}$ MC samples. An additional 50% uncertainty is assigned to events originating from $t\bar{t} + \geq 1b$ and $t\bar{t} + \geq 1c$, decorrelated between flavours. Validation regions with looser lepton requirements and further enriched in non-prompt leptons are defined. A good agreement between data and background prediction is observed in all kinematic variables except for the number of $b$-jets. Based on this disagreement, an $N_{b-jets}$-dependent uncertainty is added to the HF non-prompt background ranging from 6%–40%
<table>
<thead>
<tr>
<th>Systematic uncertainty</th>
<th>Components</th>
<th>Systematic uncertainty</th>
<th>Components</th>
</tr>
</thead>
<tbody>
<tr>
<td>Signal modelling</td>
<td></td>
<td>Luminosity</td>
<td>1</td>
</tr>
<tr>
<td>QCD scale</td>
<td>1</td>
<td>Pile-up reweighting</td>
<td>1</td>
</tr>
<tr>
<td>PDFs+αS</td>
<td>3</td>
<td><strong>Physics objects</strong></td>
<td></td>
</tr>
<tr>
<td>tW modelling</td>
<td></td>
<td>Electron</td>
<td>6</td>
</tr>
<tr>
<td>QCD scale</td>
<td>3</td>
<td>Muon</td>
<td>15</td>
</tr>
<tr>
<td>Generator</td>
<td>2</td>
<td>Electron Non-prompt BDT</td>
<td>14</td>
</tr>
<tr>
<td>Electroweak cross section</td>
<td>1</td>
<td>Muon Non-prompt BDT</td>
<td>20</td>
</tr>
<tr>
<td>Additional heavy-flavour</td>
<td>1</td>
<td>Jet energy scale</td>
<td>30</td>
</tr>
<tr>
<td>tZ/γ* modelling</td>
<td></td>
<td>Jet energy resolution</td>
<td>12</td>
</tr>
<tr>
<td>QCD scale</td>
<td>2</td>
<td>Jet vertex fraction</td>
<td>1</td>
</tr>
<tr>
<td>Generator</td>
<td>2</td>
<td>Jet flavour tagging</td>
<td>62</td>
</tr>
<tr>
<td>Additional heavy-flavour</td>
<td>1</td>
<td>Emiss</td>
<td>3</td>
</tr>
<tr>
<td><strong>WZ modelling</strong></td>
<td></td>
<td><strong>Data-driven reducible background</strong></td>
<td>165</td>
</tr>
<tr>
<td>Cross section (N)</td>
<td>1</td>
<td>Material conversions modelling</td>
<td>1</td>
</tr>
<tr>
<td>Parton shower and hadronisation model</td>
<td>1</td>
<td>Internal conversions modelling</td>
<td>1</td>
</tr>
<tr>
<td>Generator</td>
<td>1</td>
<td>Charge misassignment</td>
<td>1</td>
</tr>
<tr>
<td>QCD scale</td>
<td>1</td>
<td>HF non-prompt</td>
<td>8</td>
</tr>
<tr>
<td>Additional heavy-flavour</td>
<td>1</td>
<td>tt additional heavy-flavour</td>
<td>2</td>
</tr>
<tr>
<td><strong>ttH modelling</strong></td>
<td></td>
<td>Total (Data-driven reducible background)</td>
<td>13</td>
</tr>
<tr>
<td>Cross section (N)</td>
<td>1</td>
<td>Total (Overall)</td>
<td>210</td>
</tr>
<tr>
<td>Extra-jets correction</td>
<td>1</td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Other background modelling</strong></td>
<td></td>
<td><strong>Total (Signal and background modelling)</strong></td>
<td>32</td>
</tr>
<tr>
<td>Generator</td>
<td>1</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Cross section (N)</td>
<td>1</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Cross section (N)</td>
<td>6</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

**Table 6.** Sources of systematic uncertainty considered in the analysis. “N” means that the uncertainty is taken as normalisation-only for all processes and channels affected. Some of the systematic uncertainties are split into several components, as indicated by the number in the rightmost column.

...for 1–3 additional b-jets in the non-prompt muon regions, and 10%–80% in the non-prompt electron regions.

The modelling of internal and material conversions is tested in dedicated validation regions with two tight same-sign leptons, requiring one of them to be a conversion candidate. Additional uncertainties of 10% and 50% are assigned to the material and internal conversion backgrounds, respectively, evaluated from the data to background agreement in the validation regions.

A systematic uncertainty of 10%–60% is assigned to the background from electrons with misidentified charge. The uncertainty increases with electron $p_T$ and decreases with $|\eta|$. The uncertainty is assessed combining the uncertainties from the measurement of the charge misassignment rate, the difference in rates from varying the $m_Z$ window selection, and the different rates measured in data and $Z \rightarrow e^+e^-$ MC.
8 Results

A maximum-likelihood fit is performed on all bins in the 27 signal and control regions considered in this search to simultaneously determine the background and the signal yields that are most consistent with the data. The DNN$^{SB}$ is used as the discriminating variable in the signal regions, whereas the $N_{b-jets}$, fake-lepton-candidate $p_T$ and event yields are fit in the control regions. The sum of all the g2HDM signal processes studied here (sstt, ttq, ttt, tttq, tttt) is considered as a single signal template and its acceptance in each category is predicted by the simulation. The same procedure is followed for the SUSY signals.

The likelihood function $L(\mu, \lambda, \theta)$ is constructed as a product of Poisson probability terms over all bins considered in the search, and depends on the signal-strength parameter, $\mu$, a multiplicative factor applied to the predicted yield for the g2HDM signal (depending on the coupling configuration $\rho_{tt}$, $\rho_{tc}$, $\rho_{tu}$ and on the assumed mass $m_H$), $\lambda$, the normalisation factors for several backgrounds (see section 6), and $\theta$, a set of nuisance parameters (NP) encoding systematic uncertainties in the signal and background expectations [141]. Systematic uncertainties can impact the estimated signal and background rates, the migration of events between categories, and the shape of the fitted distributions; they are summarised in table 6. Both $\mu$ and $\lambda$ are treated as free parameters in the likelihood fit. The NPs $\theta$ allow variations of the expectations for signal and background according to the systematic uncertainties, subject to Gaussian or Poisson constraints in the likelihood fit. Their fitted values represent the deviations from the nominal expectations that globally provide the best fit to the data. Statistical uncertainties in each bin due to the limited size of the simulated samples are taken into account by dedicated parameters using the Beeston-Barlow “lite” technique [142].

The test statistic $q_\mu$ is defined as the profile likelihood ratio: $q_\mu = -2 \ln(\mathcal{L}(\mu, \tilde{\lambda}_\mu, \tilde{\theta}_\mu)/\mathcal{L}(\hat{\mu}, \hat{\lambda}_\mu, \hat{\theta}_\mu))$, where $\hat{\mu}$, $\hat{\lambda}_\mu$, and $\hat{\theta}_\mu$ are the values of the parameters that maximise the likelihood function, and $\tilde{\lambda}_\mu$ and $\tilde{\theta}_\mu$ are the values of the parameters that maximise the likelihood function for a given value of $\mu$. The test statistic $q_\mu$ is evaluated with the RooFit package [143]. A related statistic is used to determine the probability that the observed data are incompatible with the background-only hypothesis (i.e. the discovery test) by setting $\mu = 0$ in the profile likelihood ratio ($q_0$). The $p$-value (referred to as $p_0$) representing the probability of the data being compatible with the background-only hypothesis is estimated by integrating the distribution of $q_0$ from background-only pseudo-experiments, approximated using the asymptotic formulae given in ref. [144], above the observed value of $q_0$. Some model dependence exists in the estimation of the $p_0$, as a given signal scenario must be assumed in the calculation of the denominator of $q_0$, even if the overall signal normalisation is allowed to float and is fit to data. The observed $p_0$ is checked for each explored signal scenario. Upper limits on the signal production cross section for each of the signal scenarios considered are derived by using $q_\mu$ in the CL$_S$ method [145, 146]. For a given signal scenario, values of the production cross section (parameterised by $\mu$) yielding CL$_S < 0.05$, where CL$_S$ is computed using the asymptotic approximation [144], are excluded at $\geq 95\%$ confidence level (CL).
The smallest $p_0$ value is observed when assuming a signal with $m_H = 900$ GeV and couplings $\rho_t=0.6$, $\rho_c=0.0$, and $\rho_u=1.1$, corresponding to a local significance of 2.8 standard deviations. The fitted signal strength is $\mu = 0.07 \pm 0.03$, pointing to an incompatibility of the model prediction with the size of the excess or else the need for additional undetected decay modes taking up 93% of the branching ratio. The signal cross section resulting from the fit to data for this g2HDM signal hypothesis is 154 fb, with fractional contributions of 55% ttq, 31% stt, and 14% ttt. The signal with the fitted signal strength closest to unity ($\mu = 0.9 \pm 0.4$) corresponds to $m_H = 900$ GeV and couplings $\rho_t=0.2$, $\rho_c=0.4$, and $\rho_u=0.4$ and a local significance of 2.4 $\sigma$. Figure 10 shows the local significance as a function of the three couplings normalised to the sum of the couplings. This normalisation eliminates one degree of freedom related to the total normalisation of the signal, which is not relevant for the computation of the significance. However, a residual dependency on the actual value of the coupling remains as the normalization of the stt process scales as the fourth power of the couplings, while the rest of the processes scale as a function of the couplings squared. The star indicates the coupling configuration leading to the highest observed significance of 2.8 standard deviations.

A comparison of the distributions of observed and expected yields is shown figure 11a for the 17 SRs, and figure 11b for the 10 CRs, after the combined likelihood fit for the signal-plus-background hypothesis. The corresponding post-fit yields for the SRs can be found in tables 7, 8, and 9 for the $2\ell$SS positively charged, $2\ell$SS negatively charged, and $3\ell$ and $4\ell$ SRs, respectively. The signal shown in the figures and tables is the g2HDM signal
Figure 11. Comparison between data and the background prediction for the event yields in (a) the 17 signal region categories and (b) the 10 control region categories. The expected signal for $m_H = 900$ GeV and couplings $\rho_{tt}=0.6$, $\rho_{tc}=0.0$, and $\rho_{tu}=1.1$, along with the background contributions, is shown after the likelihood fit to data (“Post-Fit”) for the signal-plus-background hypothesis. The total background prediction before the likelihood fit to data (“Pre-Fit”) is shown as a dashed blue histogram in the upper panel. The ratio of the data to the total prediction is shown in the lower panel, separately for post-fit signal-plus-background (black points) and pre-fit background (dashed blue line). The size of the combined statistical and systematic uncertainty in the background prediction is indicated by the blue hatched band.
with couplings $\rho_{tt}=0.6$, $\rho_{tc}=0.0$, and $\rho_{tu}=1.1$, and mass of 900 GeV, which corresponds to the largest observed significance above the background only hypothesis. Good agreement between the data and fitted signal-plus-background yields is found across all event categories.

The systematic uncertainties with the largest impact on the signal strength originate from the modelling of $t\bar{t}W$ with and without additional heavy flavour jets, $t\bar{t}Z$, $t\bar{t}h$, and $t\bar{t}t\bar{t}$ processes. The signal strength is partially anti-correlated with $\lambda_{t\bar{t}W}$, with a linear correlation value of $-35\%$. The search is dominated by statistical uncertainties.

Comparisons between data and the background prediction for the DNN$^{SB}$ distributions used in the different SRs are shown in figures 12 and 13. The binning used for the DNN$^{SB}$ distributions in the different SRs represents a compromise between preserving enough discrimination in the fit between the background and the signal for the different values of the heavy $H$ mass considered and keeping the MC statistical uncertainty of the background prediction per bin well below 30%. The signal regions with the largest pre-fit tension between data and the background yields (shown in the blue dashed line) at high values of the DNN$^{SB}$ are the $2\ell SS++$ CAT tttq, the $2\ell SS++$ CAT tttt, the $2\ell SS++$ CAT sstt, and the $2\ell SS++$ CAT ttq regions. Within this model, the $\hat{\lambda}_{t\bar{t}W}$ remains higher than 1, as observed by other analyses [33, 42, 147–149]. Since the largest discrepancies between data and the background expectation before the fit are observed in signal categories with positive total lepton charge, this tension cannot be explained by the lepton-charge-symmetric SM $t\bar{t}t\bar{t}$ production. The goodness-of-fit based on the saturated model [150] for the best fit g2HDM signal plus background is 62%, which shows a better fit to data than the background-only hypothesis with a goodness-of-fit of 45%.

Exclusion limits on the heavy Higgs boson mass are set for different choices of the couplings, as shown in figure 14. Masses of an additional scalar boson $m_H$ between 200–620 (200–840) GeV with couplings $\rho_{tt} = 0.4$, $\rho_{tc} = 0.2$, and $\rho_{tu} = 0.2$ are observed (expected) to be excluded at 95% confidence level. Limits on the heavy Higgs boson mass are also set for a scenario without coupling to two top quarks, $\rho_{tt} = 0$, $\rho_{tc} = 0.2$, $\rho_{tu} = 0.2$, resulting on an observed (expected) limit of 200–320 (200–560) GeV on the heavy Higgs boson mass. No limits can be set on scenarios without off-diagonal couplings, leading only to four-top final states with a coupling set $\rho_{tt} = 1$, $\rho_{tc} = 0$, $\rho_{tu} = 0$. The sensitivity of the analysis on the four-top final state is similar to previous ATLAS analyses [39]. The excluded mass is also presented as a function of two couplings for different assumptions, as shown in figure 15.

The search is also used to set limits on RPV SUSY models using the existing DNNs that were trained for the g2HDM model. Figures 16a and 16b show the exclusion limits obtained on the higgsino and wino models, respectively. Higgsinos (winos) with masses up to 585 (670) GeV are excluded. Figure 17 shows limits on the smuon-bino model. Smuon masses up to 460 GeV are excluded, with weaker exclusion limits for small mass splittings between the smuon and the lightest SUSY particle (LSP), or for LSP masses close to the top-quark threshold.
Figure 12. Comparison between data and prediction for the DNN$^{SB}$ distribution used in different signal region categories of the 2$\ell$SS channel: (a) 2$\ell$SS $^{--}$ CAT sstt, (b) 2$\ell$SS $^{++}$ CAT sstt, (c) 2$\ell$SS $^{--}$ CAT ttq, (d) 2$\ell$SS $^{++}$ CAT ttq, (e) 2$\ell$SS $^{--}$ CAT ttt, (f) 2$\ell$SS $^{++}$ CAT ttt, (g) 2$\ell$SS $^{--}$ CAT tttq, (h) 2$\ell$SS $^{++}$ CAT tttq, (i) 2$\ell$SS $^{--}$ CAT tttt, and (j) 2$\ell$SS $^{++}$ CAT tttt. The expected signal for $m_H = 900$ GeV and couplings $\rho_{tt} = 0.6$, $\rho_{tc} = 0.0$, and $\rho_{tu} = 1.1$, along with the background contributions, is shown after the likelihood fit to data (“Post-Fit”) for the signal-plus-background hypothesis. The ratio of the data to the prediction (“Pred.”) is shown in the lower panel, separately for post-fit signal-plus-background (black points) and pre-fit background (dashed blue line). The size of the combined statistical and systematic uncertainty in the signal-plus-background prediction is indicated by the blue hatched band.
and a multi-output deep neural network classifier. The dominant backgrounds originate from signal regions. The best-fit signal for $m_H = 900$ GeV and couplings $\rho_{tt}=0.6$, $\rho_{tc}=0.0$, and $\rho_{tu}=1.1$, is shown for a signal strength $\mu = 0.07 \pm 0.03$.

<table>
<thead>
<tr>
<th></th>
<th>2fSS ++ CAT sstt</th>
<th>2fSS ++ CAT ttq</th>
<th>2fSS ++ CAT ttt</th>
</tr>
</thead>
<tbody>
<tr>
<td>Signal</td>
<td>35 ±13</td>
<td>19 ±7</td>
<td>5.8 ±2.2</td>
</tr>
<tr>
<td>$ttt$</td>
<td>0.19 ±0.05</td>
<td>0.45 ±0.09</td>
<td>2.3 ±0.5</td>
</tr>
<tr>
<td>$tW$</td>
<td>220 ±20</td>
<td>135 ±10</td>
<td>214 ±16</td>
</tr>
<tr>
<td>$tH$</td>
<td>17.2 ±2.6</td>
<td>15.4 ±2.6</td>
<td>37 ±5</td>
</tr>
<tr>
<td>$tZ/\gamma^*$</td>
<td>43 ±5</td>
<td>23.6 ±2.8</td>
<td>53 ±5</td>
</tr>
<tr>
<td>$t\gamma^*$ (LM)</td>
<td>5.6 ±3.1</td>
<td>4.5 ±2.7</td>
<td>5.5 ±3.0</td>
</tr>
<tr>
<td>$VV$</td>
<td>22 ±7</td>
<td>11 ±4</td>
<td>5.2 ±2.0</td>
</tr>
<tr>
<td>$t\ell$</td>
<td>23.7 ±1.4</td>
<td>9.3 ±0.6</td>
<td>5.3 ±0.4</td>
</tr>
<tr>
<td>Non-prompt $\ell$</td>
<td>42 ±15</td>
<td>20 ±7</td>
<td>19 ±6</td>
</tr>
<tr>
<td>Mat Conv</td>
<td>19 ±5</td>
<td>6.1 ±1.6</td>
<td>7.3 ±2.2</td>
</tr>
<tr>
<td>QM ISD</td>
<td>7.4 ±2.7</td>
<td>1.7 ±0.6</td>
<td>1.4 ±0.5</td>
</tr>
<tr>
<td>Other</td>
<td>5.9 ±2.0</td>
<td>4.1 ±1.4</td>
<td>8.1 ±2.2</td>
</tr>
<tr>
<td>Total</td>
<td>441 ±16</td>
<td>250 ±8</td>
<td>365 ±12</td>
</tr>
<tr>
<td>Data</td>
<td>434</td>
<td>261</td>
<td>342</td>
</tr>
</tbody>
</table>

Table 7. Post-fit yields of the 2fSS positively charged signal regions. The best-fit signal for $m_H = 900$ GeV and couplings $\rho_{tt}=0.6$, $\rho_{tc}=0.0$, and $\rho_{tu}=1.1$, is shown for a signal strength $\mu = 0.07 \pm 0.03$.

<table>
<thead>
<tr>
<th></th>
<th>2fSS -- -- CAT sstt</th>
<th>2fSS -- -- CAT ttq</th>
<th>2fSS -- -- CAT ttt</th>
</tr>
</thead>
<tbody>
<tr>
<td>Signal</td>
<td>3.2 ±2.1</td>
<td>1.4 ±0.7</td>
<td>0.48 ±0.26</td>
</tr>
<tr>
<td>$ttt$</td>
<td>0.20 ±0.05</td>
<td>0.49 ±0.11</td>
<td>2.4 ±0.6</td>
</tr>
<tr>
<td>$tW$</td>
<td>110 ±9</td>
<td>70 ±5</td>
<td>124 ±9</td>
</tr>
<tr>
<td>$tH$</td>
<td>17.1 ±2.5</td>
<td>15.3 ±2.2</td>
<td>37 ±6</td>
</tr>
<tr>
<td>$tZ/\gamma^*$</td>
<td>42 ±5</td>
<td>23.4 ±2.6</td>
<td>53 ±5</td>
</tr>
<tr>
<td>$t\gamma^*$ (LM)</td>
<td>10 ±5</td>
<td>3.7 ±2.1</td>
<td>7 ±4</td>
</tr>
<tr>
<td>$VV$</td>
<td>21 ±6</td>
<td>7.5 ±2.6</td>
<td>3.7 ±1.5</td>
</tr>
<tr>
<td>$t\ell$</td>
<td>13.0 ±0.8</td>
<td>5.53 ±0.33</td>
<td>3.25 ±0.28</td>
</tr>
<tr>
<td>Non-prompt $\ell$</td>
<td>54 ±16</td>
<td>22 ±8</td>
<td>25 ±9</td>
</tr>
<tr>
<td>Mat Conv</td>
<td>16 ±5</td>
<td>4.0 ±1.1</td>
<td>6.9 ±1.9</td>
</tr>
<tr>
<td>QM ISD</td>
<td>7.4 ±2.7</td>
<td>1.7 ±0.6</td>
<td>1.4 ±0.5</td>
</tr>
<tr>
<td>Other</td>
<td>6.5 ±2.1</td>
<td>3.9 ±1.2</td>
<td>7.8 ±2.3</td>
</tr>
<tr>
<td>Total</td>
<td>300 ±10</td>
<td>159 ±5</td>
<td>271 ±8</td>
</tr>
<tr>
<td>Data</td>
<td>296</td>
<td>158</td>
<td>282</td>
</tr>
</tbody>
</table>

Table 8. Post-fit yields of the 2$\ell$ negatively charged signal regions. The best-fit signal for $m_H = 900$ GeV and couplings $\rho_{tt}=0.6$, $\rho_{tc}=0.0$, and $\rho_{tu}=1.1$, is shown for a signal strength $\mu = 0.07 \pm 0.03$.

9 Conclusion

A search for a general two Higgs doublet model is presented, where the heavy Higgs bosons feature flavour changing couplings. Such couplings allow for same-sign top and three-top production among others, with a sizeable charge asymmetry. The targeted final state is characterised by multiple leptons and multiple $b$-jets. To improve the sensitivity of the search, events are categorised according to the lepton multiplicity, total lepton charge, and a multi-output deep neural network classifier. The dominant backgrounds originate from $t\bar{t}W$, $t\bar{t}Z$, and $tt\bar{t}$, and are estimated from Monte-Carlo simulation and normalised...
Table 9. Post-fit yields of the $3\ell$ and $4\ell$ signal regions. The best-fit signal for $m_H = 900$ GeV and couplings $\rho_{tt}=0.6$, $\rho_{tc}=0.0$, and $\rho_{tu}=1.1$, is shown for a signal strength $\mu = 0.07 \pm 0.03$.

Figure 13. Comparison between data and prediction for the DNN$^{SB}$ distribution used in different signal region categories of the $3\ell$ and $4\ell$ channels: (a) $3\ell --$ CAT ttt, (b) $3\ell ++$ CAT ttt, (c) $3\ell --$ CAT ttq, (d) $3\ell ++$ CAT ttq, (e) $3\ell --$ CAT tttt, (f) $3\ell ++$ CAT tttt, and (g) $4\ell$. The expected signal for $m_H = 900$ GeV and couplings $\rho_{tt}=0.6$, $\rho_{tc}=0.0$, and $\rho_{tu}=1.1$, along with the background contributions, is shown after the likelihood fit to data (“Post-Fit”) for the signal-plus-background hypothesis. The ratio of the data to the prediction (“Pred.”) is shown in the lower panel, separately for post-fit signal-plus-background (black points) and pre-fit background (dashed blue line). The size of the combined statistical and systematic uncertainty in the signal-plus-background prediction is indicated by the blue hatched band.
Figure 14. Observed and expected exclusion limits at 95% confidence level on the heavy Higgs boson mass for the g2HDM signal model for different couplings choices: (a) $\rho_T = 0.4$, $\rho_C = 0.2$, $\rho_T = 0.2$, and $\rho_T = 0$, $\rho_C = 0$, and (d) $\rho_T = 0.6$, $\rho_C = 0$, $\rho_T = 1.1$, the latter corresponding to the couplings yielding the most significant excess. The yellow and green contours of the band around the expected limit are the $\pm 1\sigma$ and $\pm 2\sigma$ variations including all uncertainties, respectively. The theoretical prediction for the signal production cross section is also shown as a red line. The production cross section is the sum of the five production modes considered in the search.
Figure 15. Observed (solid line) and expected (dashed line) exclusion limits on the scalar mass as a function of the coupling for different assumptions: (a) $\rho_{tc} = \rho_{tu}$, (b) $\rho_{tu} = 0$, (c) $\rho_{tc} = 0$, and (d) $\rho_{tt} = 0$. 
Figure 16. Expected and observed exclusion limits on the cross section × branching ratio of electroweakino production in RPV SUSY models as a function of the sparticle masses: (a) higgsino model, and (b) wino model. The yellow and green contours of the band around the expected limit are the ±1σ and ±2σ variations including all uncertainties, respectively. The theoretical prediction for the signal production cross section is also shown as a red line.

Figure 17. Expected and observed exclusion limits on the smuon plus bino RPV SUSY model. The yellow and green contours of the band around the expected limit are the ±1σ and ±2σ variations including all uncertainties, respectively. The diagonal grey lines indicates the allowed kinematic limit for the decays. The neutralino is assumed to decay to tbl or bbν with equal probability.
to data. The analysis is performed with proton-proton collision data at $\sqrt{s} = 13$ TeV collected from 2015 to 2018 with the ATLAS detector at the LHC, corresponding to an integrated luminosity of 139 fb$^{-1}$. This search is the first collider result on general two Higgs doublet model with flavour violation. It also represents the first search to target explicitly beyond-the-standard-model production of three top quarks.

The largest deviation observed with respect to the Standard Model expectation corresponds to a local significance of 2.8 standard deviations for a signal with $m_H = 900$ GeV and couplings $\rho_{tt} = 0.6$, $\rho_{tc} = 0.0$, and $\rho_{tu} = 1.1$. Exclusion limits are set on the mass and couplings of the heavy Higgs bosons, where an additional scalar boson with couplings $\rho_{tt} = 0.4$, $\rho_{tc} = 0.2$, and $\rho_{tu} = 0.2$ is excluded at 95% confidence level observed (expected) for masses $m_H$ between 200–620 (200–840) GeV. Additional mass limits are set for different coupling choices. For a fixed mass of $m_H = 400$ GeV, exclusion limits are set on the allowed coupling strengths as low as $\rho_{tt} = 0.3$, $\rho_{tc} = \rho_{tu} = 0.18$. Different assumptions are tested to set 2-dimensional exclusion limits on the three couplings. Additional models based on R-parity-violating supersymmetry with the lepton-number-violating coupling $\lambda_{33}^i$ (with $i \in 2, 3$), are used to further interpret the results of the search. Scenarios with direct electroweak production of higgsinos (winos) are excluded for masses between 200–585 (200–670) GeV. Smuons with masses between 225 and 460 GeV are excluded in a model with direct smuon production and decay to a bino-like neutralino, which in turn decays via the $\lambda_{33}^i$ coupling. These are the first collider limits for these models.

Acknowledgments

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

We acknowledge the support of ANPCyT, Argentina; YerPhI, Armenia; ARC, Australia; BMWFW and FWF, Austria; ANAS, Azerbaijan; CNPq and FAPESP, Brazil; NSERC, NRC and CFI, Canada; CERN; ANID, Chile; CAS, MOST and NSFC, China; Minciencias, Colombia; MEYS CR, Czech Republic; DNRF and DNSRC, Denmark; IN2P3-CNRS and CEA-DRF/IRFU, France; SRNSFG, Georgia; BMBF, HGF and MPG, Germany; GSRI, Greece; RGC and Hong Kong SAR, China; ISF and Benoziyo Center, Israel; INFN, Italy; MEXT and JSPS, Japan; CNRST, Morocco; NWO, Netherlands; RCN, Norway; MEiN, Poland; FCT, Portugal; MNE/IFA, Romania; MESTD, Serbia; MSSR, Slovakia; ARRS and MIZŠ, Slovenia; DSI/NRF, South Africa; MICINN, Spain; SRC and Wallenberg Foundation, Sweden; SERI, SNSF and Cantons of Bern and Geneva, Switzerland; MOST, Taiwan; TENMAK, Türkiye; STFC, United Kingdom; DOE and NSF, United States of America. In addition, individual groups and members have received support from BCKDF, CANARIE, Compute Canada and CRC, Canada; PRIMUS 21/SCI/017 and UNCE SCI/013, Czech Republic; COST, ERC, ERDF, Horizon 2020 and Marie Skłodowska-Curie Actions, European Union; Investissements d’Avenir Labex, Investissements d’Avenir Idex and ANR, France; DFG and AvH Foundation, Germany; Herakleitos, Thales and Aristeia programmes co-financed by EU-ESF and the Greek NSRF, Greece; BSF-NSF and MINERVA, Israel; Norwegian Financial Mechanism 2014-2021, Nor-
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. [151].

Open Access. This article is distributed under the terms of the Creative Commons Attribution License (CC-BY 4.0), which permits any use, distribution and reproduction in any medium, provided the original author(s) and source are credited.

References


[20] CMS collaboration, Search for new neutral Higgs bosons through the $H \rightarrow Z A \rightarrow \ell^+\ell^- b\bar{b}$ process in pp collisions at $\sqrt{s} = 13$ TeV, JHEP 03 (2020) 055 [arXiv:1911.03781] [inSPIRE].


[22] CMS collaboration, Search for a charged Higgs boson decaying into top and bottom quarks in events with electrons or muons in proton-proton collisions at $\sqrt{s} = 13$ TeV, JHEP 01 (2020) 096 [arXiv:1908.09206] [inSPIRE].

[23] CMS collaboration, Search for heavy Higgs bosons decaying to a top quark pair in proton-proton collisions at $\sqrt{s} = 13$ TeV, JHEP 04 (2020) 171 [Erratum ibid. 03 (2022) 187] [arXiv:1908.01115] [inSPIRE].

[24] CMS collaboration, Search for charged Higgs bosons in the $H^\pm \rightarrow \tau^\pm \nu$, decay channel in proton-proton-collisions at $\sqrt{s} = 13$ TeV, JHEP 07 (2019) 142 [arXiv:1903.04560] [inSPIRE].


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


[148] ATLAS collaboration, Search for $t\bar{t}H/A \to t\bar{t}A$ production in the multilepton final state in proton–proton collisions at $\sqrt{s} = 13$ TeV with the ATLAS detector, *JHEP* **07** (2023) 203 [arXiv:2211.01136] [inSPIRE].


ac Also at Institute of Physics, Azerbaijan Academy of Sciences, Baku; Azerbaijan
ad Also at Institute of Theoretical Physics, Ilia State University, Tbilisi; Georgia
ae Also at L2IT, Université de Toulouse, CNRS/IN2P3, UPS, Toulouse; France
af Also at Lawrence Livermore National Laboratory, Livermore; United States of America
ag Also at National Institute of Physics, University of the Philippines Diliman (Philippines); Philippines
ah Also at Technical University of Munich, Munich; Germany
ai Also at The Collaborative Innovation Center of Quantum Matter (CICQM), Beijing; China
aj Also at TRIUMF, Vancouver BC; Canada
ak Also at Università di Napoli Parthenope, Napoli; Italy
al Also at University of Colorado Boulder, Department of Physics, Colorado; United States of America
am Also at Washington College, Chestertown, MD; United States of America
an Also at Yeditepe University, Physics Department, Istanbul; Türkiye

* Deceased