ATLAS CONF Note

CONF-HDBS-2022-35

June 2, 2022

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

The ATLAS Collaboration

A search for non-resonant Higgs boson pair production in the $b\bar{b}b\bar{b}$ final state is presented. The analysis uses 126 fb^{-1} of pp collision data at $\sqrt{s} = 13$ TeV collected with the ATLAS detector at the Large Hadron Collider, and targets both the gluon-gluon fusion and vector-boson fusion production modes. No evidence for signal is found and the observed (expected) upper limit on the cross section for non-resonant Higgs boson pair production is determined to be 5.4 (8.1) times the Standard Model predicted cross-section at 95% confidence level. Constraints are placed upon modifiers to the HHH and $HHVV$ couplings, κ_λ and κ_{2V}. The observed (expected) constraints on κ_λ are determined to be [-3.9, 11.1] ([-4.6, 10.8]) at 95% confidence level, while the corresponding constraints for κ_{2V} are [-0.03, 2.11] ([-0.05, 2.12]).

ATLAS-CONF-2022-035
03 June 2022

1 Introduction

The discovery of the 125 GeV Higgs boson (H) [1–4] at the Large Hadron Collider (LHC) has prompted a broad research programme to investigate its properties and compare the results with the Standard Model (SM) predictions. Of particular interest is the search for non-resonant Higgs boson pair, or di-Higgs (HH), production. This process has a direct dependence on the Higgs self-coupling, which is a key ingredient of the electroweak symmetry breaking mechanism and a sensitive probe of both the SM and new physics [5–11].

The dominant SM HH production process is gluon–gluon fusion (ggF). Its cross-section, for a Higgs boson mass m_H =125 GeV, calculated at next-to-next-to-leading order (NNLO) including finite top-quark mass effects [12], is 31.05 fb at $\sqrt{s}$ =13 TeV. The two leading order Feynman diagrams contributing to this process are shown in Figure 1, where Figure 1(a) is commonly referred to as the *box* diagram and Figure 1(b) as the *triangle* diagram. The triangle diagram is dependent on the trilinear Higgs self-coupling, λ (the interaction between three Higgs bosons), which can be expressed in terms of its modifier, κ_λ.[1] In the SM, these two diagrams interfere destructively. As a result, the HH production cross-section and kinematic properties depend critically on the value of κ_λ.

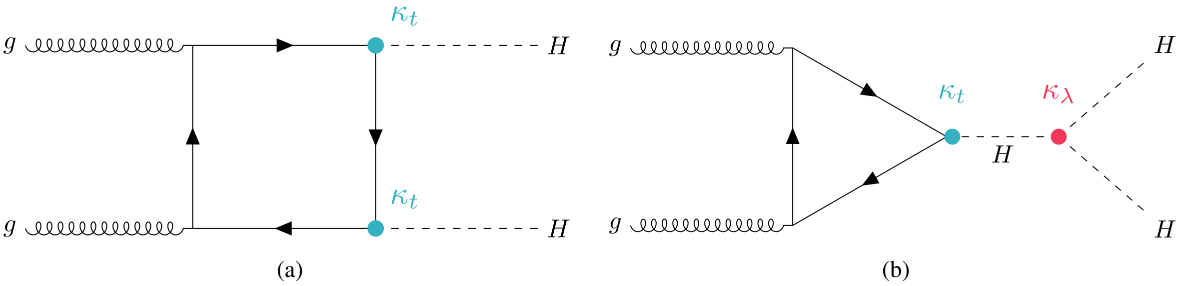


Figure 1: The two tree-level gluon-gluon fusion di-Higgs production Feynman diagrams: (a) the box diagram; (b) the triangle diagram.

The second-leading HH production process in the SM is vector-boson fusion (VBF), with a cross-section of 1.73 fb at next-to-next-to-next-to-leading order (N^{3}LO) [13], for m_H =125 GeV. The three tree-level Feynman diagrams describing this process are shown in Figure 2. Figures 2(a) and 2(b) show the dependence on κ_λ and the $HHVV$ interaction vertex modifier, κ_{2V}, respectively. In the SM, perturbative unitarity leads to exact cancellation of divergences between the diagrams in Figures 2(b) and 2(c). As κ_V (the HVV coupling modifier) and κ_{2V} depart from their SM value of one, the cross-section grows linearly with the effective center-of-mass energy of the incoming vector-bosons, resulting in harder kinematics for the produced Higgs bosons [14], up to the scale of some new physics that would unitarize the total amplitude.

The analysis described in this paper targets the $b\bar{b}b\bar{b}$ final state, in both the ggF and VBF production modes, using the full ATLAS Run 2 dataset. Assuming the SM branching ratio of 58.2% for $H \rightarrow b\bar{b}$ [15, 16], about one third of di-Higgs events decay to $b\bar{b}b\bar{b}$, making it the most abundant di-Higgs final state. However, as this is a fully hadronic final state, the analysis faces the challenge of large backgrounds from multijet production processes.

---

[1] Coupling modifiers, κ, are defined as the ratio of the modified coupling to its SM value, $\kappa = c/c^{\mathrm{SM}}$.

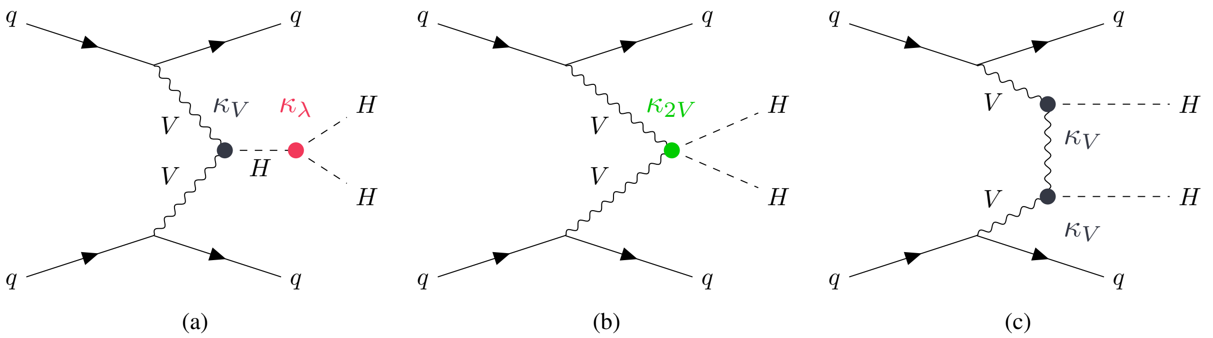


Figure 2: The three tree-level vector-boson fusion di-Higgs production Feynman diagrams.

The ATLAS collaboration has previously published search results for non-resonant $HH \rightarrow b\bar{b}b\bar{b}$ production, using 27 fb^{-1} of early Run 2 data in the search for ggF HH production [17] and the full Run 2 dataset in the search for VBF HH production [18]. The present analysis, in addition to using the full Run 2 dataset for both production modes and combining the results, is optimized specifically for the non-resonant HH event topology, and it also benefits from improvements in jet reconstruction and the identification of jets arising from the decays of b-quarks in ATLAS ("b-tagging"). In addition, the analysis employs a fully data-driven technique for the background estimation, which uses an artificial neural network to perform a kinematic reweighting of data from an alternative phase space to model the background in the region of interest. The CMS collaboration has also published results on the search for non-resonant $HH \rightarrow b\bar{b}b\bar{b}$ with the full Run 2 data [19, 20], and both ATLAS and CMS have published results on the search for resonant and non-resonant HH production in a multitude of other final states [21–30].

This document is structured as follows: the ATLAS detector and the data and simulated events used in the analysis are described in Sections 2 and 3, respectively. Section 4 presents the reconstruction and identification of physics objects in this analysis and Section 5 details the event selection and categorization. The background modeling method is described in Section 6, the systematic uncertainties are detailed in Section 7 and, finally, the results are reported in Section 8 and a summary is given in Section 9.

2 ATLAS detector

The ATLAS detector [31] 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 (ID) system is immersed in a 2 T axial magnetic field and provides charged-particle tracking in the range $|\eta| < 2.5$. The high-granularity silicon pixel detector covers the vertex region and typically provides four space-point measurements per track, the first hit normally being in the insertable

---

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

B-layer installed before Run 2 [32, 33]. Following the pixel detector is the silicon microstrip tracker, which usually provides eight measurements per track. These silicon detectors are complemented by the transition radiation tracker, which enables radially extended track reconstruction up to $|\eta| = 2.0$.

The calorimeter system covers the pseudorapidity range $|\eta| < 4.9$. Within $|\eta| < 3.2$, 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 optimized 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 T·m across most of the detector. A set of precision chambers covers the region $|\eta| < 2.7$ with three layers of monitored drift tubes, 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.

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 [34]. The first-level trigger accepts events from the 40 MHz bunch crossings at a rate below 100 kHz, which the high-level trigger further reduces in order to record events to disk at about 1 kHz.

An extensive software suite [35] 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 samples

3.1 Data sample

This analysis is performed in LHC proton–proton (pp) collision data at $\sqrt{s} = 13$ TeV collected in 2016–2018. Only data collected during stable beam conditions are used, with all relevant detector systems functional [36], corresponding to an integrated luminosirty of 126 fb^{-1}.

The analysis uses events that satisfy two types of trigger signature requirements with different criteria on the number of jets and their b-tagging status [37]. The jets used are reconstructed with the anti-k_t algorithm [38, 39], with a radius parameter of $R = 0.4$. b-tagging is performed at the trigger level with the MV2c20 algorithm in 2016 and the MV2c10 algorithm in 2017-2018 [37], with a range of b-jet identification efficiency operating points from 40% to 70%. The first of the two trigger signatures used for selecting $b\bar{b}b\bar{b}$ events requires two b-jets plus one additional jet ("2b1j"), while the second requires two b-jets plus two additional jets ("2b2j"). The minimum transverse energy (E_{T}) requirement on the jets is 35 GeV for all jets used in the 2b2j trigger and the b-jets in the 2b1j trigger, and 100 to 150 GeV (depending on the year of data taking) for the additional jet used in the 2b1j triggers.

During 2016 data taking, a fraction of the data (8.3 fb^{-1}) was affected by an inefficiency in the online vertex reconstruction, which reduced the efficiency of b-tagging algorithms in the trigger; those events were not retained for further analysis, resulting in an integrated luminosity of 24.6 fb^{-1} for the 2016 dataset. The integrated luminosities of the 2017 and 2018 datasets are 43.7 fb^{-1} and 57.7 fb^{-1}, respectively.

3.2 Simulated samples

Monte Carlo (MC) simulation is used for the modeling of signal events, as well as to produce event samples of background processes for cross-checks and validation studies. The Higgs boson mass is set to 125 GeV in the simulation. All samples are processed by the ATLAS simulation framework [40] and the detector response is simulated with GEANT4 [41].

The ggF signal process was simulated using the POWHEG BOX v2 generator [42–44] at next-to-leading order (NLO), including finite top-quark mass effects, using the PDF4LHC15 [45] parton distribution function (PDF) set. Parton showers and hadronization are simulated with PYTHIA 8.244 [46] with the A14 set of tuned parameters [47] and the NNPDF2.3LO PDF set [48]. The SM ggF HH cross-section was taken as $\sigma_{\text{ggF}} = 31.05$ fb, calculated at NNLO including finite top-quark mass effects [12]. Signal samples for the ggF process were generated explicitly for modifier values of $\kappa_\lambda = 1$ and 10. A reweighting method is used to determine the ggF signal yield at each κ_λ value. Scale factors are derived as a function of κ_λ in bins of the truth invariant mass of the HH system by performing a linear combination of generator-level samples at three different κ_λ values [49]. The $\kappa_\lambda = 10$ ggF signal sample is used to validate the derived scale factors; the agreement between this generated sample and the signal sample obtained from the reweighting method is found to be within the statistical precision of the simulated sample. Additional truth-level ggF HH signal samples without parton showering were generated with POWHEG BOX v2 for the $\kappa_\lambda = 0$ and 20 coupling modifier configurations to provide a basis for the κ_λ reweighting, along with the SM ggF sample. For the reweighted ggF signal, the NNLO cross-section as a function of κ_λ is taken from Ref. [12]. Alternative ggF samples were generated using the POWHEG BOX v2 generator at NLO using the PDF4LHC15 PDF set, interfaced to HERWIG 7.1.6 [50] for parton showering and hadronization and using the H7.1-Default set of tuned parameters [51] and MMHT2014LO PDF set [52], in order to assess parton showering uncertainties.

The VBF signal process was simulated using MADGRAPH 2.7.3 [53] at leading order (LO) using the NNPDF3.0NLO PDF set, interfaced with PYTHIA 8.244 for parton showering and hadronization using the A14 set of tuned parameters and NNPDF2.3LO PDF set. Signal samples for the VBF process were generated explicitly for coupling modifier values of $(\kappa_\lambda, \kappa_{2V}, \kappa_V) = (1, 1, 1)$, $(1, 1.5, 1)$, $(2, 1, 1)$, $(10, 1, 1)$, $(1, 1, 0.5)$, $(-5, 1, 0.5)$, $(0, 1, 1)$, $(1, 0, 1)$, and $(1, 3, 1)$. Interference between various diagrams contributing to the non-resonant signal is considered in the simulation. A linear combination of the first six of the listed samples is used to derive distributions for a finer granularity of κ_{2V} values, following a technique used previously to generate κ_λ distributions [54]. The specific basis of six samples utilized is chosen to avoid large statistical uncertainties in the reweighted signal samples resulting from sparsely populated areas of kinematic phase space. The generated VBF signal samples not included in the linear combination basis – $(\kappa_\lambda, \kappa_{2V}, \kappa_V) = (0, 1, 1)$, $(1, 0, 1)$, and $(1, 3, 1)$ – were used to validate the performance of the combination method. The agreement between these generated samples and the corresponding signal samples obtained from the combination method was found to be within the statistical precision of the simulated samples. The cross-section of the VBF HH process, evaluated at N^3LO in QCD, is 1.73 fb in the SM [13, 55–57]. For the reweighted VBF signal points, the N^3LO to LO cross-section ratio at the SM value is calculated, and this factor is applied to the cross-sections at each κ_λ, κ_{2V}, and κ_V point. Alternative LO samples were generated using MADGRAPH 2.7.3 with the NNPDF3.0NLO PDF set, interfaced to HERWIG 7.0.4 with the

H7.1-Default set of tuned parameters and MMHT2014LO PDF set for parton showering and hadronization, in order to assess parton showering uncertainties.

Top quark pair production ($t\bar{t}$) and multijet background processes were simulated in order to validate the background modeling procedure. The $t\bar{t}$ sample was simulated at NLO in α_s using POWHEG BOX v2 [58]. Parton showering, hadronization, and the underlying event were modeled using PYTHIA 8.230. The matrix element calculation uses NNPDF3.0NLO [59] as the PDF set, while the parton shower and underlying-event modeling uses NNPDF2.3LO [48] and the A14 set of tuned parameters. The damping parameter h_{damp}, which effectively regulates radiation at high p_{T}, was set to 1.5 times the top quark mass. The $t\bar{t}$ simulation is normalized using the value of the inclusive cross-section calculated with TOP++ 2.0 [60, 61]. This accounts for NNLO corrections in α_s, including next-to-next-to-leading logarithmic (NNLL) resummation of soft gluon terms. The multijet background samples were modeled using PYTHIA 8.235. This simulates pure QCD 2-to-2 interactions at LO in α_s. Events were showered using the parton shower native to PYTHIA, which includes radiation and splitting that can result in additional jets. The A14 set of tuned underlying-event parameters and the NNPDF2.3LO PDF set were used.

Other background processes, such as SM Higgs boson, HH (in other final states) and electroweak diboson production, have been estimated to give negligible contributions to the selected event yields and are therefore not included.

The effect of multiple interactions in the same and neighboring bunch crossings (pileup) was modeled by overlaying the simulated hard-scattering event with inelastic pp events generated with PYTHIA 8.186 using the NNPDF2.3LO PDF set and the A3 set of tuned parameters [62]. Additionally, for all HH signal samples, heavy-flavour decays were modeled using EVTGEN 1.7.0 [63].

4 Object reconstruction

Primary vertices from pp interactions are reconstructed using at least two charged-particle tracks with transverse momentum (p_{T}) above 500 MeV measured with the ID [64]. Events are required to contain a primary vertex with at least two associated tracks. The vertex with the largest sum of squared track momenta ($\sum p_{\mathrm{T}}^2$) is chosen as the hard-scatter primary vertex.

Hadronic jets are reconstructed using the anti-k_t algorithm with radius parameter $R = 0.4$. The jet clustering uses particle-flow objects as inputs [65]. Particle-flow objects are charged-particle tracks matched to the hard-scatter vertex and calorimeter energy clusters following an energy subtraction algorithm that removes the calorimeter deposits associated with good-quality tracks from any vertex. The tracking information helps to improve the energy resolution of the calorimeter clusters and reduce the impact from pileup. The momenta of reconstructed jets are calibrated in a multi-step procedure [66]. Jets with $p_{\mathrm{T}} < 60$ GeV and $|\eta| < 2.4$ must also satisfy a requirement based on the output of the multivariate “jet vertex tagger” (JVT) algorithm [67], which is used to identify and reject jets in which much of the energy originates from pileup interactions. Correction factors are applied to the simulated events to compensate for differences between the JVT efficiencies in data and simulation. In the $HH \to b\bar{b}b\bar{b}$ analysis, jets are discarded if they fail the “Tight” JVT working point, corresponding to an average efficiency of 96% for jets from the hard-scatter vertex.

Jets with radius parameter $R = 0.4$ are also reconstructed from topological clusters of energy deposits in the calorimeter [68] and calibrated in the same way as the jets reconstructed from particle-flow objects. These jets are used exclusively for the purpose of applying quality criteria to identify events which are consistent with noise in the calorimeter or non-collision background [69]. Events containing at least one such jet, which has $p_{\mathrm{T}} > 20\,\mathrm{GeV}$, passes the JVT, and fails to meet these quality criteria, are rejected.

Jets with $|\eta| < 2.5$ originating from b-quarks are b-tagged by applying the DL1r algorithm [70], which is based on a multivariate classification technique combining information from the impact parameters of ID tracks, the presence of displaced secondary vertices, and the reconstructed flight paths of b- and c-hadrons inside the jet [71]. The DL1r working point used in the $HH \to b\bar{b}b\bar{b}$ analysis is the one that gives 77% efficiency for jets associated with true b-hadrons in simulated $t\bar{t}$ events. At this working point, the light-jet (charm-jet) rejection measured in $t\bar{t}$ simulation is about a factor of 130 (4.9). The calibration of the DL1r algorithm is performed separately for each jet type [72, 73] and correction factors are derived and applied to the simulated samples to compensate for differences between the b-tagging efficiencies in data and simulation.

Muons are reconstructed by matching ID tracks with either MS tracks or aligned individual hits in the MS and performing a combined track fit. They are required to have $p_{\mathrm{T}} > 4\,\mathrm{GeV}$ and $|\eta| < 2.5$, and to satisfy "Medium" identification criteria based on track-quality variables [74]. Muons are used only to apply energy corrections to jets.

A momentum correction is applied to b-tagged jets to account for energy lost to soft out-of-cone radiation and to muons and neutrinos in semileptonic b-hadron decays. This correction follows the procedure used in Ref. [75] and consists of two steps. First, if any muon is within $\Delta R = 0.4$ of a b-tagged jet, the four-momentum of the muon is added to that of the jet. Any energy deposited in the calorimeter by the muon is then subtracted from the jet to avoid double counting; this is computed according to the description in Ref. [76]. In the second step, a global scale factor is applied to each b-tagged jet based on its p_{T} and whether or not it has a muon within $\Delta R = 0.4$ of the jet axis. These scale factors are derived from simulation.

5 Analysis selection and categorization

The analysis utilizes a set of criteria to select $HH \to b\bar{b}b\bar{b}$ candidate events, including dedicated cuts to separate events into orthogonal ggF and VBF signal regions. "Forward" and "central" jets are used with the following selection criteria:

- central jets: $|\eta| < 2.5$ and $p_{\mathrm{T}} > 40\,\mathrm{GeV}$;
- forward jets: $2.5 < |\eta| < 4.5$ and $p_{\mathrm{T}} > 30\,\mathrm{GeV}$.

An initial "preselection" is applied to all events in order to conserve computing resources, which requires at least 4 jets with $p_{\mathrm{T}} > 40\,\mathrm{GeV}$, at least 2 of which are b-tagged. As described in Section 3, the events considered in this analysis are selected online through the $2b2$j and $2b1$j trigger signatures. In order to simplify the modeling of trigger efficiencies, a further selection is applied using offline kinematic quantities. Events are selected if they have a leading jet (i.e. jet with the highest p_{T}) with $p_{\mathrm{T}} > 170\,\mathrm{GeV}$, the third leading jet with $p_{\mathrm{T}} > 70\,\mathrm{GeV}$, and pass the "$2b1$j" trigger; or if they fail either of the two jet p_{T} requirements and pass the "$2b2$j" trigger. This selection step leads to about 10% loss of signal efficiency, but enables the reliable calculation of simulation-to-data correction factors for estimating the trigger efficiency in the remaining $HH \to b\bar{b}b\bar{b}$ signal events, depending on which of the above two *trigger classes* they belong to.

Events are required to contain at least four central jets passing the b-tagging requirement outlined in Section 4. The four highest-p_{T} b-tagged jets are chosen to reconstruct the decays of the two Higgs bosons. These four jets can be matched one-to-one (within $\Delta R < 0.3$) to the four b-quarks from the decays of the Higgs bosons in about 75% of simulated signal events reaching this selection stage. Events for which this matching fails are dominated by those in which one of the b-quarks from the Higgs boson decays leads to a jet that is out of acceptance.

From the four selected b-tagged jets, there are three possible combinatorial pairings to form the two Higgs boson candidates. For a given pairing, the four-momentum of the Higgs boson is defined as the sum of the four-momenta of the two jets used to construct it. The analysis selects the pairing configuration in which the higher-p_{T} jet pair has the smallest ΔR separation. In the simulated samples with SM coupling values, for which the analysis was mainly optimized, this method gives the correct pairing in around 90% of those signal events in which the four b-tagged jets are correctly matched to the b-quarks from the decays of the Higgs bosons. While the pairing accuracy drops for values of the coupling modifiers κ_λ and κ_{2V} that result in softer p_{T} spectra for the produced Higgs bosons, this pairing method leads to a smoothly varying distribution of the expected background in the plane of the invariant masses of the two Higgs boson candidates, which facilitates the derivation of the reweighting function used in the data-driven background estimation described in Section 6.

Events are then subjected to additional selections designed to isolate those consistent with the VBF production mode. For this, events must contain at least two additional jets, central or forward; b-tagged jets are excluded. The two jets forming the pair with the largest invariant mass (m_{jj}) are chosen as the "VBF jets". The VBF jet pair is required to satisfy $m_{\mathrm{jj}} > 1$ TeV, and the pseudorapidity separation between the two jets, $|\Delta\eta_{\mathrm{jj}}|$, must satisfy $|\Delta\eta_{\mathrm{jj}}| > 3$. Lastly, the transverse component of the momentum vector sum of the two VBF jets and the four jets forming the Higgs boson candidates is required to be less than 65 GeV. Events failing any of the above criteria are considered further in the ggF-targeted selection.

Events satisfying either the ggF or VBF selections are required to pass additional selection criteria designed to reduce the background and improve the analysis sensitivity. In order to suppress the $t\bar{t}$ background, a top-veto discriminant X_{Wt} is defined as:

$$X_{Wt} = \sqrt{\left(\frac{m_W - 80.4\,\mathrm{GeV}}{0.1\,m_W}\right)^2 + \left(\frac{m_t - 172.5\,\mathrm{GeV}}{0.1\,m_t}\right)^2}\,, \tag{1}$$

where m_W and m_t are the invariant masses of W boson and top quark candidates formed from jet combinations in each event. W candidates are formed from any pair of central jets in the event and the top quark candidates are then reconstructed by pairing the W candidates with any remaining b-tagged Higgs boson candidate jets. The X_{Wt} discriminant is designed to quantify the compatibility of an event with containing a hadronic top quark decay. Events with any jet combination producing $X_{Wt} < 1.5$ are rejected. This reduces the $t\bar{t}$ background by a factor of about 2 in simulated events, for a small loss of signal efficiency, of order 10%, and a similar reduction in the non-$t\bar{t}$, multijet background.

In order to reduce the overall background contamination, events in the ggF selection are also required to have a pseudorapidity separation of the reconstructed Higgs bosons, $|\Delta\eta_{HH}| < 1.5$. No such requirement is imposed in the VBF selection, as SM VBF HH signal events tend to have a larger $|\Delta\eta_{HH}|$.

A final analysis selection criterion to quantify the compatibility of events with the HH decay is applied in both the ggF and VBF selections. A discriminant X_{HH} is defined as:

$$X_{HH} = \sqrt{\left(\frac{m_{H1} - 124\,\text{GeV}}{0.1\, m_{H1}}\right)^2 + \left(\frac{m_{H2} - 117\,\text{GeV}}{0.1\, m_{H2}}\right)^2}\,, \tag{2}$$

where m_{H1} and m_{H2} are the masses of the leading (highest-p_{T}) and subleading reconstructed Higgs boson candidates respectively. The values of 124 and 117 GeV in the X_{HH} definition are chosen to approximately match the centers of the m_{H1} and m_{H2} distributions for correctly paired signal events. Events are required to have $X_{HH} < 1.6$ to be included in the *signal region* (SR) of the analysis.

All the selection steps of the analysis are summarized in Figure 3. The yields in the data and in the simulated signal samples for some typical coupling values are shown in Table 1. This sample of data events will be referred to as $4b$ events hereafter.

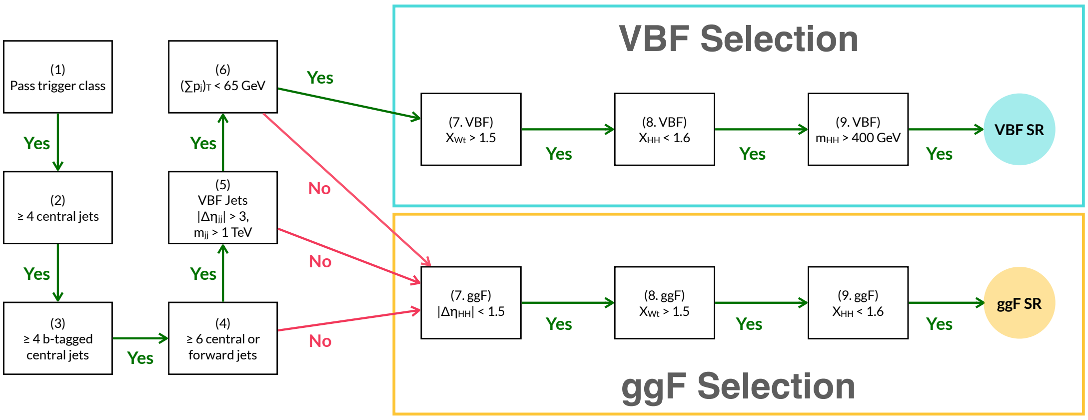


Figure 3: A flowchart summarizing the analysis selection. Events must pass selection criteria 1-3 in order to be considered for either analysis signal region. Events failing any of the selection criteria 4-6 are considered for inclusion in the ggF signal region, while those passing selection criteria 4-6 are considered for the VBF signal region. The X_{W_t} variable in the figure denotes the minimum value of the X_{W_t} variable as obtained from the different combinations of central and b-tagged jets in the event.

Both the ggF and VBF signal regions are further subdivided into a number of orthogonal categories in order to better isolate the HH signal and improve the analysis sensitivity. The $|\Delta\eta_{HH}|$ and X_{HH} quantities are used to define six orthogonal ggF categories. Three categories are defined in $|\Delta\eta_{HH}|$, with boundaries of 0, 0.5, 1, and 1.5, and two in X_{HH} with boundaries of 0, 0.95, and 1.6 . In the VBF signal region, two categories are defined using the $|\Delta\eta_{HH}|$ quantity, with the dividing boundary at $|\Delta\eta_{HH}| = 1.5$. The $|\Delta\eta_{HH}| < 1.5$ category is more sensitive to VBF signals with non-SM couplings, while the $|\Delta\eta_{HH}| > 1.5$ category is more sensitive to the SM VBF production.

The reconstructed di-Higgs boson mass m_{HH} is used as the discriminating variable for all analysis regions and categories when extracting results, as detailed in Section 8. The m_{HH} variable was found to have a strong separation power between signal and background, as well as between signal hypotheses arising from different values of coupling modifiers. The binning in m_{HH} varies between categories and is chosen in order to both maintain discrimination power and limit the expected statistical uncertainty in each bin to less than approximately 30%. This 30% threshold ensures that the assumptions used in the statistical procedure, outlined in Section 8, are satisfied. In the VBF signal region, only events with $m_{HH} > 400$ GeV are considered, as this region was found to be inadequately modeled by the background estimation technique in the $3b1$f region (as described in Section 7). For the ggF signal region, no selection on m_{HH} is applied.

Table 1: The yields of data and various example ggF and VBF HH signal models at each step of the analysis selection. The "Preselection" entry denotes an initial selection requiring at least 4 jets with $p_{\mathrm{T}} > 40\,\mathrm{GeV}$, at least 2 of which are b-tagged. Events which satisfy the "VBF selection" requirements are considered as part of the VBF signal region of the analysis, while the rest are considered for the ggF signal region. The signal yields are taken from simulation and are normalized by their theoretical cross-sections and the integrated luminosity of $126\,\mathrm{fb}^{-1}$. Corrections for differences in the b-tagging efficiency and trigger acceptance between data and simulation are applied beginning from the "Trigger class" requirement. The X_{Wt} variable in the table denotes the minimum value of the X_{Wt} variable as obtained from the different combinations of central and b-tagged jets in the event.

	Data	**ggF Signal**		**VBF Signal**	
		SM	$\kappa_\lambda = 10$	**SM**	$\kappa_{2V} = 0$
Common preselection					
Preselection	5.70×10^8	526.6	7337.7	22.3	626.1
Trigger class	2.49×10^8	381.8	5279.1	16.1	405.2
ggF selection					
Fail VBF selection	2.46×10^8	376.6	5198.0	13.9	334.4
At least 4 b-tagged central jets	1.89×10^6	86.0	1001.7	1.9	65.2
$\lvert\Delta\eta_{HH}\rvert < 1.5$	1.03×10^6	71.9	850.6	0.9	46.4
$X_{Wt} > 1.5$	7.51×10^5	60.4	569.0	0.7	43.1
$X_{HH} < 1.6$ (ggF signal region)	1.62×10^4	29.1	182.7	0.2	23.0
VBF selection					
Pass VBF selection	3.30×10^6	5.2	81.1	2.2	70.7
At least 4 b-tagged central jets	2.71×10^4	1.1	15.3	0.7	27.6
$X_{Wt} > 1.5$	2.18×10^4	1.0	11.2	0.7	26.5
$X_{HH} < 1.6$	5.02×10^2	0.5	3.1	0.3	17.3
$m_{HH} > 400\,\mathrm{GeV}$ (VBF signal region)	3.57×10^2	0.4	1.8	0.3	16.4

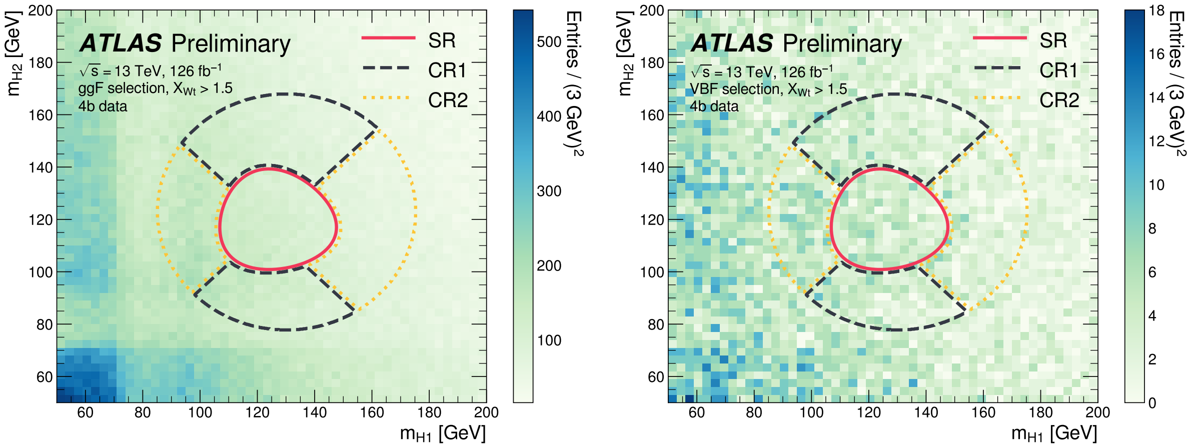
6 Background modeling

After the selection described above, about 90% of the background events come from multijet processes (excluding top quark production), with the approximately 10% remainder almost entirely composed of $t\bar{t}$ events. This background composition was determined by applying the full event selection to simulated samples of the various processes and comparing to the total background estimate in the SR; it is purely meant to be indicative and is not used for deriving any results. The only simulated datasets used in the statistical analysis are those for the signal HH processes. The background is modeled using the fully data-driven technique described below.

The background estimation makes use of an alternative set of events, which pass the same b-jet triggers and satisfy all the same selection criteria as the $4b$ events, with one difference: they are required to contain exactly two b-tagged jets. This sample, referred to hereafter as "$2b$", has about two orders of magnitude more events than the $4b$ sample, hence the presence in it of any $HH \rightarrow b\bar{b}b\bar{b}$ signal is negligible, making it suitable for the background estimation. The jets selected to form the two Higgs boson candidates in the $2b$ events are the two b-tagged jets and the two untagged jets with the highest p_{T} (excluding the VBF jets in the VBF categories).

The kinematic properties of the $2b$ and $4b$ events are not expected to be identical, partly due to different processes contributing to the two samples, but also due to differences in the trigger acceptance and because the performance of b-tagging varies as a function of jet p_{T} and η. Therefore, a reweighting function is required, which, when applied to the $2b$ events, maps their kinematic distributions onto the corresponding $4b$ distributions. This function is derived using the $2b$ and $4b$ events in a *control region* (CR) surrounding the SR in the reconstructed (m_{H1}, m_{H2}) plane and then applied to the $2b$ events in the SR to produce the background estimate. The "inner edge" of the CR is defined by $X_{HH} = 1.6$ and the "outer edge" by:

$$\sqrt{\left(m_{H1} - 1.05 \cdot 124\,\mathrm{GeV}\right)^2 + \left(m_{H2} - 1.05 \cdot 117\,\mathrm{GeV}\right)^2} = 45\,\mathrm{GeV}\,. \tag{3}$$



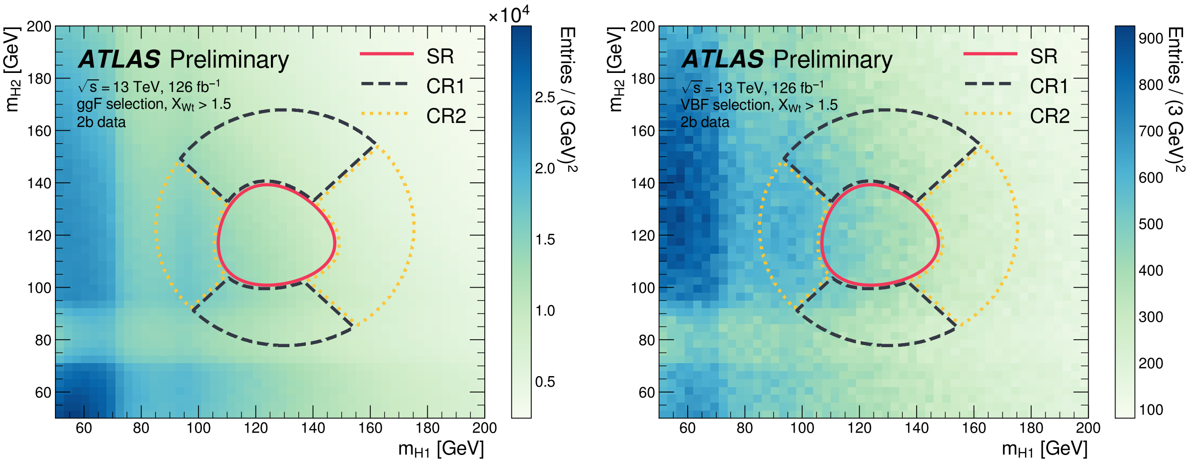
(a)

(b)

Figure 4: The mass planes of the reconstructed Higgs boson candidates for the (a) ggF and (b) VBF signal regions of the analysis, shown for the $4b$ data events. The X_{Wt} variable in the figure denotes the minimum value of the X_{Wt} variable as obtained from the different combinations of central and b-tagged jets in the event. In (a), the analysis selection up to step 8 (as outlined in Figure 3) of the ggF selection has been applied, while in (b), the the analysis selection up to step 7 of the VBF selection has been applied.

The shift by a factor of 1.05 of the center of the above circle, compared to X_{HH}, is found to be the optimal trade-off between having balanced statistics all around the SR and avoiding the low m_{H1}/m_{H2} regions, where the differences between $2b$ and $4b$ kinematic distributions tend to amplify. The CR is further split into four roughly-equal directional quadrants, defined by 45° and 135° lines passing through the SR center. The four quadrants are labeled based on compass directions: the upper quadrant $\mathrm{Q_N}$, the lower $\mathrm{Q_S}$, the left $\mathrm{Q_W}$, and the right $\mathrm{Q_E}$. Events in $\mathrm{Q_N}$ and $\mathrm{Q_S}$, hereafter referred to as CR1, are used to derive the reweighting function for the nominal background estimate, while an alternative reweighting function, derived from the events in $\mathrm{Q_E}$ and $\mathrm{Q_W}$ (referred to hereafter as CR2), is used to define a systematic uncertainty related to the reweighting function interpolation into the SR, as detailed in Section 7. The boundaries of the SR, CR1, and CR2 in the reconstructed (m_{H1}, m_{H2}) plane are shown in Figures 4 and 5, for the $4b$ and $2b$ data, respectively. The horizontal and vertical bands of lower event density around 80 GeV visible in these plots are caused by the X_{Wt} selection criterion. The reweighting function has the form:

$$w(\vec{x}) = \frac{p_{4b}(\vec{x})}{p_{2b}(\vec{x})}, \tag{4}$$



(a) (b)

Figure 5: The mass planes of the reconstructed Higgs boson candidates for the (a) ggF and (b) VBF signal regions of the analysis, shown for the $2b$ data events. The X_{Wt} variable in the figure denotes the minimum value of the X_{Wt} variable as obtained from the different combinations of central and b-tagged jets in the event. In (a), the analysis selection up to step 8 (as outlined in Figure 3) of the ggF selection has been applied, while in (b), the the analysis selection up to step 7 of the VBF selection has been applied. In both (a) and (b), the requirement in step 3 has been replaced with a selection requiring exactly two b-tagged jets in the event.

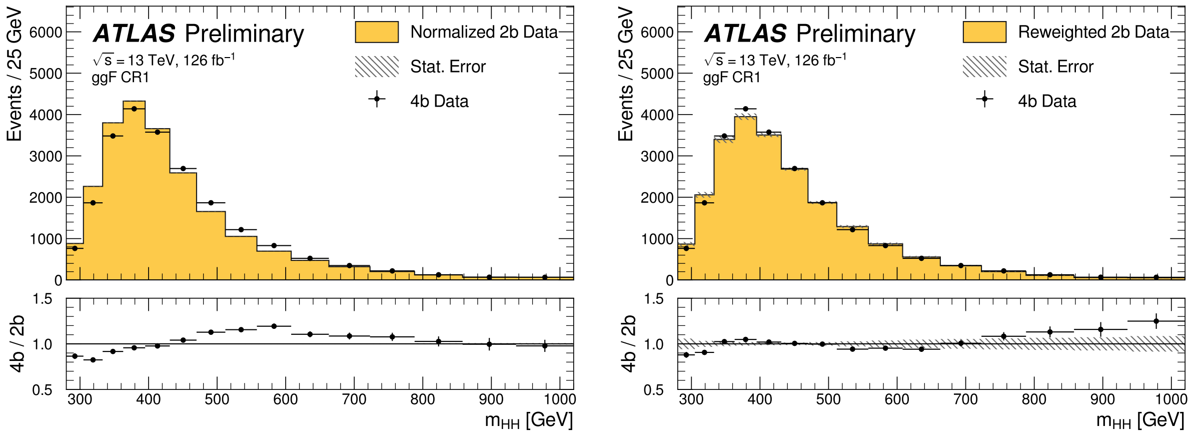
where $p_{2b}(\vec{x})$ and $p_{4b}(\vec{x})$ are the probability density functions for $2b$ and $4b$ data, respectively, over a set of kinematic variables $\vec{x}$. The computation of $w(\vec{x})$ is a density ratio estimation problem, for which a variety of approaches exist. The method employed in this analysis is modified from Refs. [77, 78] and makes use of an artificial neural network (NN). This NN is trained on $2b$ and $4b$ CR1 (or CR2) data to minimize the loss function:

$$\mathcal{L}(w(\vec{x})) = \int d\vec{x} \left[\sqrt{w(\vec{x})}\, p_{2b}(\vec{x}) + \frac{1}{\sqrt{w(\vec{x})}} p_{4b}(\vec{x}) \right].$$

The function in Eq. (4) optimizes this loss by equalizing the contributions from the two terms. The kinematic variables used to make up $\vec{x}$ are listed in Table 2 for the ggF and VBF selections; they are chosen to be sensitive to differences between the $2b$ and $4b$ events. The NN used in the ggF selection has three densely connected hidden layers of 50 nodes, each with a rectified linear unit activation function [79], and a single node linear output. A similar architecture is chosen for the NN used in the VBF selection, except for the smaller number of nodes - 20 in each of the three hidden layers, reflecting the significantly lower event statistics in the VBF selection relative to the ggF one, by nearly two orders of magnitude. This is also the reason behind the choice to perform the NN training in the VBF selection for all years together (with the year index as one-hot encoded input features), while for the ggF selected events, a dedicated reweighting is derived for each year separately, which, thanks to the adequate statistics, deals better with the different levels of disparity between $2b$ and $4b$ distributions, due to the differences in the trigger conditions from year to year. Finally, in order to ensure that there are adequate statistics of $4b$ events for the NN training, this is performed inclusively, before splitting the events into the $|\Delta\eta_{HH}|$ and X_{HH} categories. Both $|\Delta\eta_{HH}|$ and X_{HH} are found to be insensitive to the kinematic reweighting, and so the inclusive training is not expected to introduce any additional bias when splitting the events into the various categories.

Table 2: The set of input variables used for the $2b$ to $4b$ reweighting in the ggF and VBF channels respectively. The X_{Wt} variable in the table denotes the minimum value of the X_{Wt} variable as obtained from the different combinations of central and b-tagged jets in the event.

ggF	**VBF**
1. $\log(p_{\mathrm{T}})$ of the 2^{nd} leading Higgs boson candidate jet	1. Maximum di-jet mass out of the possible pairings of the four Higgs boson candidate jets
2. $\log(p_{\mathrm{T}})$ of the 4^{th} leading Higgs boson candidate jet	2. Minimum di-jet mass out of the possible pairings of the four Higgs boson candidate jets
3. $\log(\Delta R)$ between the closest two Higgs boson candidate jets	3. Energy of the leading Higgs boson candidate
4. $\log(\Delta R)$ between the other two Higgs boson candidate jets	4. Energy of the subleading Higgs boson candidate
5. Average absolute η value of the Higgs boson candidate jets	5. Second smallest ΔR between the jets in the leading Higgs boson candidate (out of the three possible pairings for the leading Higgs candidate)
6. $\log(p_{\mathrm{T}})$ of the di-Higgs system	6. Average absolute η value of Higgs boson candidate jets
7. ΔR between the two Higgs boson candidates	7. $\log(X_{Wt})$
8. $\Delta\phi$ between jets in the leading Higgs boson candidate	8. Trigger class index as one-hot encoder
9. $\Delta\phi$ between jets in the subleading Higgs boson candidate	9. Year index as one-hot encoder (for years inclusive training)
10. $\log(X_{Wt})$	
11. Number of jets in the event	
12. Trigger class index as one-hot encoder	

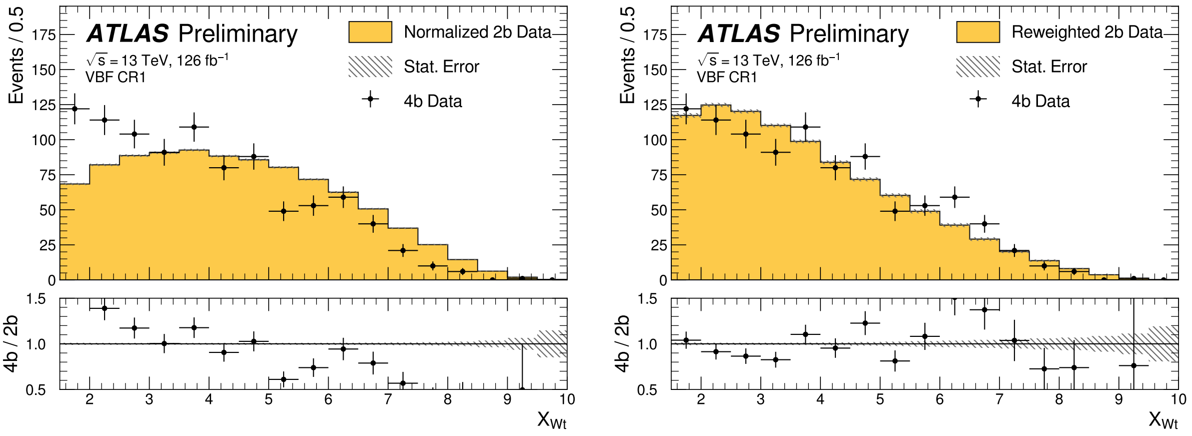


(a) (b)

Figure 6: Distributions of the m_{HH} variable as observed in the subset of $4b$ data falling within Control Region 1 (CR1) in the ggF signal region, shown with black points. In (a), the yellow histogram shows the m_{HH} distribution in the $2b$ CR1 dataset, with only a normalization factor applied. In (b), the yellow histogram shows the m_{HH} distribution in the $2b$ CR1 dataset after the NN reweighting. The hashes indicate the statistical uncertainty on the $2b$ data. In (a), the statistical uncertainty is only the Poisson uncertainty on the $2b$ data, while in (b), the statistical uncertainty also includes the uncertainty from the bootstrap procedure, as described in Section 7. Note that the hashes in (a) are smaller than the width of the plotted histogram.

In order to estimate and mitigate the impact of varying initial conditions and limited size of the training samples on the NN training, the Deep Ensembles technique [80] is used together with a bootstrap resampling [81] of the training data. This entails constructing a set of training datasets by sampling with replacement from the original dataset. In this analysis, this is approximated by the usage of different random training weights, following a Poisson distribution with $\mu = 1$, for each event in each training. The NN is trained independently on each element of this set, using different initial conditions each time. This results in an ensemble of reweighting functions. Each reweighting function is further multiplied by a normalization factor, such that the number of reweighted $2b$ events is equal to the $4b$ events in the region where the NN is trained. In this analysis, the ensembles comprise 100 reweighting functions each, hence 100 weights are calculated for each $2b$ event in the SR. The background estimate uses the mean of these weights for each event, and the variation of the background predictions from the ensemble of reweighting functions is used to derive a systematic uncertainty on the stability of the NN training procedure, as described in Section 7.

The effect of the above reweighting procedure in CR1, where the reweighting function is derived, is illustrated in Figure 6 for the m_{HH} distribution of the ggF-selected events and in Figure 7 for the X_{W_t} distribution of the VBF-selected events. The reweighted "$2b$" distributions agree with the corresponding "$4b$" distributions to within about 10% for the majority of the phase space, with some larger deviations observed in bins near the tails of the distributions where available data statistics are limited. While possible disagreements could be present in Figure 6, their impact on the ability to model the SR will be reflected in the non-closure uncertainties derived directly in the SR in the validation regions described below. A large number of additional kinematic variables are also studied before and after the reweighting was applied in order to validate the performance of the NN. For all variables, the agreement as quantified by the χ^2 metric either improves after the reweighting or, for variables where the "$2b$" and "$4b$" distributions were already similar, changes only slightly.



(a)

(b)

Figure 7: Distributions of the minimum value of the X_{Wt} variable, obtained from the different combinations of central and b-tagged jets in the event, as observed in the subset of $4b$ data falling within Control Region 1 (CR1) in the VBF signal region, shown with black points. In (a), the yellow histogram shows the X_{Wt} distribution in the $2b$ CR1 dataset, with only a normalization factor applied. In (b), the yellow histogram shows the X_{Wt} distribution in the $2b$ CR1 dataset after the NN reweighting. The hashes indicate the statistical uncertainty on the $2b$ data. In (a), the statistical uncertainty is only the Poisson uncertainty on the $2b$ data, while in (b), the statistical uncertainty also includes the uncertainty from the bootstrap procedure, as described in Section 7.

The robustness of the background modeling procedure was tested and shown to produce the expected results in a high-statistics simulated $t\bar{t}$ sample and a sample of simulated (non-$t\bar{t}$) multijet events, as well as in several control data samples, orthogonal to the nominal event selection, where the presence of any HH signal is negligible and the $4b$ events can be compared to the reweighted $2b$ events avoiding any bias. These samples, summarized in Table 3, include: (a) events satisfying all the $2b/4b$ ggF selection criteria, with the difference that the $|\Delta\eta_{HH}| < 1.5$ cut is inverted (requiring $|\Delta\eta_{HH}| > 1.5$); (b) events satisfying all the same $2b/4b$ selection criteria, except that the center of the SR (and hence also of CR1 and CR2) is shifted, so as to avoid any overlap with the nominal SR; and (c) events that satisfy all the same $4b$ selection criteria, except that, in terms of b-tagging, they contain exactly three b-tagged jets, and the fourth jet is taken as the highest-p_{T} jet that fails a looser working point of the b-tagging algorithm (one that gives 85% efficiency for b-jets in simulated $t\bar{t}$ events), hereafter referred to as $3b1$f. The $3b1$f sample has about one order of magnitude higher statistics than the $4b$ sample and a negligible amount of HH signal, hence it is used to derive a non-closure systematic uncertainty for the reweighting procedure, as discussed in Section 7.

Table 3: A summary of the different alternative data samples used in the analysis. For each sample, the nominal analysis selection is applied, except where noted explicitly.

Control Data Sample	**Definition**	**Usage**
Control Region (CR)	Events with $X_{HH} > 1.6$ and within the circle defined by: $\sqrt{\left(m_{H1} - 1.05 \cdot 124\,\text{GeV}\right)^2 + \left(m_{H2} - 1.05 \cdot 117\,\text{GeV}\right)^2} = 45\,\text{GeV}$	Background estimation (ggF and VBF)
$2b$	Remove the ≥ 4 b-tagged central jets selection and require exactly 2 b-tagged central jets plus two additional untagged central jets	Background estimation (ggF and VBF)
$3b1$f	Remove the ≥ 4 b-tagged central jets selection and require exactly 3 b-tagged central jets plus one central jet failing a looser b-tagging requirement	Background estimation validation (ggF and VBF), additional background modeling uncertainty (ggF only)
Reverse $\lvert\Delta\eta_{HH}\rvert$	Remove the $\lvert\Delta\eta_{HH}\rvert < 1.5$ selection and require $\lvert\Delta\eta_{HH}\rvert > 1.5$	Background estimation validation (ggF only)
Shifted region	Shift the center of the SR in the m_{H1}-m_{H2} plane to avoid overlap with the nominal SR	Background estimation validation (ggF only)

7 Systematic uncertainties

The uncertainties with the greatest impact on the analysis sensitivity are those arising from the data-driven background estimate described in Section 6. There are two main sources: uncertainties from the limited sample sizes in the CR and SR, and physical differences between the CR, where the $2b$ reweighting function is derived, and the SR, where it is applied. An additional uncertainty is obtained by comparing the result of the background estimation procedure to the $3b1\mathrm{f}$ dataset.

As described in Section 6, the ensemble of 100 reweighting functions results in 100 separate background predictions. An m_{HH} histogram can be constructed from each of these predictions, and the standard deviation of the predictions in each bin is taken as the bootstrap uncertainty. The uncertainty is treated as uncorrelated across m_{HH} bins.

An additional statistical uncertainty results from the limited sample size of the $2b$ SR dataset in which the trained background reweighting network is applied to obtain the final background estimation. A Poisson uncertainty is taken for each m_{HH} bin, which is combined in quadrature with the bootstrap uncertainty described above.

The uncertainty component in the background estimation relating to the kinematic differences between the SR and CR1 is derived by obtaining alternative predictions using the CR2 region. Using the same quadrants $\mathrm{Q_N}$, $\mathrm{Q_S}$, $\mathrm{Q_E}$, and $\mathrm{Q_W}$ defined in Section 6, four alternative background estimates are derived by applying weights derived in CR1 to three of the SR quadrants and weights derived in the CR2 to the one remaining SR quadrant. For example, one alternative background estimate is derived by applying CR1-derived weights to $\mathrm{Q_S}$, $\mathrm{Q_E}$ and $\mathrm{Q_W}$, and CR2-derived weights to $\mathrm{Q_N}$. Each of these four background predictions is symmetrized around the nominal m_{HH} distribution to construct a two-sided uncertainty. In the ggF signal region, these uncertainties are taken to be uncorrelated across the datasets from the three different years. In both the ggF and VBF signal regions, the uncertainty is treated as correlated across the analysis categories.

An additional closure uncertainty is derived by applying the full background modeling procedure to the $3b1\mathrm{f}$ sample instead of the $4b$ sample. The predicted $3b1\mathrm{f}$ m_{HH} distribution in the various analysis categories is then compared to the observed $3b1\mathrm{f}$ data in the SR. For the VBF signal region, no statistically significant deviation between the prediction and observation is found, hence no additional uncertainty is applied. For the ggF signal region, an additional uncertainty is derived in each category by taking the deviation between the predicted and observed $3b1\mathrm{f}$ m_{HH} distributions. For the ggF signal region, an additional uncertainty is derived in each category based on the observed deviations between the predicted and observed $3b1\mathrm{f}$ m_{HH} distributions. For m_{HH} bins in which the predicted and observed values differ by less than 1σ, where σ refers to all other background modeling uncertainties combined, no additional uncertainty is applied. For m_{HH} bins where the predicted and observed values differ by more than 1σ, the amount beyond 1σ is averaged with the corresponding amounts in the two adjacent bins, to limit the impact of statistical fluctuations, and is symmetrized around the nominal prediction to construct a two-sided uncertainty. This non-closure uncertainty has a much smaller impact on the analysis sensitivity than the other sources of background modeling uncertainty.

Several detector modeling uncertainties are evaluated and included. These affect only the signal description, as the background is estimated entirely from data. Uncertainties in the jet energy scale and resolution, as well as the JVT, are treated according to the prescription in Refs. [66, 67]. Additional uncertainties arising from the correction of the simulated pile-up distribution are treated according to the prescription in Ref. [82]. Uncertainties in the b-tagging efficiency are treated according to the prescription in Ref. [70].

Uncertainties in the trigger efficiencies are evaluated from measurements of per-jet online efficiencies for both jet reconstruction and b-tagging, which are used to compute event-level uncertainties. These are then applied to the simulated events as overall weight variations. The uncertainty in the total integrated luminosity used in this analysis is 1.7% [83], obtained using the LUCID-2 detector for the primary luminosity measurements [84].

Several sources of theoretical uncertainty affecting the signal models are considered as described below. Uncertainties due to modeling of the parton shower and underlying event are evaluated by comparing results between two generators for these parts of the calculation: the nominal Pythia 8 and the alternative Herwig 7. This is found to have an effect of roughly 10% on the ggF and VBF signal acceptances, and a negligible impact on the shape of the m_{HH} distributions. The parton showering uncertainty is derived within each analysis SR category; the uncertainty is observed to reach up to approximately 40% for a given production mode in some categories in which the acceptance is small for that mode. Uncertainties in the matrix element calculation are evaluated by raising and lowering the factorization and renormalization scales used in the generator by a factor of two, both independently and simultaneously. This results in an effect of typically 2% for both ggF and VBF, with a maximum effect of about 6% in certain analysis categories. PDF uncertainties are evaluated using the PDF4LHC_NLO_MC set [45] by calculating the signal acceptance for each replica and taking the standard deviation. The magnitude of this uncertainty is found to typically be less than 1% in both the ggF and VBF signal acceptances, with a maximum magnitude of approximately 2%. Theoretical uncertainties in the $H \to b\bar{b}$ branching ratio [15] are included; they amount to approximately a 3.5% overall uncertainty on the signal normalization. The dependence of the branching ratio uncertainty on κ_λ is neglected. Theoretical uncertainties on the ggF and VBF HH cross-sections arising from uncertainties on the PDF and α_s, as well as the choice of renormalization scheme and the scale of the top-quark mass, are taken from Refs. [12, 15, 85]. The cross-section uncertainties are included in the measurements of the upper limits on the ggF, VBF, and combined HH signal strengths, as well as the likelihood-based constraints on the values of the κ_λ and κ_{2V} modifiers, as presented in Section 8.

8 Results

The analysis results are obtained using a maximum likelihood fit performed in bins of reconstructed m_{HH}. For the ggF signal region, the fit is performed simultaneously across the different data-taking years (2016, 2017, and 2018), while for the VBF signal region, the fit is performed inclusively on the data from all years.

The likelihood function used to construct the test statistic has a standard form, consisting of a product of Poisson distributions for the yields in each bin and constraint functions for nuisance parameters describing systematic uncertainties. For uncertainties due to the limited sample size in data or simulation, the constraint is a Poisson distribution. For all other systematic uncertainties, the constraint is a Gaussian distribution. Any systematic uncertainty which is treated as uncorrelated between different categories or bins has a separate independent nuisance parameter for each of them. Uncertainties in the luminosity and signal modeling are treated as fully correlated between the analysis categories and, for ggF, the data-taking years. Each component of the quadrant-derived uncertainty covering the kinematic differences between the SR and CR1 regions is correlated across the data-taking years for the ggF region. The components are correlated across analysis categories within the ggF and VBF signal regions, but not between the ggF and

VBF signal regions. All other uncertainties in the background model are treated as uncorrelated across the different categories and data-taking years. The statistical model is implemented using RooFit [86].

For each configuration of coupling modifiers (κ_λ, κ_{2V}, and κ_V), the hypothesis of the presence of a signal is tested using the profile likelihood ratio [87]. The dependence of the $H \to b\bar{b}$ branching ratio on κ_λ and κ_V is ignored, and the value is fixed to the SM prediction. The signal strength of the combined ggF and VBF signal process $\mu_{\text{ggF+VBF}}$ is chosen as the parameter of interest (POI) and is a free parameter in the fit. The relative contributions of the ggF and VBF signals to the total signal model are fixed to their predicted values. The profile likelihood ratio takes the following form:

$$-2\Delta \ln \lambda(\mu) = -2 \ln \left(\frac{L(\mu, \hat{\hat{\boldsymbol{\theta}}})}{L(\hat{\mu}, \hat{\boldsymbol{\theta}})} \right),$$

where μ is the POI and $\boldsymbol{\theta}$ represents the nuisance parameters. The numerator represents the conditional maximum likelihood fit, in which the nuisance parameters are set to their profiled values $\hat{\hat{\boldsymbol{\theta}}}$ for which the likelihood is maximized for a fixed value of μ. The denominator represents the unconditional likelihood fit, where both μ and $\boldsymbol{\theta}$ are set to the values which jointly maximize the likelihood, $\hat{\mu}$ and $\hat{\boldsymbol{\theta}}$ respectively.

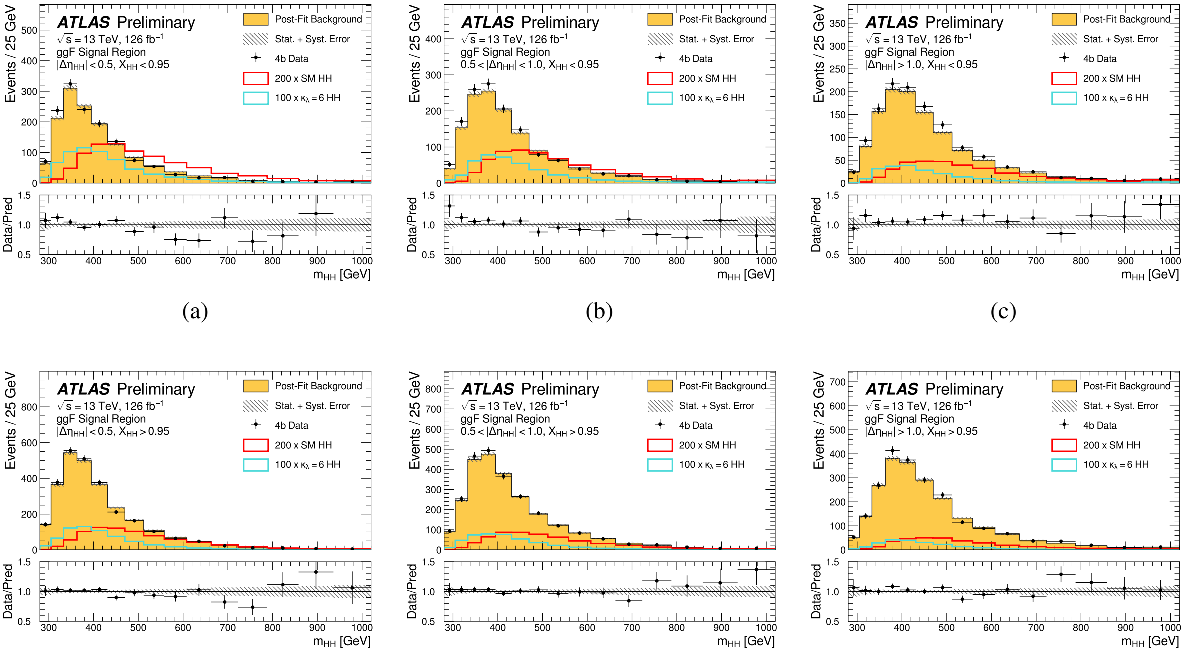
The observed distributions in m_{HH}, as well as the predicted background and example signal shapes, are presented in Figure 8 for each of the six ggF categories (with all data-taking years combined, for presentation purposes) and in Figure 9 (with each data-taking year presented separately). The distributions of the expected background are obtained using the best-fit values of the nuisance parameters in the background-only fit to the data.

The m_{HH} distributions in the two VBF categories are shown in Figure 10. The observed number of data events, predicted number of background events, and expected number of signal events for the SM ggF and VBF signals are summarized for each of the analysis categories in Table 4.

Table 4: The yields in each analysis category of the data, expected background, and expected SM ggF and VBF signals. The expected background yields are obtained using a background-only fit to the data. The expected signal yields are obtained from simulation.

Category	**Data**	**Expected Background**	**ggF Signal SM**	**VBF Signal SM**
ggF signal region				
$\lvert\Delta\eta_{HH}\rvert < 0.5$, $X_{HH} < 0.95$	1940	1940 ± 130	6.99	0.038
$\lvert\Delta\eta_{HH}\rvert < 0.5$, $X_{HH} > 0.95$	3602	3620 ± 200	6.49	0.036
$0.5 < \lvert\Delta\eta_{HH}\rvert < 1.0$, $X_{HH} < 0.95$	1924	1870 ± 120	5.15	0.037
$0.5 < \lvert\Delta\eta_{HH}\rvert < 1.0$, $X_{HH} > 0.95$	3540	3490 ± 190	4.75	0.040
$\lvert\Delta\eta_{HH}\rvert > 1.0$, $X_{HH} < 0.95$	1880	1740 ± 120	2.92	0.043
$\lvert\Delta\eta_{HH}\rvert > 1.0$, $X_{HH} > 0.95$	3285	3210 ± 200	2.81	0.041
VBF signal region				
$\lvert\Delta\eta_{HH}\rvert < 1.5$	116	125 ± 12	0.37	0.090
$\lvert\Delta\eta_{HH}\rvert > 1.5$	241	231 ± 20	0.06	0.207

In the absence of an observed signal, exclusion limits on the cross-section as a function of the coupling modifiers $\sigma_{\text{ggF+VBF}}(\kappa_\lambda, \kappa_V, \kappa_{2V})$ are computed using the asymptotic formula [87] and based on the CL_s method [88]. A cross-section value is considered to be excluded at the 95% confidence level (CL) when



(a) (b) (c)

(d) (e) (f)

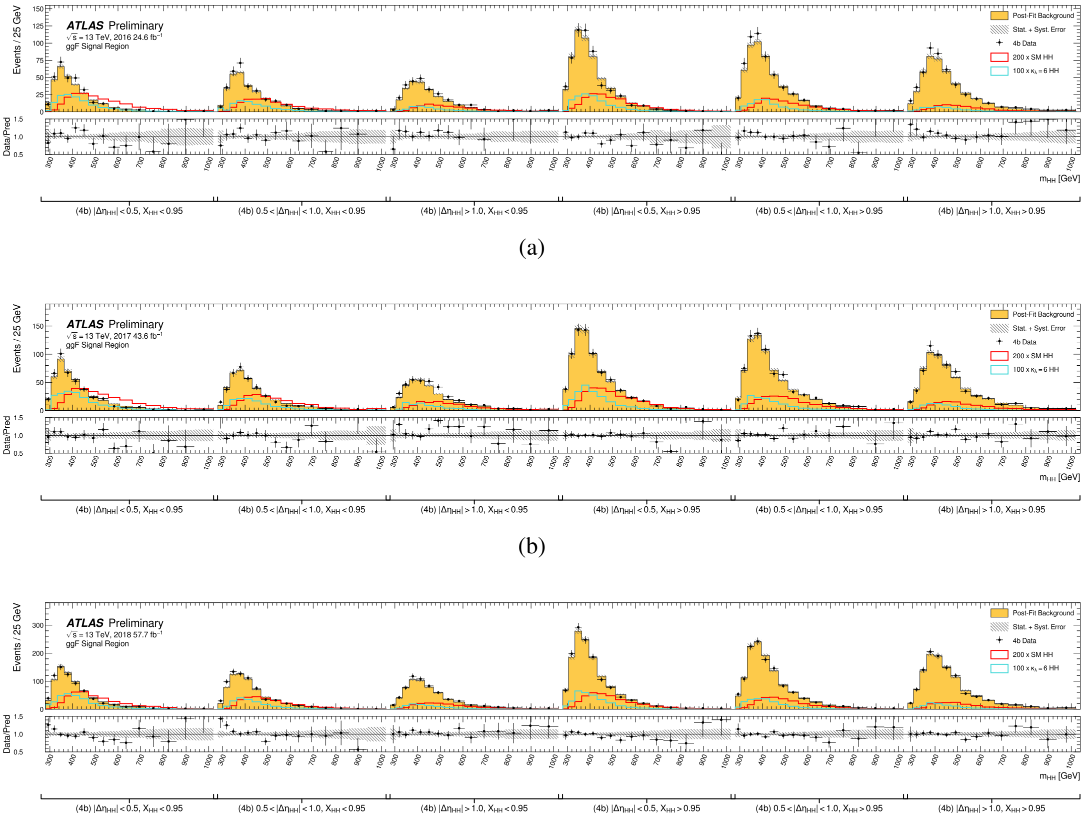
Figure 8: Distributions of the reconstructed m_{HH} in data (shown by the black points) and the estimated background (shown by the yellow histograms), in each of the six $|\Delta\eta_{HH}|/X_{HH}$ categories in the ggF signal region. The contribution from the different data-taking years are combined in each category for presentation purposes. The black hashed band shows the total uncertainty on the background estimation. The distribution of the expected background is obtained using the best-fit values of the nuisance parameters in the background-only fit to the data. Distributions of the SM and $\kappa_\lambda = 6$ signal models are overlaid, scaled so as to be visible on the plot, and the scaling for each signal model is the same across the six categories. The lower panels show the ratio of the predicted background to the observed data yield in each bin. Events in the underflow and overflow bins are counted in the yields of the initial and final bins respectively.

the $\mathrm{CL_s}$ is less than 0.05 . By extension, values of the coupling modifiers $(\kappa_\lambda, \kappa_{2V})$ are excluded if the predicted cross-section of the signal model obtained with that configuration is excluded with the $\mathrm{CL_s}$ method. Upper limits on the HH cross-section as a function of κ_λ and κ_{2V} are shown in Figure 11, and the exclusion boundaries are summarized in Table 5. With the values of the other modifiers (κ_V and either κ_{2V} or κ_λ, respectively) fixed to their SM value of 1, values of κ_λ beyond $[-3.9, 11.1]$ and values of κ_{2V} beyond $[-0.03, 2.11]$ are excluded.

Table 5: The observed and expected constraints on the κ_λ and κ_{2V} coupling modifiers at 95% CL. For each modifier, the constraints were extracted with all other modifiers fixed to the SM prediction.

Parameter	**Expected Constraint**		**Observed Constraint**	
	Lower	Upper	Lower	Upper
κ_λ	−4.6	10.8	−3.9	11.1
κ_{2V}	−0.05	2.12	−0.03	2.11

Alternative exclusion limits on the coupling modifiers are obtained by using the profile likelihood ratio



(a)

(b)

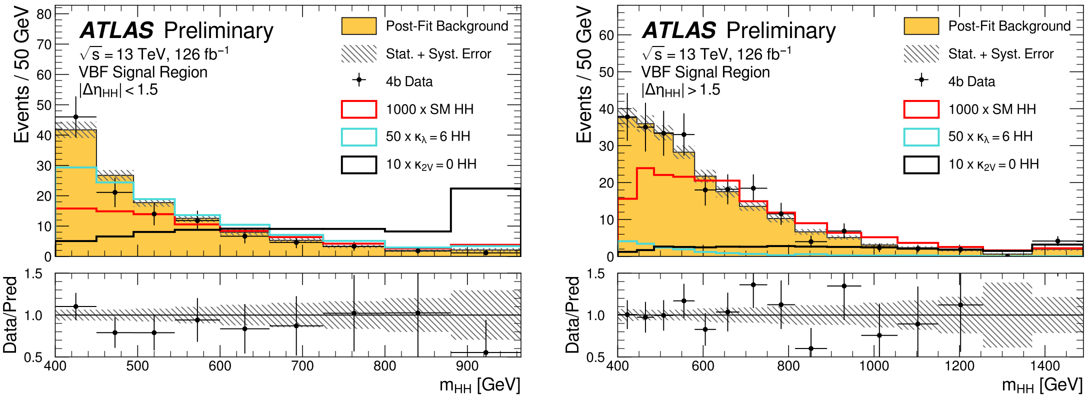
(c)

Figure 9: Distributions of the reconstructed m_{HH} in data (shown by the black points) and the estimated background (shown by the yellow histograms), in each of the six $|\Delta\eta_{HH}|/X_{HH}$ categories in the ggF signal region, for the (a) 2016, (b) 2017, and (c) 2018 data. The black hashed band shows the total uncertainty on the background estimation. The distribution of the expected background is obtained using the best-fit values of the nuisance parameters in the background-only fit to the data. Distributions of the SM and $\kappa_\lambda = 6$ signal models are overlaid, scaled so as to be visible on the plot, and the scaling for each signal model is the same across the six categories. The lower panels show the ratio of the predicted background to the observed data yield in each bin. Events in the underflow and overflow bins are counted in the yields of the initial and final bins respectively.

with the coupling modifiers $\boldsymbol{\kappa} = (\kappa_\lambda, \kappa_V, \kappa_{2V})$ as the POIs, rather than the signal strength μ:

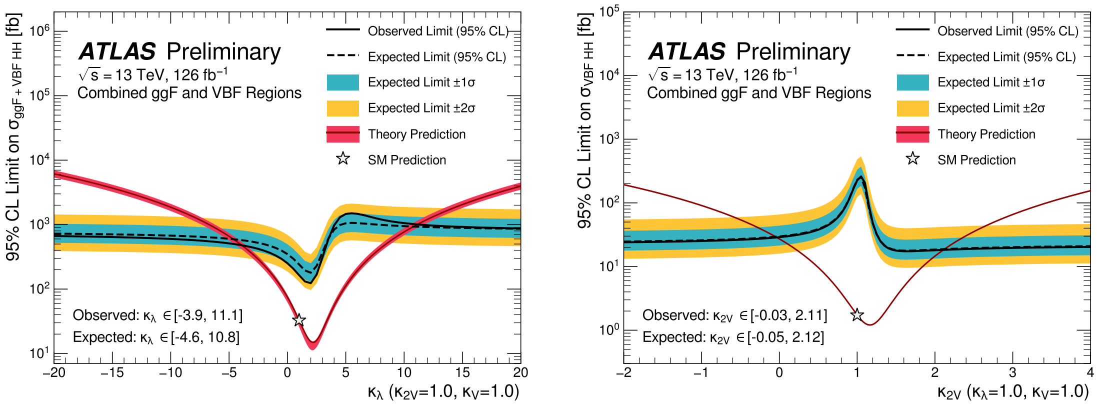
$$-2\Delta \ln \lambda(\boldsymbol{\kappa}) = -2 \ln \left(\frac{L(\boldsymbol{\kappa}, \hat{\hat{\boldsymbol{\theta}}})}{L(\hat{\boldsymbol{\kappa}}, \hat{\boldsymbol{\theta}})} \right),$$

A scan of the profile likelihood ratio is taken as a function of the coupling modifiers at discrete points to produce the curves shown in Figure 12. The exclusion constraints obtained from the profile likelihood ratio scan are not expected to be identical to those of the 95% CL methodology described above, as the two strategies employ slightly different physical assumptions. In the profile likelihood ratio scan, the signal strength is fixed to the prediction obtained for a specific coupling modifier configuration, while for the 95% CL scan, the signal strength is allowed to float. Additionally, 2σ-level constraints are quoted from the profile likelihood ratio scan, as opposed to 95%-level constraints. The best-fit value of κ_λ is found to be 6.2 from the profile likelihood scan, although the SM value is not excluded from the scan. The observed pull of the best-fit κ_λ value away from the SM value is understood to be due to an upward fluctuation in the



(a) (b)

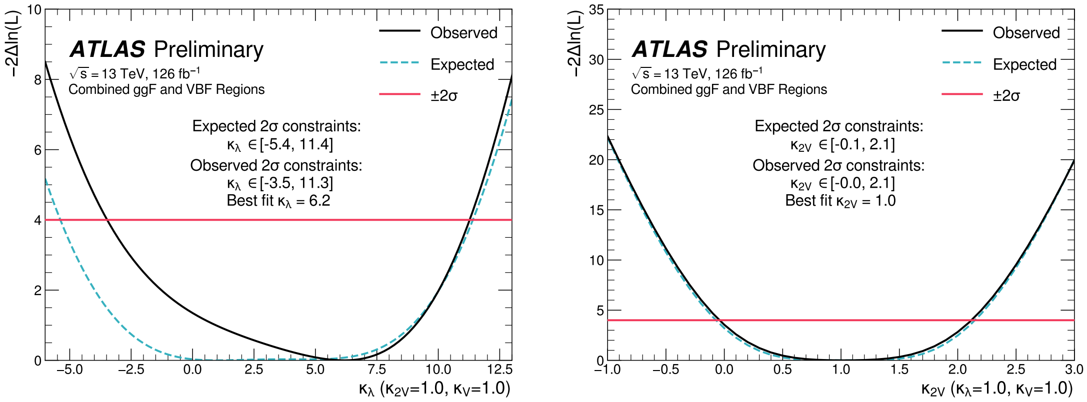
Figure 10: Distributions of the reconstructed m_{HH} in data (shown by the black points), the estimated background (shown by the yellow histograms), in each of the two $|\Delta\eta_{HH}|$ categories in the VBF signal region. The black hashed band shows the total uncertainty on the background estimation. The distribution of the expected background is obtained using the best-fit values of the nuisance parameters in the background-only fit to the data. Distributions for three choices of couplings are shown: the SM, $\kappa_\lambda = 6$, and $\kappa_{2V} = 0$ (with all other couplings set to their SM values in the latter two models), scaled so as to be visible on the plot, and the scaling for each signal model is the same across the two categories. The lower panels show the ratio of the predicted background to the observed data yield in each bin. Events in the overflow bins are counted in the yields of the final bins.



(a) (b)

Figure 11: The observed 95% $\mathrm{CL_s}$ exclusion limits as a function of (a) κ_λ (obtained using the signal strength $\mu_{\mathrm{ggF+VBF}}$ as the POI) and (b) κ_{2V} (obtained using the signal strength μ_{VBF} as the POI) from the combined ggF and VBF signal regions, as shown by the solid black line. The values of the other two parameters (of κ_λ, κ_{2V}, and κ_V) are fixed to 1. The blue and yellow bands show the 1σ and 2σ bands respectively on the expected exclusion limits, which are shown by the dashed black line. The expected exclusion limits are obtained using a background-only fit to the data. The red line shows (a) the predicted combined ggF and VBF HH cross-section as a function of κ_λ and (b) the predicted VBF HH cross-section as a function of κ_{2V}. The magenta band surrounding the red cross-section lines indicate the theoretical uncertainty on the cross-section, as taken from Ref. [85]. Note that the band in (b) is smaller than the width of the plotted line.

observed data in the ggF signal region, specifically in the low-m_{HH} range. This fluctuation also results in the deviation of the observed limits in Figure 11(a) with respect to the expected limits of about 1σ. No such fluctuation is observed in the VBF signal region, and the best-fit value of κ_{2V} from the likelihood scan is found to be 1.



(a) (b)

Figure 12: The observed profile likelihood ratio scans for the (a) κ_{λ} and (b) κ_{2V} coupling modifiers, shown by the solid black line, using the coupling modifiers $\boldsymbol{\kappa}$ as the POIs. The values of the other two parameters (κ_V and κ_{2V} in (a) and κ_V and κ_{λ} in (b)) are fixed to 1. The dashed blue line shows the expected exclusion limits, as obtained using a background-only fit to the data. The pink line indicates the 2σ exclusion boundary.

Upper limits on the cross-sections of the individual *HH* production modes considered, ggF and VBF, assuming SM coupling modifiers, are presented in Table 6. These limits are obtained using the asymptotic formula and the CL_s method, using the signal strengths μ_{ggF} or μ_{VBF} as the POI. The table also includes the upper limits on the combined ggF and VBF *HH* cross-section, assuming SM coupling modifiers. For the individual σ_{ggF} and σ_{VBF} limits, the results are derived by treating the other production mode (VBF when placing limits on σ_{ggF}, and vice-versa) as a background process, with its normalization only loosely constrained in the fit. For the combined $\sigma_{\text{ggF+VBF}}$ upper limit, the relative ratio of the two production modes is fixed to the SM prediction. The total uncertainty on the upper limit of the cross-section is dominated by the uncertainty sources related to the background modelling procedure and theoretical predictions. With only the statistical uncertainties of the reweighted $2b$ data, observed $4b$ data, and simulated signal samples included in the fit, the expected upper limit on the combined $\sigma_{\text{ggF+VBF}}$ is found to be 6.0 times the SM prediction. Including the uncertainty sources resulting from the background estimation (including the bootstrap uncertainty, the uncertainty from the kinematic differences between the SR and CR1, and, in the ggF signal region, the $3b1$f non-closure uncertainty), the expected upper limit on the combined $\sigma_{\text{ggF+VBF}}$ is relaxed to 7.1 times the SM prediction. Including all sources of uncertainty results in the upper limit relaxing to 8.1 times the SM prediction, as quoted in Table 6, driven primarily by the uncertainties arising from theoretical predictions.

Table 6: The observed and expected upper limits on the SM ggF HH production cross-section σ_{ggF}, SM VBF HH production cross-section σ_{VBF}, and combined SM ggF and VBF HH production cross-section $\sigma_{\text{ggF+VBF}}$ at the 95% CL, expressed as ratios to the corresponding SM cross-sections. The expected values are shown with corresponding one and two standard deviation error bounds, and they are obtained using a background-only fit to the data. When extracting the limits on $\sigma_{\text{ggF+VBF}}$, the relative contributions of ggF and VBF production to the total cross-section are fixed to the SM prediction.

	Observed Limit	**-2σ**	**-1σ**	**Expected Limit**	**$+1\sigma$**	**$+2\sigma$**
$\sigma_{\text{ggF}}/\sigma_{\text{ggF}}^{\text{SM}}$	5.5	4.4	5.9	8.2	12.4	19.6
$\sigma_{\text{VBF}}/\sigma_{\text{VBF}}^{\text{SM}}$	130.5	71.6	96.1	133.4	192.9	279.3
$\sigma_{\text{ggF+VBF}}/\sigma_{\text{ggF+VBF}}^{\text{SM}}$	5.4	4.3	5.8	8.1	12.2	19.1

9 Conclusion

A search for non-resonant pair production of Higgs bosons in the $b\bar{b}b\bar{b}$ final state was carried out, with dedicated analyses in the ggF and VBF production modes, using 126 fb^{-1} of LHC pp collision data collected by the ATLAS detector at $\sqrt{s} = 13$ TeV. The sensitivity of the analyses is improved relative to previous iterations by using more sophisticated background modeling techniques, event categorization and improved jet reconstruction and flavour identification algorithms, in addition to the increased integrated luminosity of the analysed data.

No evidence for signal is found and the observed (expected) upper limit on the cross section for non-resonant Higgs boson pair production is determined to be 5.4 (8.1) times the Standard Model predicted cross-section at 95% confidence level. Constraints are placed upon modifiers to the HHH and $HHVV$ couplings. The observed (expected) constraints on the HHH coupling modifier, κ_{λ}, are determined to be [-3.9, 11.1] ([-4.6, 10.8]) at 95% confidence level, while the corresponding constraints for the $HHVV$ coupling modifier, κ_{2V}, are [-0.03, 2.11] ([-0.05, 2.12]).

References

[1] ATLAS Collaboration, *Observation of a new particle in the search for the Standard Model Higgs boson with the ATLAS detector at the LHC*, Phys. Lett. B **716** (2012) 1, arXiv: `1207.7214 [hep-ex]` (cit. on p. 2).

[2] CMS Collaboration, *Observation of a new boson at a mass of 125 GeV with the CMS experiment at the LHC*, Phys. Lett. B **716** (2012) 30, arXiv: `1207.7235 [hep-ex]` (cit. on p. 2).

[3] ATLAS Collaboration, *Observation of $H \rightarrow b\bar{b}$ decays and VH production with the ATLAS detector*, Phys. Lett. B **786** (2018) 59, arXiv: `1808.08238 [hep-ex]` (cit. on p. 2).

[4] CMS Collaboration, *Observation of Higgs Boson Decay to Bottom Quarks*, Phys. Rev. Lett. **121** (2018) 121801, arXiv: `1808.08242 [hep-ex]` (cit. on p. 2).

[5] G. C. Branco et al., *Theory and phenomenology of two-Higgs-doublet models*, Phys. Rept. **516** (2012) 1, arXiv: `1106.0034 [hep-ph]` (cit. on p. 2).

[6] A. Djouadi, *The Anatomy of electro-weak symmetry breaking. II. The Higgs bosons in the minimal supersymmetric model*, Phys. Rept. **459** (2008) 1, arXiv: `hep-ph/0503173` (cit. on p. 2).

[7] A. Djouadi and J. Quevillon, *The MSSM Higgs sector at a high M_{SUSY}: reopening the low tanβ regime and heavy Higgs searches*, JHEP **10** (2013) 028, arXiv: `1304.1787 [hep-ph]` (cit. on p. 2).

[8] A. Djouadi et al., *The post-Higgs MSSM scenario: habemus MSSM?* Eur. Phys. J. C **73** (2013) 2650, arXiv: `1307.5205 [hep-ph]` (cit. on p. 2).

[9] A. Djouadi, L. Maiani, A. Polosa, J. Quevillon, and V. Riquer, *Fully covering the MSSM Higgs sector at the LHC*, JHEP **06** (2015) 168, arXiv: `1502.05653 [hep-ph]` (cit. on p. 2).

[10] A. Hill and J. J. van der Bij, *Strongly interacting singlet-doublet Higgs model*, Phys. Rev. D **36** (11 1987) 3463, URL: `https://link.aps.org/doi/10.1103/PhysRevD.36.3463` (cit. on p. 2).

[11] T. Robens and T. Stefaniak, *Status of the Higgs Singlet Extension of the Standard Model after LHC Run 1*, Eur. Phys. J. C **75** (2015) 104, arXiv: `1501.02234 [hep-ph]` (cit. on p. 2).

[12] M. Grazzini et al., *Higgs boson pair production at NNLO with top quark mass effects*, JHEP **05** (2018) 059, arXiv: `1803.02463 [hep-ph]` (cit. on pp. 2, 5, 18).

[13] F. A. Dreyer and A. Karlberg, *Vector-boson fusion Higgs pair production at N^3LO*, Phys. Rev. D **98** (2018) 114016, arXiv: `1811.07906 [hep-ph]` (cit. on pp. 2, 5).

[14] F. Bishara, R. Contino, and J. Rojo, *Higgs pair production in vector-boson fusion at the LHC and beyond*, Eur. Phys. J. C **77** (2017) 481, arXiv: `1611.03860 [hep-ph]` (cit. on p. 2).

[15] D. de Florian et al., *Handbook of LHC Higgs Cross Sections: 4. Deciphering the Nature of the Higgs Sector*, (2016), arXiv: `1610.07922 [hep-ph]` (cit. on pp. 2, 18).

[16] A. Djouadi, J. Kalinowski, and M. Spira, *HDECAY: A program for Higgs boson decays in the Standard Model and its supersymmetric extension*, Comput. Phys. Commun. **108** (1998) 56, arXiv: `hep-ph/9704448` (cit. on p. 2).

[17] ATLAS Collaboration, *Search for pair production of Higgs bosons in the $b\bar{b}b\bar{b}$ final state using proton–proton collisions at $\sqrt{s} = 13$ TeV with the ATLAS detector*, JHEP **01** (2019) 030, arXiv: `1804.06174 [hep-ex]` (cit. on p. 3).

[18] ATLAS Collaboration, *Search for the $HH \rightarrow b\bar{b}b\bar{b}$ process via vector-boson fusion production using proton–proton collisions at $\sqrt{s} = 13$ TeV with the ATLAS detector*, JHEP **07** (2020) 108, arXiv: `2001.05178 [hep-ex]`, [Erratum: JHEP **01** (2021) 145] (cit. on p. 3).

[19] CMS Collaboration, *Search for Higgs boson pair production in the four b quark final state in proton-proton collisions at $\sqrt{s} = 13$ TeV*, (2022), arXiv: `2202.09617 [hep-ex]` (cit. on p. 3).

[20] CMS Collaboration, *Search for nonresonant pair production of highly energetic Higgs bosons decaying to bottom quarks*, (2022), arXiv: `2205.06667 [hep-ex]` (cit. on p. 3).

[21] ATLAS Collaboration, *Search for Higgs boson pair production in the two bottom quarks plus two photons final state in pp collisions at $\sqrt{s} = 13$ TeV with the ATLAS detector*, (2021), arXiv: `2112.11876 [hep-ex]` (cit. on p. 3).

[22] ATLAS Collaboration, *Search for non-resonant Higgs boson pair production in the $bb\ell\nu\ell\nu$ final state with the ATLAS detector in pp collisions at $\sqrt{s} = 13$ TeV*, Phys. Lett. B **801** (2020) 135145, arXiv: `1908.06765 [hep-ex]` (cit. on p. 3).

[23] ATLAS Collaboration, *Search for Higgs boson pair production in the $WW^{(*)}WW^{(*)}$ decay channel using ATLAS data recorded at $\sqrt{s} = 13$ TeV*, JHEP **05** (2019) 124, arXiv: `1811.11028 [hep-ex]` (cit. on p. 3).

[24] ATLAS Collaboration, *Search for Higgs boson pair production in the $b\bar{b}WW^*$ decay mode at $\sqrt{s} = 13$ TeV with the ATLAS detector*, JHEP **04** (2019) 092, arXiv: `1811.04671 [hep-ex]` (cit. on p. 3).

[25] ATLAS Collaboration, *Search for resonant and non-resonant Higgs boson pair production in the $b\bar{b}\tau^+\tau^-$ decay channel using 13 TeV pp collision data from the ATLAS detector*, ATLAS-CONF-2021-030, 2021, URL: `https://cds.cern.ch/record/2777236` (cit. on p. 3).

[26] ATLAS Collaboration, *Search for Higgs boson pair production in the $\gamma\gamma WW^*$ channel using pp collision data recorded at $\sqrt{s} = 13$ TeV with the ATLAS detector*, Eur. Phys. J. C **78** (2018) 1007, arXiv: `1807.08567 [hep-ex]` (cit. on p. 3).

[27] CMS Collaboration, *Search for Higgs boson pair production in the four b quark final state in proton–proton collisions at $\sqrt{s} = 13$ TeV*, (2022), arXiv: `2202.09617 [hep-ex]` (cit. on p. 3).

[28] CMS Collaboration, *Search for heavy resonances decaying to a pair of Lorentz-boosted Higgs bosons in final states with leptons and a bottom quark pair at $\sqrt{s} = 13$ TeV*, (2021), arXiv: `2112.03161 [hep-ex]` (cit. on p. 3).

[29] CMS Collaboration, *Search for nonresonant Higgs boson pair production in final states with two bottom quarks and two photons in proton–proton collisions at $\sqrt{s} = 13$ TeV*, JHEP **03** (2021) 257, arXiv: `2011.12373 [hep-ex]` (cit. on p. 3).

[30] CMS Collaboration, *Search for resonant and nonresonant Higgs boson pair production in the $b\bar{b}\ell\nu\ell\nu$ final state in proton–proton collisions at $\sqrt{s} = 13$ TeV*, JHEP **01** (2018) 054, arXiv: `1708.04188 [hep-ex]` (cit. on p. 3).

[31] ATLAS Collaboration, *The ATLAS Experiment at the CERN Large Hadron Collider*, JINST **3** (2008) S08003 (cit. on p. 3).

[32] ATLAS Collaboration, *ATLAS Insertable B-Layer: Technical Design Report*, ATLAS-TDR-19; CERN-LHCC-2010-013, 2010, URL: `https://cds.cern.ch/record/1291633` (cit. on p. 4), Addendum: ATLAS-TDR-19-ADD-1; CERN-LHCC-2012-009, 2012, URL: `https://cds.cern.ch/record/1451888`.

[33] B. Abbott et al., *Production and integration of the ATLAS Insertable B-Layer*, JINST **13** (2018) T05008, arXiv: `1803.00844 [physics.ins-det]` (cit. on p. 4).

[34] ATLAS Collaboration, *Performance of the ATLAS trigger system in 2015*, Eur. Phys. J. C **77** (2017) 317, arXiv: `1611.09661 [hep-ex]` (cit. on p. 4).

[35] ATLAS Collaboration, *The ATLAS Collaboration Software and Firmware*, ATL-SOFT-PUB-2021-001, 2021, URL: `https://cds.cern.ch/record/2767187` (cit. on p. 4).

[36] ATLAS Collaboration, *ATLAS data quality operations and performance for 2015–2018 data-taking*, JINST **15** (2020) P04003, arXiv: `1911.04632 [physics.ins-det]` (cit. on p. 4).

[37] ATLAS Collaboration, *Configuration and performance of the ATLAS b-jet triggers in Run 2*, Eur. Phys. J. C **81** (2021) 1087, arXiv: `2106.03584 [hep-ex]` (cit. on p. 4).

[38] M. Cacciari, G. P. Salam, and G. Soyez, *The anti-k_t jet clustering algorithm*, JHEP **04** (2008) 063, arXiv: `0802.1189 [hep-ph]` (cit. on p. 4).

[39] M. Cacciari, G. P. Salam, and G. Soyez, *FastJet user manual*, Eur. Phys. J. C **72** (2012) 1896, arXiv: `1111.6097 [hep-ph]` (cit. on p. 4).

[40] ATLAS Collaboration, *The ATLAS Simulation Infrastructure*, Eur. Phys. J. C **70** (2010) 823, arXiv: `1005.4568 [physics.ins-det]` (cit. on p. 5).

[41] GEANT4 Collaboration, S. Agostinelli, et al., *Geant4 – a simulation toolkit*, Nucl. Instrum. Meth. A **506** (2003) 250 (cit. on p. 5).

[42] P. Nason, *A New method for combining NLO QCD with shower Monte Carlo algorithms*, JHEP **11** (2004) 040, arXiv: `hep-ph/0409146 [hep-ph]` (cit. on p. 5).

[43] S. Frixione, P. Nason and C. Oleari, *Matching NLO QCD computations with parton shower simulations: the POWHEG method*, JHEP **11** (2007) 070, arXiv: `0709.2092 [hep-ph]` (cit. on p. 5).

[44] S. Alioli, P. Nason, C. Oleari, and E. Re, *A general framework for implementing NLO calculations in shower Monte Carlo programs: the POWHEG BOX*, JHEP **06** (2010) 043, arXiv: `1002.2581 [hep-ph]` (cit. on p. 5).

[45] J. Butterworth et al., *PDF4LHC recommendations for LHC Run II*, J. Phys. G **43** (2016) 023001, arXiv: `1510.03865 [hep-ph]` (cit. on pp. 5, 18).

[46] T. Sjostrand et al., *An Introduction to PYTHIA 8.2*, Comput.Phys.Commun. **191** (2015) 159, arXiv: `1410.3012 [hep-ph]` (cit. on p. 5).

[47] ATLAS Collaboration, *ATLAS Pythia 8 tunes to 7 TeV data*, ATL-PHYS-PUB-2014-021, 2014, URL: `https://cds.cern.ch/record/1966419` (cit. on p. 5).

[48] R. D. Ball et al., *Parton distributions with LHC data*, Nucl. Phys. B **867** (2013) 244, arXiv: `1207.1303 [hep-ph]` (cit. on pp. 5, 6).

[49] ATLAS Collaboration, *Validation of signal Monte Carlo event generation in searches for Higgs boson pairs with the ATLAS detector*, ATL-PHYS-PUB-2019-007, 2019, URL: `https://cds.cern.ch/record/2665057` (cit. on p. 5).

[50] M. Bähr et al., *Herwig++ Physics and Manual*, Eur. Phys. J. **C58** (2008) 639, arXiv: `0803.0883 [hep-ph]` (cit. on p. 5).

[51] S. Gieseke, C. Röhr, and A. Siodmok, *Colour reconnections in Herwig++*, Eur. Phys. J. C **72** (2012) 2225, arXiv: `1206.0041 [hep-ph]` (cit. on p. 5).

[52] L. Harland-Lang, A. Martin, P. Motylinski, and R. Thorne, *Parton distributions in the LHC era: MMHT 2014 PDFs*, Eur. Phys. J. C **75** (2015) 204, arXiv: `1412.3989 [hep-ph]` (cit. on p. 5).

[53] J. Alwall et al., *The automated computation of tree-level and next-to-leading order differential cross sections, and their matching to parton shower simulations*, JHEP **07** (2014) 079, arXiv: `1405.0301 [hep-ph]` (cit. on p. 5).

[54] ATLAS Collaboration, *Combination of searches for Higgs boson pairs in pp collisions at $\sqrt{s} = 13$ TeV with the ATLAS detector*, Phys. Lett. B **800** (2020) 135103, arXiv: `1906.02025 [hep-ex]` (cit. on p. 5).

[55] J. Baglio et al., *The measurement of the Higgs self-coupling at the LHC: theoretical status*, JHEP **04** (2013) 151, arXiv: `1212.5581 [hep-ph]` (cit. on p. 5).

[56] L.-S. Ling et al., *NNLO QCD corrections to Higgs pair production via vector boson fusion at hadron colliders*, Phys. Rev. D **89** (2014) 073001, arXiv: `1401.7754 [hep-ph]` (cit. on p. 5).

[57] F. A. Dreyer and A. Karlberg, *Fully differential Vector-Boson Fusion Higgs Pair Production at Next-to-Next-to-Leading Order*, Phys. Rev. D **99** (2019) 074028, arXiv: `1811.07918 [hep-ph]` (cit. on p. 5).

[58] J. M. Campbell, R. K. Ellis, P. Nason, and E. Re, *Top-pair production and decay at NLO matched with parton showers*, JHEP **04** (2015) 114, arXiv: `1412.1828 [hep-ph]` (cit. on p. 6).

[59] R. D. Ball et al., *Parton distributions for the LHC run II*, JHEP **04** (2015) 040, arXiv: `1410.8849 [hep-ph]` (cit. on p. 6).

[60] M. Czakon and A. Mitov, *Top++: A program for the calculation of the top-pair cross-section at hadron colliders*, Comput. Phys. Commun. **185** (2014) 2930, arXiv: `1112.5675 [hep-ph]` (cit. on p. 6).

[61] ATLAS Collaboration, *Improvements in $t\bar{t}$ modelling using NLO+PS Monte Carlo generators for Run 2*, ATL-PHYS-PUB-2018-009, 2018, URL: `https://cds.cern.ch/record/2630327` (cit. on p. 6).

[62] ATLAS Collaboration, *The Pythia 8 A3 tune description of ATLAS minimum bias and inelastic measurements incorporating the Donnachie–Landshoff diffractive model*, ATL-PHYS-PUB-2016-017, 2016, URL: `https://cds.cern.ch/record/2206965` (cit. on p. 6).

[63] D. J. Lange, *The EvtGen particle decay simulation package*, Nucl. Instrum. Meth. A **462** (2001) 152 (cit. on p. 6).

[64] ATLAS Collaboration, *Vertex Reconstruction Performance of the ATLAS Detector at $\sqrt{s} = 13$ TeV*, ATL-PHYS-PUB-2015-026, 2015, URL: `https://cds.cern.ch/record/2037717` (cit. on p. 6).

[65] ATLAS Collaboration, *Jet reconstruction and performance using particle flow with the ATLAS Detector*, Eur. Phys. J. C **77** (2017) 466, arXiv: `1703.10485 [hep-ex]` (cit. on p. 6).

[66] ATLAS Collaboration, *Jet energy scale and resolution measured in proton–proton collisions at $\sqrt{s} = 13$ TeV with the ATLAS detector*, Eur. Phys. J. C **81** (2020) 689, arXiv: `2007.02645 [hep-ex]` (cit. on pp. 6, 17).

[67] ATLAS Collaboration, *Performance of pile-up mitigation techniques for jets in pp collisions at $\sqrt{s} = 8$ TeV using the ATLAS detector*, Eur. Phys. J. C **76** (2016) 581, arXiv: `1510.03823 [hep-ex]` (cit. on pp. 6, 17).

[68] ATLAS Collaboration, *Topological cell clustering in the ATLAS calorimeters and its performance in LHC Run 1*, Eur. Phys. J. C **77** (2017) 490, arXiv: `1603.02934 [hep-ex]` (cit. on p. 6).

[69] ATLAS Collaboration, *Selection of jets produced in 13 TeV proton–proton collisions with the ATLAS detector*, ATLAS-CONF-2015-029, 2015, URL: `https://cds.cern.ch/record/2037702` (cit. on p. 7).

[70] ATLAS Collaboration, *ATLAS b-jet identification performance and efficiency measurement with $t\bar{t}$ events in pp collisions at $\sqrt{s} = 13$ TeV*, Eur. Phys. J. C **79** (2019) 970, arXiv: `1907.05120 [hep-ex]` (cit. on pp. 7, 17).

[71] ATLAS Collaboration, *Identification of Jets Containing b-Hadrons with Recurrent Neural Networks at the ATLAS Experiment*, ATL-PHYS-PUB-2017-003, 2017, URL: `https://cds.cern.ch/record/2255226` (cit. on p. 7).

[72] ATLAS Collaboration, *Calibration of light-flavour b-jet mistagging rates using ATLAS proton–proton collision data at $\sqrt{s} = 13$ TeV*, ATLAS-CONF-2018-006, 2018, URL: `https://cds.cern.ch/record/2314418` (cit. on p. 7).

[73] ATLAS Collaboration, *Measurement of the c-jet mistagging efficiency in $t\bar{t}$ events using pp collision data at $\sqrt{s} = 13$ TeV collected with the ATLAS detector*, Eur. Phys. J. C **82** (2022) 95, arXiv: `2109.10627 [hep-ex]` (cit. on p. 7).

[74] ATLAS Collaboration, *Muon reconstruction and identification efficiency in ATLAS using the full Run 2 pp collision data set at $\sqrt{s} = 13$ TeV*, Eur. Phys. J. C **81** (2021) 578, arXiv: `2012.00578 [hep-ex]` (cit. on p. 7).

[75] ATLAS Collaboration, *Evidence for the $H \rightarrow b\bar{b}$ decay with the ATLAS detector*, JHEP **12** (2017) 024, arXiv: `1708.03299 [hep-ex]` (cit. on p. 7).

[76] ATLAS Collaboration, *Muon reconstruction performance of the ATLAS detector in proton–proton collision data at $\sqrt{s} = 13$ TeV*, Eur. Phys. J. C **76** (2016) 292, arXiv: `1603.05598 [hep-ex]` (cit. on p. 7).

[77] G. V. Moustakides and K. Basioti, *Training Neural Networks for Likelihood/Density Ratio Estimation*, 2019, arXiv: `1911.00405 [eess.SP]` (cit. on p. 12).

[78] T. Kanamori, S. Hido, and M. Sugiyama, *A Least-squares Approach to Direct Importance Estimation*, Journal of Machine Learning Research **10** (2009) 1391, URL: `http://jmlr.org/papers/v10/kanamori09a.html` (cit. on p. 12).

[79] V. Nair and G. E. Hinton, "Rectified Linear Units Improve Restricted Boltzmann Machines," *Proceedings of the 27th International Conference on International Conference on Machine Learning*, ICML'10, Haifa, Israel: Omnipress, 2010 807, ISBN: 9781605589077, URL: `https://www.cs.toronto.edu/~fritz/absps/reluICML.pdf` (cit. on p. 12).

[80] B. Lakshminarayanan, A. Pritzel, and C. Blundell, *Simple and Scalable Predictive Uncertainty Estimation using Deep Ensembles*, 2017, arXiv: `1612.01474 [stat.ML]` (cit. on p. 14).

[81] B. Efron, *Bootstrap Methods: Another Look at the Jackknife*, Ann. Statist. **7** (1979) 1 (cit. on p. 14).

[82] ATLAS Collaboration, *Measurement of the Inelastic Proton–Proton Cross Section at $\sqrt{s} = 13$ TeV with the ATLAS Detector at the LHC*, Phys. Rev. Lett. **117** (2016) 182002, arXiv: `1606.02625 [hep-ex]` (cit. on p. 17).

[83] ATLAS Collaboration, *Luminosity determination in pp collisions at $\sqrt{s} = 13$ TeV using the ATLAS detector at the LHC*, ATLAS-CONF-2019-021, 2019, URL: `https://cds.cern.ch/record/2677054` (cit. on p. 18).

[84] G. Avoni et al., *The new LUCID-2 detector for luminosity measurement and monitoring in ATLAS*, JINST **13** (2018) P07017 (cit. on p. 18).

[85] J. Baglio et al., *$gg \to HH$: Combined uncertainties*, Phys. Rev. D **103** (2021) 056002, arXiv: `2008.11626 [hep-ph]` (cit. on pp. 18, 22).

[86] W. Verkerke and D. Kirkby, *The RooFit toolkit for data modeling*, 2003, arXiv: `physics/0306116 [physics.data-an]` (cit. on p. 19).

[87] G. Cowan, K. Cranmer, E. Gross, and O. Vitells, *Asymptotic formulae for likelihood-based tests of new physics*, Eur. Phys. J. C **71** (2011) 1554, arXiv: `1007.1727 [physics.data-an]` (cit. on p. 19), Erratum: Eur. Phys. J. C **73** (2013) 2501.

[88] A. L. Read, *Presentation of search results: the CL_S technique*, J. Phys. G **28** (2002) 2693 (cit. on p. 19).