ATLAS CONF Note

ATLAS-CONF-2023-071

25th November 2023

Search for the non-resonant production of Higgs boson pairs via gluon fusion and vector-boson fusion in the $b\bar{b}\tau^+\tau^-$ final state in proton-proton collisions at $\sqrt{s}$ = 13 TeV with the ATLAS detector

The ATLAS Collaboration

A search for the non-resonant production of Higgs boson pairs in the $HH \to b\bar{b}\tau^+\tau^-$ channel is performed using 140 fb^{-1} of proton-proton collisions at a centre-of-mass energy of 13 TeV recorded by the ATLAS detector at the CERN Large Hadron Collider. The analysis strategy is optimised to probe anomalous values of the Higgs boson self-coupling modifier κ_λ and of the quartic $HHVV$ ($V = W, Z$) coupling modifier κ_{2V}. No significant excess above the expected background from Standard Model processes is observed. An observed (expected) upper limit $\mu_{HH} < 5.9$ (3.1) is set at 95% confidence-level on the Higgs boson pair production cross-section normalised to its Standard Model prediction. The coupling modifiers are constrained to an observed (expected) 95% confidence interval of $-3.2 < \kappa_\lambda < 9.1$ ($-2.5 < \kappa_\lambda < 9.2$) and $-0.4 < \kappa_{2V} < 2.6$ ($-0.2 < \kappa_{2V} < 2.4$), assuming all other Higgs boson couplings are fixed to the Standard Model prediction. The results are also interpreted in the context of effective field theories.

ATLAS-CONF-2023-071
26 November 2023

Contents

1 Introduction

After the discovery of the Higgs boson (H) in 2012 [1, 2], the ATLAS [3] and CMS [4] Collaborations at the Large Hadron Collider (LHC) have been focusing on the measurement of its properties [5–11]. All features probed so far for this new particle are consistent with the Standard Model (SM) predictions [12–17] for a Higgs boson with an observed mass m_H near 125 GeV [18, 19].

In this extensive measurement programme, the properties of interactions involving multiple Higgs bosons remain to be verified. In the SM, these can be characterised by the trilinear and quartic self-couplings λ_{HHH} and λ_{HHHH}, which are both equal to the coefficient λ of the quartic term of the Higgs field potential $V(\phi) = \mu^2|\phi|^2 + \lambda|\phi|^4$. The quartic couplings g_{HHVV} ($V = W, Z$) characterise the interactions between two Higgs bosons and two W or Z bosons, and are related in the SM to the HWW and HZZ couplings g_{HVV} through the relation $g_{HHVV} = g_{HVV}/2v$, where v is the vacuum expectation value of the Higgs field. Significant deviations from the SM predictions for these couplings would provide a strong indication of beyond the Standard Model (BSM) physics [20].

The most sensitive test of Higgs boson self-interactions comes from processes of Higgs boson pair production (HH) such as gluon fusion (ggF) and vector-boson fusion (VBF) HH production. Measuring the cross-section of these processes offers a direct probe of the values of these couplings, through their scale factors with respect to the SM predictions: $\kappa_\lambda = \lambda_{HHH}/\lambda^{SM}_{HHH}$, affecting both ggF and VBF production, and $\kappa_V = g_{HVV}/g^{SM}_{HVV}$ and $\kappa_{2V} = g_{HHVV}/g^{SM}_{HHVV}$, which only impact VBF production.

The dominant mode for HH production is ggF, with a cross-section of $\sigma^{\text{SM}}_{\text{ggF}} = 31.05^{+2.1}_{-7.2}$ fb [21–28], calculated at next-to-next-to-leading order (NNLO) including finite top-quark-mass effects for $m_H = 125$ GeV at $\sqrt{s} = 13$ TeV, resulting from the destructive interference between the leading-order (LO) Feynman diagrams shown in Figure 1. The second most common HH production mechanism at the LHC is VBF, with

a total cross-section of $\sigma_{\text{VBF}}^{\text{SM}} = 1.73 \pm 0.04$ fb [29–33] calculated at next-to-next-to-next-to-leading order (N^{3}LO) for $m_H = 125$ GeV at $\sqrt{s} = 13$ TeV. The LO diagrams for this process are shown in Figure 2.

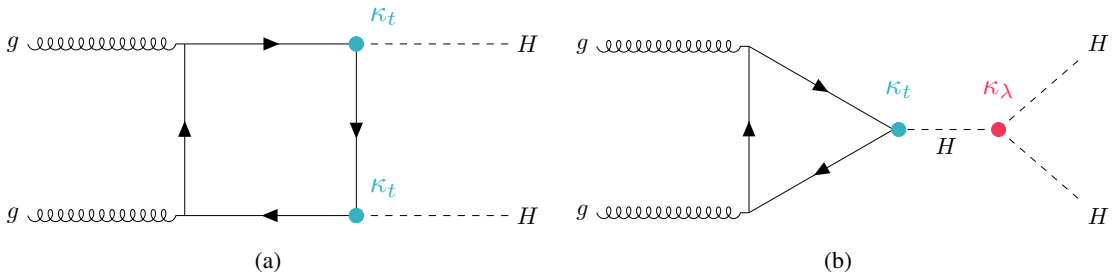


Figure 1: Leading-order Feynman diagrams of the ggF HH production process: (a) box and (b) triangle diagrams. The Higgs self-coupling modifier is shown as κ_λ, while the modifier for the coupling of the Higgs boson to the top quark is shown as κ_t.

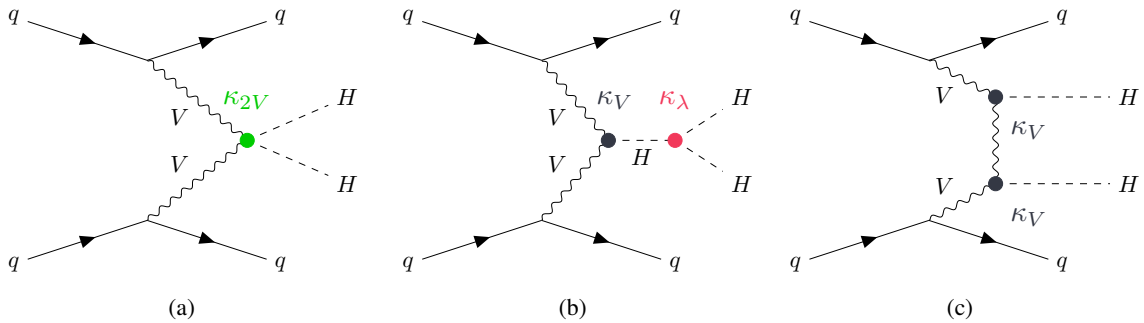


Figure 2: Leading-order Feynman diagrams of the VBF HH production process. The g_{HHVV} coupling modifier is shown as κ_{2V}, the Higgs self-coupling modifier as κ_λ, and the modifier for the coupling of the Higgs boson to the SM vector bosons V as κ_V.

The most precise measurement of SM HH production by the ATLAS Collaboration was obtained through a statistical combination [34] of the results in the $b\bar{b}\gamma\gamma$ [35], $b\bar{b}\tau^+\tau^-$ [36] and $b\bar{b}b\bar{b}$ [37] final states, exploiting the entire sample of proton-proton (pp) collisions provided by the LHC during its second phase of data-taking (Run 2, 2015-2018). The 95% confidence level (CL) observed (expected) upper limit on the HH production signal strength $\mu_{HH} = (\sigma_{\text{ggF}} + \sigma_{\text{VBF}})/(\sigma_{\text{ggF}}^{\text{SM}} + \sigma_{\text{VBF}}^{\text{SM}})$ from the combination is 2.4 (2.9) with respect to the case $\mu_{HH} = 0$, which is subsequently referred to as the *background-only hypothesis*. Using the values of the combination test statistics as a function of either κ_λ or κ_{2V}, when all other coupling modifiers are set to unity, the observed (expected with respect to the SM prediction) 95% confidence intervals (CIs) $\kappa_\lambda \in [-0.6, 6.6]$ ($[-2.1, 7.8]$) and $\kappa_{2V} \in [0.1, 2.0]$ ($[0.0, 2.1]$) are measured. The CMS Collaboration also performed a combination of HH analyses in multiple final states, based on their full Run 2 dataset [11]. The observed (expected) 95% CL upper limit on μ_{HH} is 2.5 (3.4). The observed allowed ranges of κ_λ and κ_{2V} are restricted to $\kappa_\lambda \in [-1.24, 6.49]$ and $\kappa_{2V} \in [0.67, 1.38]$. No excess over the background-only hypothesis is observed by any of these analyses.

With a branching ratio of 7.3%, the $b\bar{b}\tau^+\tau^-$ decay channel is able to provide a compromise between expected signal event yield and background contamination. This leads to a sensitivity similar to that of the $b\bar{b}b\bar{b}$ and $b\bar{b}\gamma\gamma$ decay modes. The latest ATLAS results for $HH \to b\bar{b}\tau^+\tau^-$ with the full LHC Run 2 dataset are documented in Ref. [36]. They result in an observed (expected) 95% CL upper

limit on the total HH production cross-section of 4.7 (3.9) times the SM prediction with respect to the background-only hypothesis. Moreover, observed (expected) 95% CIs on κ_λ and κ_{2V} are measured at $\kappa_\lambda \in [-2.7, 9.5]$ ($[-3.1, 10.2]$) and $\kappa_{2V} \in [-0.6, 2.7]$ ($[-0.5, 2.7]$), respectively [34]. Recent results from the CMS Collaboration [38] set an observed (expected) upper limit of 3.3 (5.2) times the SM production cross-section at 95% CL over the background-only hypothesis. Additionally, constraints are set on κ_λ and κ_{2V} with respect to the background-only hypothesis, leading to the observed (expected) 95% CL exclusion limits $\kappa_\lambda \in [-1.7, 8.7]$ ($[-2.9, 9.8]$) and $\kappa_{2V} \in [-0.4, 2.6]$ ($[-0.6, 2.8]$).

In this note, an updated search for non-resonant Higgs boson pair production in the $b\bar{b}\tau^+\tau^-$ final state with the full Run 2 ATLAS dataset is presented. Compared to the previous ATLAS result in the $b\bar{b}\tau^+\tau^-$ channel [36], the overall object identification, trigger strategy and event selection in the signal-enriched regions is unchanged, but an optimised classification of the selected events is implemented to enhance the sensitivity to κ_λ as well as to the VBF production mode. Improved multivariate classifiers are used to exploit the kinematic features of SM VBF HH production to define a dedicated VBF category, which improves the sensitivity to anomalous values of the coupling modifiers κ_λ and κ_{2V}. Updated Monte Carlo (MC) predictions are used for describing the main backgrounds of top-quark pair production ($t\bar{t}$) and Z boson production in association with heavy flavour quarks, leading to a more accurate modeling of these processes and enhancing the statistical power of the simulation. The event selection for the auxiliary measurement of the background from Z boson production in association with heavy flavour quarks is adapted to improve the consistency of the kinematic properties of this process with the signal-enriched regions. Finally, the obtained results are reinterpreted in the context of effective field theories (EFTs). These changes in analysis setup and strategy lead to an improvement of the expected 95% CL limit on μ_{HH} of 20%. The widths of the 95% CIs on κ_λ and κ_{2V} with respect to the SM expectation are reduced by 10% and 19%, respectively, compared to the analysis described in Ref. [36]. The results documented in this note are meant to supersede those from Ref. [36].

This note is organised as follows. Section 2 describes the ATLAS detector. The collected dataset and the simulated event samples are covered in Section 3. The event selection and categorisation, along with the description of the multivariate classifiers, are described in Section 4. Section 5 lists the systematic uncertainties. Finally, Section 6 summarises the results, Section 7 covers the conducted EFT interpretations, and the study is concluded in Section 8.

2 ATLAS detector

The ATLAS experiment at the LHC is a multipurpose particle detector with a forward–backward symmetric cylindrical geometry and a near 4π coverage in solid angle.[1] It consists of an inner tracking detector surrounded by a thin superconducting solenoid providing a 2 T axial magnetic field, electromagnetic and hadron calorimeters, and a muon spectrometer. The inner tracking detector covers the pseudorapidity range $|\eta| < 2.5$. It consists of silicon pixel, silicon microstrip, and transition radiation tracking detectors. Lead/liquid-argon (LAr) sampling calorimeters provide electromagnetic (EM) energy measurements with high granularity. A steel/scintillator-tile hadron calorimeter covers the central pseudorapidity range ($|\eta| < 1.7$). The endcap and forward regions are instrumented with LAr calorimeters for both the EM and

[1] ATLAS uses a right-handed coordinate system with its origin at the nominal interaction point (IP) in the centre of the detector and the z-axis along the beam pipe. The x-axis points from the IP to the centre of the LHC ring, and the y-axis points upwards. Polar 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 $\Delta R \equiv \sqrt{(\Delta\eta)^2 + (\Delta\phi)^2}$.

hadronic energy measurements up to $|\eta| = 4.9$. The muon spectrometer surrounds the calorimeters and is based on three large superconducting air-core toroidal magnets with eight coils each. The field integral of the toroids ranges between 2.0 and 6.0 T m across most of the detector. The muon spectrometer includes a system of precision tracking chambers and fast detectors for triggering. A two-level trigger system is used to select events. The first-level trigger is implemented in hardware and uses a subset of the detector information to accept events at a rate below 100 kHz. This is followed by a software-based trigger that reduces the accepted event rate to 1 kHz on average depending on the data-taking conditions. An extensive software suite [39] is used in simulation, in the reconstruction and analysis of real and simulated events, in detector operations, and in the trigger and data acquisition systems of the experiment.

3 Data and simulated event samples

The measurements presented in this note use the pp collision dataset collected by the ATLAS experiment during the LHC Run 2 at a centre-of-mass energy of $\sqrt{s} = 13$ TeV. After data quality requirements [40], the integrated luminosity of the dataset is 140.1 ± 1.2 fb^{-1} [41].

The samples of simulated events used for this study are summarised in Table 1, and they correspond to those already employed in the previous publication [36], with a few important updates aimed at improving the statistical power of the simulated samples for the main background processes providing a more accurate simulation of the HH signal. Dedicated VBF HH samples are produced for additional κ_{2V} and κ_V variations, while additional ggF HH samples are employed to model κ_λ variations.

The production of $t\bar{t}$ events, and of single-top-quark events in the $Wt-$, $s-$ and $t-$channels is simulated using the Powheg Box v2 [42] generator together with the NNPDF3.0nlo parton distribution functions (PDF) set [43]. The simulated events are interfaced to Pythia 8.230 [44] for parton showering and hadronisation using the A14 tune [45, 46] together with the NNPDF2.3lo PDF set. The setup of the MC event generator for these production processes is unchanged with respect to Ref. [36]. For $t\bar{t}$ processes the size of the sample is increased by including two statistically independent sets of events. The first sample is simulated with the requirement of at least one top quark decaying to a leptonic final state, while in the second sample both top quarks are forced to decay leptonically. The combination of these samples results in a decrease of the statistical uncertainty related to the $t\bar{t}$ simulation approximately by a factor of 2 in the selected phase space.

For events containing a W or a Z boson produced in association with jets, new samples are produced with the Sherpa 2.2.11 generator [47]. Events are generated using NLO matrix elements for up to two partons, and LO matrix elements for up to five partons calculated using the OpenLoops [48–51] and Comix [52] matrix-element generators. For these samples, the NNPDF3.0nnlo PDF set is used with dedicated parton shower, matched to the matrix element via the MEPS@NLO prescription [53]. The electroweak input scheme is updated compared to the previous version, along with corrected heavy-flavour hadron production fractions. The Sherpa 2.2.11 generator allows faster per-event generation time, with significant reduction of the negative weight fraction, resulting in an improvement of the MC statistical uncertainty ranging from 30% to 60%, depending on the flavour composition of the events and the analysis region.

Diboson (WW, WZ and ZZ) and $t\bar{t}Z$ production processes are simulated with the Sherpa 2.2.1 generator, whereas the $t\bar{t}W$ production process is simulated with the Sherpa 2.2.8 generator. These samples use the NNPDF3.0nnlo PDF set with dedicated parton shower tuning developed by the Sherpa authors. Single Higgs boson production is considered as part of the background in this search, and its production modes

are simulated using the POWHEG BOX v2 generator and either the PDF4LHC15NLO with the AZNLO [54] tune or the NNPDF3.0NLO PDF sets with the A14 tune. The setup of the MC event generator for these production processes is unchanged with respect to Ref. [36].

Signal samples consist of simulated events from non-resonant ggF and VBF production of Higgs boson pairs, with one Higgs boson decaying to $b\bar{b}$ and the other one to $\tau^+\tau^-$. The simulated ggF events are generated with the POWHEG BOX v2 generator at NLO with finite top-quark mass [55], and using the PDF4LHC15NLO PDF set. The VBF events are generated at LO using the MADGRAPH5_AMC@NLO 2.7.3 [56] generator with the NNPDF3.0NLO PDF set. Parton shower and hadronisation are simulated using PYTHIA 8.244 with the A14 tune and the NNPDF2.3LO PDF set. In addition to the SM case of $\kappa_\lambda = 1$, a ggF HH sample is generated with the same settings as the nominal sample but with the non-SM value of the self-coupling modifier $\kappa_\lambda = 10$, and then passed through detector simulation and reconstruction algorithms. A reweighting technique based on the particle-level invariant mass m_{HH} of the Higgs boson pair is applied to the $\kappa_\lambda = 1$ sample to determine the ggF HH signal yield and kinematic distributions for any value of κ_λ [57]. The particle-level m_{HH} spectrum for any generic value of κ_λ is calculated from the m_{HH} distributions of three ggF HH samples generated for $\kappa_\lambda = 0$, 1 or 20. In order to determine the potential non-closure in the reweighting process from residual kinematic effects, the procedure is validated by comparing the predicted event yields and kinematic distributions of the simulated sample generated with $\kappa_\lambda = 1$ and reweighted to $\kappa_\lambda = 10$ with those of the simulated sample generated under the hypothesis $\kappa_\lambda = 10$. Furthermore, 12 additional VBF HH samples are generated and simulated with the same setup as the nominal VBF sample, but using non-SM combinations of the coupling modifiers κ_λ, κ_{2V} and κ_V. A basis for a linear combination is formed by the SM sample and five of the other 12 samples, corresponding to the combinations of the κ_λ, κ_{2V} and κ_V couplings (1, 1.5, 1), (0, 1, 1), (10, 1, 1), (1, 3, 1), (-5, 1, 0.5). This approach is used to determine the expected VBF HH yields and distributions for any value of κ_λ, κ_{2V} and κ_V. The remaining seven samples are compared to the corresponding predictions from the interpolation procedure for validation purposes. The same procedure was used in the measurements presented in Ref. [34]. For both ggF and VBF production modes good closure between the simulated and reweighted samples for alternative κ_λ, κ_V and κ_{2V} values is observed within statistical uncertainties.

All generated samples are passed through a detailed simulation of the ATLAS detector response [58] based on GEANT4 [59]. The Higgs boson mass is assumed to be 125 GeV in both the simulation and the analysis of the data. All samples are normalised to the same cross-section calculations detailed in Ref [36]. The impacts of the differences with respect to the best-fit values of the m_H measurements reported in Refs. [18, 19], as well as the effects of the corresponding experimental uncertainties on m_H, are negligible.

4 Event selection and categorisation

Events are selected in three separate signal regions (SRs), which remain unchanged with respect to Ref. [36]. The $\tau_{\text{had}}\tau_{\text{had}}$ signal region (SR) targets fully-hadronic decay modes of the τ-lepton pair, where the presence of two oppositely charged hadronically decaying τ leptons (τ_{had}) is determined by detector signatures compatible with the expected visible decay products ($\tau_{\text{had-vis}}$). Two $\tau_{\text{lep}}\tau_{\text{had}}$ SRs target events with a leptonic decay of a τ lepton (τ_{lep}) into an electron or a muon, and an oppositely charged $\tau_{\text{had-vis}}$. The decay channel with both τ leptons decaying leptonically is not studied in this report but is instead covered in Ref. [61]. A control region (CR) is defined to constrain the background from Z bosons produced in association with two jets initiated by b or c quarks (referred to as Z + HF in the following), and top-quark pair production processes. In all regions the presence of two b-jets is also required. These four regions are

Table 1: Summary of the nominal HH signal as well as the background event samples used in this note. The generator used in the simulation, the parton distribution function (PDF) set, the parton shower (PS) and the set of tuned parameters (tune) are also provided. More details are given in the text and in Ref. [36].

Process	Matrix element generator	PDF set	Showering	Tune
Signal				
$gg \to HH$ (ggF)	Powheg Box v2 [42]	PDF4LHC15nlo [60]	Pythia 8.244 [44]	A14 [45, 46]
$qq \to qqHH$ (VBF)	MadGraph5_aMC@NLO 2.7.3 [56]	NNPDF3.0nlo [43]	Pythia 8.244	A14
Top-quark				
$t\bar{t}$	Powheg Box v2	NNPDF3.0nlo	Pythia 8.230	A14
t-channel	Powheg Box v2	NNPDF3.0nlo	Pythia 8.230	A14
s-channel	Powheg Box v2	NNPDF3.0nlo	Pythia 8.230	A14
Wt	Powheg Box v2	NNPDF3.0nlo	Pythia 8.230	A14
$t\bar{t}Z$	Sherpa 2.2.1 [47]	NNPDF3.0nnlo	Sherpa 2.2.1	-
$t\bar{t}W$	Sherpa 2.2.8	NNPDF3.0nnlo	Sherpa 2.2.8	-
Vector boson + jets				
W/Z+jets	Sherpa 2.2.11	NNPDF3.0nnlo	Sherpa 2.2.11	-
Diboson				
WW, WZ, ZZ	Sherpa 2.2.1	NNPDF3.0nnlo	Sherpa 2.2.1	-
Single Higgs boson				
ggF	Powheg Box v2	PDF4LHC15nlo	Pythia 8.212	AZNLO [54]
VBF	Powheg Box v2	PDF4LHC15nlo	Pythia 8.212	AZNLO
$qq \to WH$	Powheg Box v2	PDF4LHC15nlo	Pythia 8.212	AZNLO
$qq \to ZH$	Powheg Box v2	PDF4LHC15nlo	Pythia 8.212	AZNLO
$gg \to ZH$	Powheg Box v2	PDF4LHC15nlo	Pythia 8.212	AZNLO
$t\bar{t}H$	Powheg Box v2	NNPDF3.0nlo	Pythia 8.230	A14

briefly summarised in Section 4.1. Selected events are split in different categories to enhance the sensitivity to the coupling modifiers κ_λ and κ_{2V}, as described in Section 4.2. In each category a multivariate approach based on boosted decision trees (BDTs) is adopted to build the final discriminants, as detailed in Section 4.3. The identification and reconstruction of electrons, muons, $\tau_{\text{had-vis}}$, jets from the hadronisation of quarks and gluons, b-tagged jets, and missing transverse momentum ($p_{\text{T}}^{\text{miss}}$) is identical to what was documented in Ref. [36].

4.1 Preselection

Events in the $\tau_{\text{had}}\tau_{\text{had}}$ SR are selected using a combination of single-$\tau_{\text{had-vis}}$ triggers (STTs) and di-$\tau_{\text{had-vis}}$ triggers (DTTs), and are required to have two $\tau_{\text{had-vis}}$ with opposite charge. An electron and muon veto is applied to ensure orthogonality with the $\tau_{\text{lep}}\tau_{\text{had}}$ SRs. In the $\tau_{\text{had}}\tau_{\text{had}}$ SR event selection, the offline[2] p_{T} thresholds for the $\tau_{\text{had-vis}}$ range between 100 GeV and 180 GeV for STT events depending on the data-taking period, while they are set at 40 GeV (30 GeV) for the (sub-)leading $\tau_{\text{had-vis}}$ for DTT events. For events selected by the STTs a second $\tau_{\text{had-vis}}$-candidate is required, with an offline p_{T} threshold of 25 GeV. Additional offline requirements for the DTTs are that either one extra jet with an offline p_{T} threshold set to 80 GeV is present in the event, and the $\tau_{\text{had-vis}}$ are reconstructed within $\Delta R = 2.5$ of each other, or that two extra jets with offline p_{T} thresholds set to 45 GeV are present in the event. For events that pass both the STTs and DTTs, the offline requirements used for the STTs are applied.

[2] In this note, offline objects are objects which are reconstructed after the data were collected, as opposed to trigger-level objects.

Events containing exactly an electron or muon and one $\tau_{\text{had-vis}}$ with opposite charge are split into two mutually exclusive SRs, depending on whether they pass a single-lepton trigger (SLT) or a lepton-plus-$\tau_{\text{had-vis}}$ trigger (LTT), named the $\tau_{\text{lep}}\tau_{\text{had}}$ SLT SR and the $\tau_{\text{lep}}\tau_{\text{had}}$ LTT SR respectively. Only events failing the $\tau_{\text{lep}}\tau_{\text{had}}$ SLT SR selection are considered for the $\tau_{\text{lep}}\tau_{\text{had}}$ LTT SR. Depending on the data-taking period, the offline electron (muon) selected by the SLT is required to have $p_{\text{T}}^{e} > 25$ GeV or $p_{\text{T}}^{e} > 27$ GeV ($p_{\text{T}}^{\mu} > 21$ GeV or $p_{\text{T}}^{\mu} > 27$ GeV). Events selected in the $\tau_{\text{lep}}\tau_{\text{had}}$ LTT SR are required to contain either an electron or a muon with offline p_{T} thresholds set to $p_{\text{T}}^{e} > 18$ GeV and $p_{\text{T}}^{\mu} > 15$ GeV respectively, along with a $\tau_{\text{had-vis}}$ with an with offline p_{T} threshold set to 30 GeV.

Events in all SRs are required to have $m_{\tau\tau}^{\text{MMC}} > 60$ GeV.[3] In order to target $H \to b\bar{b}$ decays, events are required to contain exactly two b-tagged jets in the pseudorapidity region of $|\eta| < 2.5$, passing the criteria of the 'DL1r' b-tagging algorithm with a nominal efficiency of 77% for b jets [63]. The two selected b-tagged jets have to pass minimum p_{T} thresholds of 45 and 20 GeV respectively, in addition to any trigger-dependent requirements. In the $\tau_{\text{lep}}\tau_{\text{had}}$ SRs the invariant mass of the b-tagged jet pair (m_{bb}) is required to be lower than 150 GeV to reject background $t\bar{t}$ events, and a $\tau_{\text{had-vis}}$ with transverse momentum (p_{T}) of at least 20 GeV is required in the $\tau_{\text{lep}}\tau_{\text{had}}$ SLT SR, while $p_{\text{T}} > 30$ GeV is required in the $\tau_{\text{lep}}\tau_{\text{had}}$ LTT SR, in addition to any trigger-dependent requirements.

Events in the CR are required to contain exactly two electrons or two muons of opposite charge with a dilepton invariant mass ($m_{\ell\ell}$) range of 75 GeV $< m_{\ell\ell} <$ 110 GeV, and exactly two b-tagged jets. The m_{bb} is required to be less than 40 GeV or greater than 210 GeV to avoid overlap with other analyses targeting $H \to b\bar{b}$ decays. Compared to Ref. [36], the cut on the transverse momentum of the selected leptons is raised to $p_{\text{T}} > 40$ GeV (from $p_{\text{T}} > 9$ GeV), while the leading b-tagged jet is required to have $p_{\text{T}} > 45$ GeV. This selection provides a closer alignment between the kinematic properties of events selected in the CR and the SRs. Figure 3 shows the predicted and observed $m_{\ell\ell}$ distributions in the CR, after the likelihood fit described in Section 6.

The main sources of background in the SRs after this preselection are from top-quark, Z+jets, W+jets, diboson, single Higgs boson and multi-jet production. The normalisations of simulated $t\bar{t}$ and Z + HF backgrounds are determined from data in the likelihood fits of signal and control regions described in Section 6. Depending on the source, the background contamination is estimated using data-driven or simulation-based techniques, or a combination of both. A reconstructed $\tau_{\text{had-vis}}$ candidate in these background events can originate either from a τ_{had} decay (true-$\tau_{\text{had-vis}}$), or from a jet, an electron or a muon misidentified as a $\tau_{\text{had-vis}}$ candidate (fake-$\tau_{\text{had-vis}}$). Simulated event samples are used to model background events containing true-$\tau_{\text{had-vis}}$ and events with an electron or a muon misidentified as a $\tau_{\text{had-vis}}$ candidate. The processes that contribute most to background events with fake-$\tau_{\text{had-vis}}$ candidate are $t\bar{t}$ and multi-jet production. In $t\bar{t}$ events, fake-$\tau_{\text{had-vis}}$ candidate typically originate from quark-initiated jets from the top-quark decay. In multi-jet events, both quark- and gluon-initiated jets are a source for fake-$\tau_{\text{had-vis}}$ candidates. Events with fake-$\tau_{\text{had-vis}}$ candidate in $t\bar{t}$ and multi-jet production are estimated from techniques relying on both simulated events and data, described in the Ref. [36], which prove to provide an accurate modelling of the variables used for the event categorisation and multivariate techniques described in this Section. The estimate of this background relies on a good description of the fundamental properties of τ leptons in the three analysis SRs, which are unchanged compared to Ref. [36]. The modelling of events with fake-$\tau_{\text{had-vis}}$ candidate from background processes other than $t\bar{t}$ and multi-jet production is performed using MC simulation, as they represent a minor contribution to the total background. The changes introduced to the MC simulation detailed in Section 3 are found to have a negligible impact on the data-driven estimate of the background.

[3] The invariant mass of the τ-lepton pair ($m_{\tau\tau}^{\text{MMC}}$) is estimated using the Missing Mass Calculator (MMC) [62].

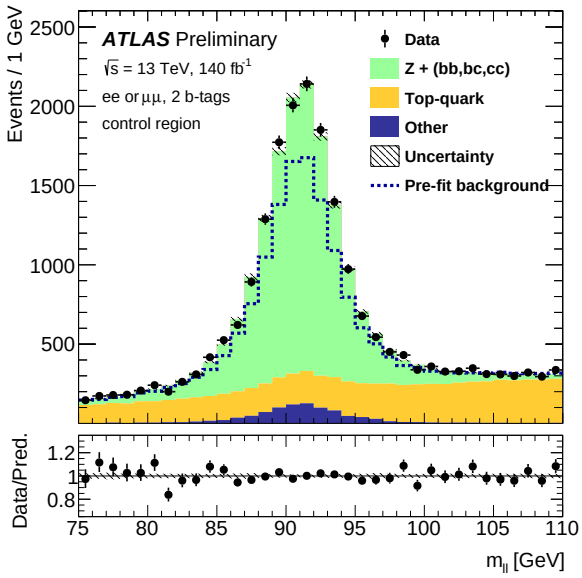


Figure 3: Predicted and observed $m_{\ell\ell}$ distributions in the CR after the fit described in Section 6. The normalisation and shape of the backgrounds and the uncertainty in the total background are shown as determined from the likelihood fit to data described in Section 6. The background processes named "Other" contain contributions from Z-boson production in association with less than two jets initiated by b or c quarks, W-boson production, vector boson pair production, $t\bar{t}W$ and $t\bar{t}Z$ production and single Higgs production processes. The uncertainty band includes the statistical and systematic uncertainties of the total background. The dashed histogram shows the total pre-fit background. The lower panels show the ratio of data to the total post-fit sum of signal and background, where the hatched bands show the statistical and systematic uncertainties of the sum of signal and background contributions.

4.2 Event categorisation

Events selected in each SR described in Section 4.1 are split into three separate categories. To enhance the sensitivity to the coupling modifier κ_{2V}, a dedicated VBF category is defined with a multivariate approach, defining a dedicated BDT to select events with characteristic features of VBF HH production, separately for each analysis SR. These are referred to as *categorisation BDTs* in the following. The distribution of the invariant mass of the HH system (m_{HH}) in ggF HH events is significantly affected by the value of κ_λ. Hence, events not falling in the VBF category are split in two m_{HH} categories, targeting ggF HH production with κ_λ values close to the SM expectation (ggF high-m_{HH}) or significantly different from it (ggF low-m_{HH}). The three categories are mutually exclusive, and they are defined separately for each SR following the procedure outlined in Figure 4, leading to a total of nine event categories. First, VBF candidate events are identified by requiring the presence of at least two jets in addition to the ones associated with the $H \to b\bar{b}$ decay. These events are used to train the categorisation BDTs to separate ggF and VBF HH production modes in event topologies with additional jets. Events passing a SR-dependent selection on the output of these BDTs are assigned to the VBF category. Events failing the VBF selection are assigned to the ggF categories, together with events not selected as VBF candidates.

The categorisation BDTs are built using the Toolkit for Multivariate Data Analysis (TMVA [64]) using ggF HH events as signal and VBF HH events as background, normalised to their respective SM cross-section. A dedicated version of the BDT is trained in each SR respectively. In order to make use of the complete set of simulated events for the BDT training, optimisation and evaluation, the events are split into three

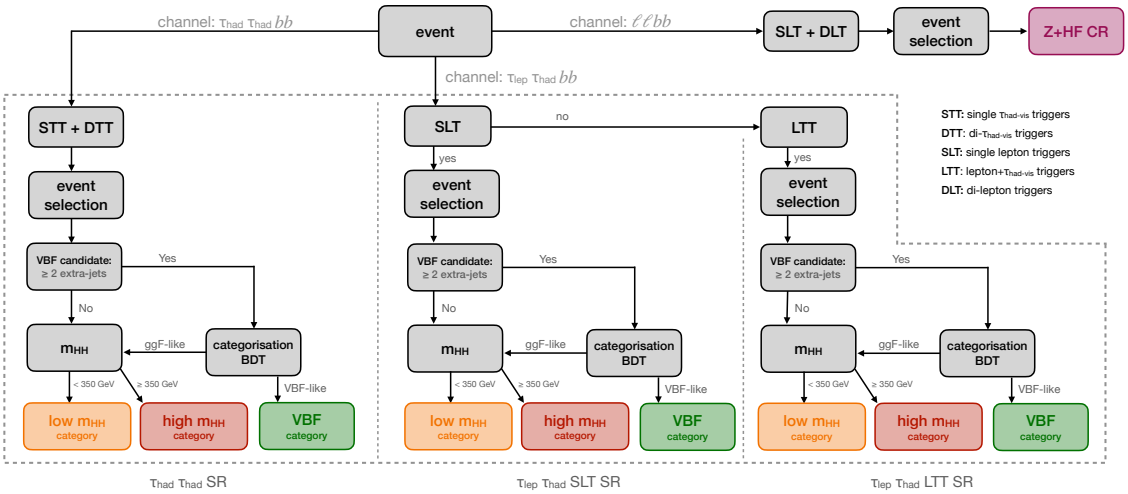


Figure 4: A flowchart summarizing the definition of the $\tau_{\text{had}}\tau_{\text{had}}$ SR, the $\tau_{\text{lep}}\tau_{\text{had}}$ SLT SR, the $\tau_{\text{lep}}\tau_{\text{had}}$ LTT SR and the dedicated CR defined in Section 4. The flowchart shows the selection criteria applied to define the corresponding ggF low-m_{HH}, ggF high-m_{HH} and VBF categories for each analysis SR, leading to a total number of nine analysis categories and an additional CR.

samples of equal size, A, B and C. The performance of the BDTs trained on sample A (respectively B, C), is optimised using sample B (C, A) and evaluated with sample C (A, B), ensuring an unbiased estimate of the performance of the BDTs. The selected data events are split in three samples, each analysed with one of the three separately trained BDTs. The output distributions of the BDTs evaluated on samples A, B and C are merged for both simulated and data events to produce the final discriminant. The number of trees and their depth are chosen to maximise the BDT separation power, quantified by the value of the number-counting significance z [65] computed from the binned distribution of the BDT discriminant.

The minimal set of training variable that optimises the BDT separation power is determined starting from a small set and iteratively adding variables one at a time from a pre-defined list of candidate variables. The variable leading to the largest increase (or minimal decrease[4]) in significance z is included, until no changes are observed. The starting set of variables includes the invariant mass of the VBF jets, defined as the two jets with the highest p_{T} not associated with the $H \to b\bar{b}$ decay, and their pseudorapidity gap ($\Delta\eta_{jj}^{\text{VBF}}$). The final set of variables for the categorisation BDTs in each SR is summarised in Table 2. It includes the product of VBF jet pseudorapidities, their angular separations ($\Delta\phi_{jj}^{\text{VBF}}$ and $\Delta R_{jj}^{\text{VBF}}$) and m_{HH}. In addition, the Fox-Wolfram moments f_i of i-th order [66] as well as their modified definitions for usage in hadron collider experiments h_i [67] are able to further increase the separation power, together with the centrality[5] (C) and the invariant mass (m_{Eff}) of the system composed by the selected τ leptons, the missing transverse momentum vector $\vec{p}_{\text{T}}^{\text{miss}}$ and the selected jets. The predicted and observed distributions of the resulting BDT scores are shown in Figure 5 for all three SRs. VBF candidate events are assigned to the VBF category if their BDT score is evaluated below a certain threshold. The value of this threshold is optimised

[4] Variables showing minimal decrease of the z significance are retained to mitigate the impact of statistical fluctuations on the optimisation, potentially leading to prematurely terminating the iterative procedure. In the final selection only variables improving the BDT sensitivity are retained.

[5] The centrality of a set of four-momenta of index i is defined as $C = \dfrac{\sum_i p_{\text{T}}(i)}{\sum_i E(i)}$

to achieve the best upper limits on HH production for ggF and VBF production modes separately and combined, along with the best exclusion limits for the coupling modifiers κ_λ and κ_{2V} from the likelihood fit described in Section 6. The categorisation BDT cut values are set to 0.1, −0.13 and −0.1 for the $\tau_{\text{had}}\tau_{\text{had}}$, $\tau_{\text{lep}}\tau_{\text{had}}$ SLT and $\tau_{\text{lep}}\tau_{\text{had}}$ LTT SRs, respectively.

Table 2: Input variables for the categorisation BDTs in each of the three SRs. The superscripts a and c specify the selection of jets that are taken into account for the calculation in addition to the two τ-lepton candidates and the $p_{\text{T}}^{\text{miss}}$ vector. For variables with a c, only the four-momenta of central jets, i.e. jets with $|\eta| < 2.5$, are included, while an a indicates that all available jets are included.

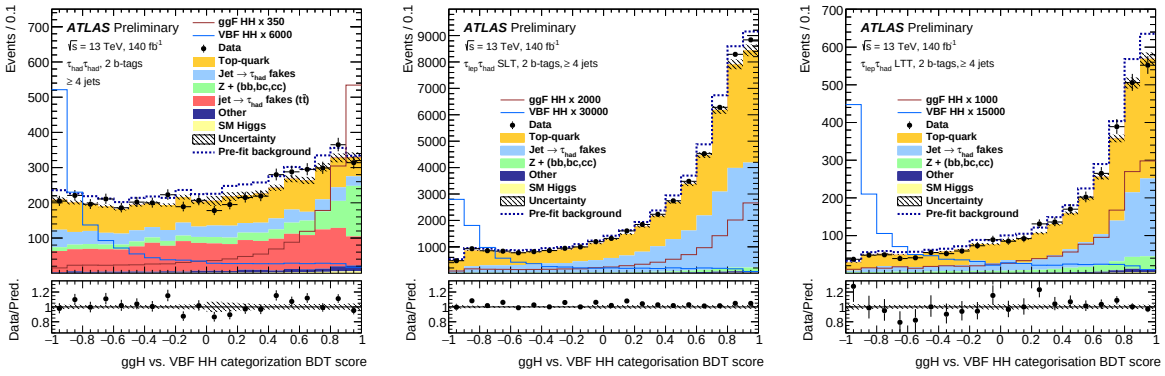
Variable	$\tau_{\text{had}}\tau_{\text{had}}$	$\tau_{\text{lep}}\tau_{\text{had}}$ SLT	$\tau_{\text{lep}}\tau_{\text{had}}$ LTT
m_{jj}^{VBF}	✓	✓	✓
$\Delta\eta_{jj}^{\text{VBF}}$	✓	✓	✓
VBF $\eta_0 \times \eta_1$	✓	✓	
$\Delta\phi_{jj}^{\text{VBF}}$	✓		
$\Delta R_{jj}^{\text{VBF}}$		✓	✓
$\Delta R_{\tau\tau}$	✓		
m_{HH}	✓		
f_2^a	✓		
C^a		✓	✓
m_{Eff}^a		✓	✓
f_0^c		✓	
f_0^a			✓
h_3^a			✓

Events not retained in the VBF category are split into low-m_{HH} and high-m_{HH} categories targeting ggF HH production. While the ggF HH cross-section increases for κ_λ values larger than the SM expectation and close to the current experimental sensitivity ($\kappa_\lambda \sim 6$ [34]), the softer m_{HH} spectrum leads to smaller detector acceptance and selection efficiency. The event split in different regions of m_{HH} allows to partially disentangle these effects, improving the sensitivity to higher κ_λ values. The splitting value is chosen to be 350 GeV since the effect of the interference between the box and triangle diagrams on the differential ggF HH production cross-section is maximal there.

4.3 Multivariate discriminants

Based on the event categorisation described in Section 4.2, an additional set of multivariate discriminants making use of BDTs is trained and evaluated in each of the analysis SRs separately to achieve optimal separation between the HH signal and the background. For each SR, a dedicated version of the BDT is constructed for each analysis category separately, leading to nine different BDTs[6]. In the ggF high-m_{HH} and

[6] One dedicated BDT is trained for each analysis SR ($\tau_{\text{had}}\tau_{\text{had}}$, $\tau_{\text{lep}}\tau_{\text{had}}$ SLT, $\tau_{\text{lep}}\tau_{\text{had}}$ LTT) and each category (ggF low-m_{HH}, ggF high-m_{HH} and VBF), leading to 9 separate BDTs.



(a) (b) (c)

Figure 5: Predicted and observed distributions of the categorisation BDT scores in the $\tau_{\text{had}}\tau_{\text{had}}$ (a), $\tau_{\text{lep}}\tau_{\text{had}}$ SLT (b) and $\tau_{\text{lep}}\tau_{\text{had}}$ LTT (c) SRs, for all candidate VBF events. The background distributions are shown at post-fit level as obtained from the combined likelihood fit to data described in Section 6. The background processes named "Other" contain contributions from Z-boson production in association with less than two jets initiated by b or c quarks, W-boson production, vector boson pair production, $t\bar{t}W$ and $t\bar{t}Z$ production. The ggF and VBF HH signal distributions are overlaid and scaled to the factor indicated in the legend times the SM expectation. The hatched band indicates the combined statistical and systematic uncertainties of this prediction. The dashed histogram shows the total pre-fit background. The lower panels show the ratio of data to the total post-fit sum of signal and background, where the hatched bands show the statistical and systematic uncertainties of this prediction.

VBF categories, the signal corresponds to the ggF and VBF SM production of HH pairs, respectively. In the ggF low-m_{HH} category the signal is defined as ggF HH production with $\kappa_\lambda = 10$. During training, the sum of all backgrounds normalised to their respective cross-sections is used. The backgrounds containing one or more $\tau_{\text{had-vis}}$ from misidentified quark- or gluon-initiated jets are modelled using simulation, except for the multi-jet background in the $\tau_{\text{had}}\tau_{\text{had}}$ category, where the data-driven estimate introduced in Section 4.1 is used. The BDTs evaluated in the low-m_{HH} and high-m_{HH} categories are trained on the event samples selected in the respective categories. The BDTs evaluated in the VBF categories are trained on events from both VBF and ggF categories to maximise the available sample size.

In order to make use of the complete set of simulated events for the BDT training, optimisation and evaluation, the events are split into three samples, following the same procedure described in Section 4.2. The number of trees and their depth are chosen to maximise the BDT separation power, quantified by the value of the signal significance z computed from the binned distribution of the BDT output.

The input variables used for the BDTs are chosen in order to maximise the BDT separation power, separately for each trained BDT. A list of variables is built for each BDT, ordered according to their impact on the signal significance z, following the same procedure as described in Section 4.2. In each analysis SR and category the starting set includes the following variables: the invariant mass of the two selected b-jets (m_{bb}), the invariant mass of the τ-lepton pair ($m_{\tau\tau}^{\text{MMC}}$), the m_{HH} reconstructed from the selected b-jet and τ-lepton pairs, the angular separation between b-jets (ΔR_{bb}) and between τ-leptons ($\Delta R_{\tau\tau}$). In the $\tau_{\text{lep}}\tau_{\text{had}}$ LTT high-m_{HH} category ΔR_{bb} is not included, and in the $\tau_{\text{lep}}\tau_{\text{had}}$ LTT VBF category both ΔR_{bb} and $\Delta R_{\tau\tau}$ are removed. Additional variables included in the BDT training through a z-based optimisation as described in Section 4.2 can be grouped in several categories. Variables that require the presence of a charged lepton are not included for the $\tau_{\text{had}}\tau_{\text{had}}$ SR. Higgs boson candidates H are reconstructed from either b-jet or τ-lepton pairs. For the low-m_{HH} and high-m_{HH} categories, variables describing the kinematic

properties of the selected b-jets and τ-leptons are included: the transverse momentum of the leading and subleading b-jets and τ leptons, along with the pseudorapidity of the τ leptons and the transverse mass of the each τ lepton and the $\vec{p}_{\mathrm{T}}^{\,\mathrm{miss}}$. The angular separation between the (sub)leading b-jet and the (sub)leading τ lepton is included, along with the angular separation between the leading b-jet and the subleading τ lepton, the pseudorapidity separation and the transverse momentum difference between the selected τ-lepton candidate and the charged lepton. Variables related to the reconstructed H candidate topologies include: the azimuthal angular separation between the b-jet pair and the τ-lepton pair, and between either the b-jet or the τ-lepton pair and the magnitude $E_{\mathrm{T}}^{\mathrm{miss}}$ of $\vec{p}_{\mathrm{T}}^{\,\mathrm{miss}}$, the azimuthal angle of the selected b-jet with respect to the H candidate rest frame, along with the transverse momentum of the b-jet and τ-lepton pairs. Variables describing the hadronic activity in candidate events, such as the number of selected hadronic jets, the scalar sum of jet transverse momenta (H_{T}), the effective mass of the τ-lepton pair and all jets with $|\eta| < 2.5$, the $E_{\mathrm{T}}^{\mathrm{miss}}$, the stransverse mass [68] (M_{T2}), the azimuthal angular separation between the selected charged lepton and the $\vec{p}_{\mathrm{T}}^{\,\mathrm{miss}}$. Additional variables are included to characterise event properties, such as the transverse mass of the W boson candidate in the $\tau_{\mathrm{lep}}\tau_{\mathrm{had}}$ SRs (defined as the transverse mass of the selected lepton and the $\vec{p}_{\mathrm{T}}^{\,\mathrm{miss}}$), the topness variable [69], the reduced invariant mass of the HH system (defined as $m_{HH}^{*} = m_{HH} - m_{bb} - m_{\tau\tau}^{\mathrm{MMC}} + 250$ GeV), the scaled invariant mass of the HH system (defined by scaling the four-momenta of both H candidates by the ratio $125\ \mathrm{GeV}/m_H$, where m_H is the H candidate invariant mass), the transverse momentum of the reconstructed HH system and the effective mass of the HH decay products. Dedicated variables describing characteristic event configurations are included via a selection of Fox-Wolfram moments (as defined in Section 4.2), circularity, sphericity and planar flow variables reconstructed from the HH decay products. Finally, the b-tagging information is provided by the quantile distribution of the DL1r tagger output for the selected b-jets, which is included as a training variable. For the VBF categories the first five variables listed in Table 2 are also included, along with circularity, sphericity and planar flow variables reconstructed from the τ-lepton pair and all selected jets, to target the specific features of VBF HH events.

5 Systematic uncertainties

While the largest source of uncertainty in this analysis comes from the limited amount of data, systematic uncertainties can affect the signal and background estimates. Uncertainties originating from the detector response in the selection and reconstruction of the objects used in this analysis are included as documented in Ref. [36]. Statistical uncertainties in the predicted background processes are modelled using a simplified version of the Beeston-Barlow technique [70], in which only the uncertainty in the total background contribution in each bin is considered. An uncertainty on the full Run 2 integrated luminosity of 0.83% [41], from measurements using the LUCID-2 [71] detector, is assigned to physics processes whose normalisations are taken from simulation.

For all processes whose estimation relies on MC simulation, the impact of various sources of theoretical uncertainties on their cross-section, on the fractional contribution to each analysis category within each SR, as well as on the shape of the BDTs introduced in Section 4.3 is considered. The total normalisation of $t\bar{t}$ and Z + HF backgrounds is determined via the likelihood fit described in Section 6, therefore no uncertainty is included for their cross-section calculation. For $t\bar{t}$ and Z + HF backgrounds, the uncertainty on the fractional contribution in each analysis category is computed as the relative variation in acceptance with respect to the dedicated CR introduced in Section 4, while for other processes this is evaluated as the absolute acceptance variation in each analysis category. To assess an uncertainty on the shape of the BDT output, a dedicated rebinning algorithm is applied to ensure that only statistically significant shape

variations are considered. In the signal-like region of the BDT distribution, the statistical uncertainty of each bin is required to be below a process- and variation-dependent threshold (ranging from 5% to 15%), while in the background-like region of the distribution the fraction of total events per bin is required to be larger than 5%. For each process, the impact of each source of uncertainty on the fractional contribution in each analysis category is correlated with the uncertainty on the shape of the BDT score in the corresponding category, in the likelihood fit described in Section 6. All sources of uncertainty are evaluated separately and correlated across the three analysis SRs. The uncertainties on the SM cross-section calculations for all processes are unchanged with respect to Ref. [36], along with the full uncertainty estimate for smaller background processes including Z+light-flavour jets, W+jets, diboson and single-top-quark production in the $s-$ and $t-$channels.

The $t\bar{t}$ hard-scatter and parton-shower uncertainties are evaluated by comparing the nominal sample with samples generated using MadGraph5_aMC@NLO +Pythia 8 and Powheg Box v2 +Herwig 7, respectively. The hard-scatter uncertainty also accounts for uncertainties in the matching and merging of the matrix-element calculation with the parton-shower algorithm. The uncertainty on missing higher-order QCD corrections and the modelling of initial state QCD radiation is assessed via independent variations of the renormalisation and factorisation scales in the hard-scatter calculation, of the showering tune Var3c parameter [45] and of the h_{damp}[7] parameter, while for the modelling of final state QCD radiation alternative choices of factorisation and renormalisation scales in the showering algorithm are considered. Finally, uncertainties on the PDF and the value of the strong coupling constant α_S are also evaluated. All sources of uncertainty have an impact on the shape of the BDT score and the $t\bar{t}$ fractional contribution in each analysis category, with the largest effect due to parton-shower variations ranging between 1% and 10% of the nominal values depending on the analysis category and SR.

For Z + HF processes the uncertainty on the modelling of the hard scatter and the parton shower are evaluated by comparing the nominal sample with a MadGraph5_aMC@NLO +Pythia 8 sample with up to three additional partons in the final state at NLO accuracy in the QCD coupling, in which additional jet multiplicities are merged together with the FxFx NLO matrix-element and parton-shower merging prescription [72]. The A14 parton-shower tune and the NNPDF2.3lo PDF set are used for this alternative sample. Uncertainties from missing higher-order QCD corrections are evaluated with renormalisation and factorisation scale variations from the nominal samples, along with PDF and α_S variations. The effect of higher-order electroweak corrections for Z + HF processes is found to be negligible, and thus not included. Uncertainties on the matching between matrix element calculation and parton shower are considered via variations of the Sherpa matching parameter (CKKW) and the resummation scale (QSF). All sources of uncertainty have an impact on the Z + HF fractional contribution in each analysis category, with the largest effect due to renormalisation and factorisation scale variations ranging up to 13% of the nominal values depending on the analysis category and SR. The uncertainty on the modelling of the hard scatter and the parton shower, from the comparison with the alternative MadGraph5_aMC@NLO +Pythia 8, is the only source of uncertainty with a significant impact on the shape of the BDT score in the analysis SRs. The same source of uncertainty is found to have a non-negligible impact on the shape of the $m_{\ell\ell}$ variable in the CR, which is included as a dedicated uncertainty uncorrelated from the BDT score shape uncertainty in the likelihood fit. Finally, an additional systematic uncertainty is included to cover the residual difference between data and MC simulation in the dedicated CR, as a function of the transverse momentum of the selected lepton pair. This uncertainty is applied in all analysis SRs as a function of the transverse momentum of the τ lepton pair selected from the MC truth record.

[7] The h_{damp} parameter regulates the p_{T} of the first additional emission beyond the Born configuration in the Powheg Box generator, controlling the matching of the matrix element to the parton shower.

For single-top-quark processes only uncertainties related to the Wt−channel are considered, as it is found to be dominant compared with $s-$ and $t-$ channel contributions. The hard-scatter and the parton-shower uncertainties are evaluated by comparing the nominal sample with alternative MadGraph5_aMC@NLO+Pythia8 and Powheg+Herwig7 samples. The uncertainty on the modelling of QCD radiation is evaluated by varying the showering tune Var3c, along with independent variations of the renormalisation and factorisation scales in the hard-scatter and parton-shower calculations. The uncertainty related to the interference between the $t\bar{t}$ and the Wt−channel single-top-quark processes is evaluated by comparing the nominal Wt-channel sample produced with the diagram removal scheme to an alternative sample produced with the diagram subtraction scheme [73]. Finally, uncertainties on the PDF are also evaluated. All sources of uncertainty have an impact on the Wt−channel single-top-quark fractional contribution in each analysis category. Variations due to the uncertainty on the Wt−channel interference scheme range from 1% to 7% in the low-m_{HH} categories, from 23% to 29% in the high-m_{HH} categories and from 14% to 34% in VBF categories. Uncertainties on the Wt−channel interference scheme are also evaluated on the shape of the BDT score, rather than on the p_{T} of the b-tagged jet pair as in Ref. [36]. The hard-scatter and parton-shower uncertainties range from 15% to 36% on the Wt−channel fractional contribution, depending on the analysis category and SR. Uncertainties on the modelling of QCD radiation are found to have a significant impact on the shape of the BDT score.

An uncertainty of 100% is applied on the normalisation of single-Higgs-boson in the ggF, VBF and associated production WH mechanisms where the Higgs boson decays into τ-lepton pairs, to account for difficulties in the modelling of these processes in association with heavy-flavour jets [74, 75]. Uncertainties from missing higher-order QCD corrections are evaluated with independent variations of the renormalisation and factorisation scales from the nominal samples, along with PDF and α_S variations. Parton-shower uncertainties are evaluated by comparing the nominal sample to alternative Powheg+Herwig7 samples for associated production ZH and $t\bar{t}H$ processes. For $t\bar{t}H$ processes the hard-scatter uncertainties are derived by comparing the nominal samples to MadGraph5_aMC@NLO+Pythia8 predictions, and uncertainties on the modelling of QCD radiation are assessed via variations of the showering tune Var3c, along with independent variations of the renormalisation and factorisation scales in the parton-shower algorithm. All sources of uncertainty have an impact on the single-Higgs-boson fractional contribution in each analysis category, while only parton-shower uncertainties are included as variations of the BDT score for ZH and $t\bar{t}H$ processes.

For the SM ggF and VBF HH signal processes, uncertainties from missing higher-order QCD corrections are assessed via independent variations of the renormalisation and factorisation scales in the hard-scatter calculation, along with PDF and α_S variations. Parton-shower uncertainties are evaluated by comparing the nominal samples with alternative Powheg+Herwig7 samples. All sources of uncertainty have an impact on the signal fractional contribution in each analysis category, while only parton-shower uncertainties are included as variations of the BDT score. Cross-section uncertainties for single-Higgs-boson and HH processes [28] follow the same approach detailed in Ref. [36], along with uncertainties on the $H \to b\bar{b}$ and $H \to \tau^+\tau^-$ branching ratios [76].

A dedicated uncertainty on the reweighting method applied to the ggF HH samples to model alternative κ_λ hypotheses, described in Section 3, is defined via a comparison of the SM ggF HH samples with an alternative sample generated assuming $\kappa_\lambda = 10$. The SM sample and the $\kappa_\lambda = 10$ sample are reweighted to a wide range of κ_λ values and the acceptance values of both predictions are compared independently in each category. The maximum of the obtained deviations is taken as an uncertainty and applied to the $\kappa_\lambda = 0$ and $\kappa_\lambda = 20$ templates used for the linear combination of signal samples in the fit. These deviations range from 2% to 4% in the ggF categories and from 20% to 40% in the VBF categories. The uncertainty in the

VBF categories arises from the m_{HH}-based reweighting method not accounting for additional radiation produced with the HH pair, to which the categorisation BDTs are highly sensitive. The linear combination of VBF HH samples described in Section 3 has been found to accurately model alternative κ_λ and κ_{2V} hypothesis, therefore no dedicated uncertainty is considered.

The estimate of systematic uncertainties affecting the data-driven background follows the approach described in Ref. [36]. Uncertainties on the fractional contribution of the data-driven background and on the shape of the BDT score are derived in all the analysis SRs and categories.

6 Results

The statistical procedures applied for extracting results are the same as in Refs. [34, 36]. The global likelihood function $L(\alpha, \theta)$ is constructed from the binned distribution of the BDT output score for signal, background and data distributions in the nine orthogonal analysis categories described in Section 4, together with the $m_{\ell\ell}$ distribution in the dedicated CR. The set α contains the parameters of interest (POI) of the measurement, while θ is the ensemble of nuisance parameters, corresponding to systematic uncertainties constrained by auxiliary measurements in control regions or by theoretical predictions, or to parameters such as the $t\bar{t}$ and Z + HF background yields, which are *a priori* unconstrained.

A dedicated procedure is applied to transform the BDT discriminant to obtain a smoother distribution for the background processes and a finer binning in the regions with the largest signal contribution, while at the same time preserving a sufficiently large number of background events in each bin, similarly to Ref. [36]. Starting from finely-binned histograms, bins are iteratively merged starting from the most signal-like bins until they satisfy the condition of $10 f_s + 5 f_b > 1$, where f_s and f_b are the fractions of signal and background contained in the bin, respectively. The relative statistical uncertainty on the predicted background contribution has to be below 20%, and at least three expected background events are required per bin. The $m_{\ell\ell}$ distribution in the CR is binned uniformly with a width of 1 GeV.

The constraints on the coupling modifiers are determined using a profile likelihood ratio $\Lambda(\alpha, \theta)$ as the test statistic, computed from the likelihood function in the asymptotic approximation [65], where the POI α are the coupling modifiers κ. The procedure adopted in Ref. [34] is used to set constraints on κ_λ and κ_{2V}, expressed as 68% and 95% CIs. Signal strength upper limits are derived using the $\mathrm{CL_s}$ method [77] with the procedure described in Ref. [36]. Upper limits are set on the overall μ_{HH} and on the separate signal strength parameters $\mu_{\mathrm{ggF}} = \sigma_{\mathrm{ggF}}/\sigma^{\mathrm{SM}}_{\mathrm{ggF}}$ and $\mu_{\mathrm{VBF}} = \sigma_{\mathrm{VBF}}/\sigma^{\mathrm{SM}}_{\mathrm{VBF}}$. The expected results are obtained with Asimov datasets [65] generated from the likelihood function after setting all nuisance parameters to their maximum likelihood estimate in the fit to the data and fixing the POIs to the values corresponding to the hypothesis under test. The asymptotic results are found to agree within 7% with the upper limits obtained using pseudo-experiments. Figure 6 shows the BDT score distribution in the nine orthogonal categories, after performing the maximum likelihood fit to data for the $L(\mu_{HH}, \theta)$ function. Good agreement between the data and the prediction is found within the assessed uncertainties.

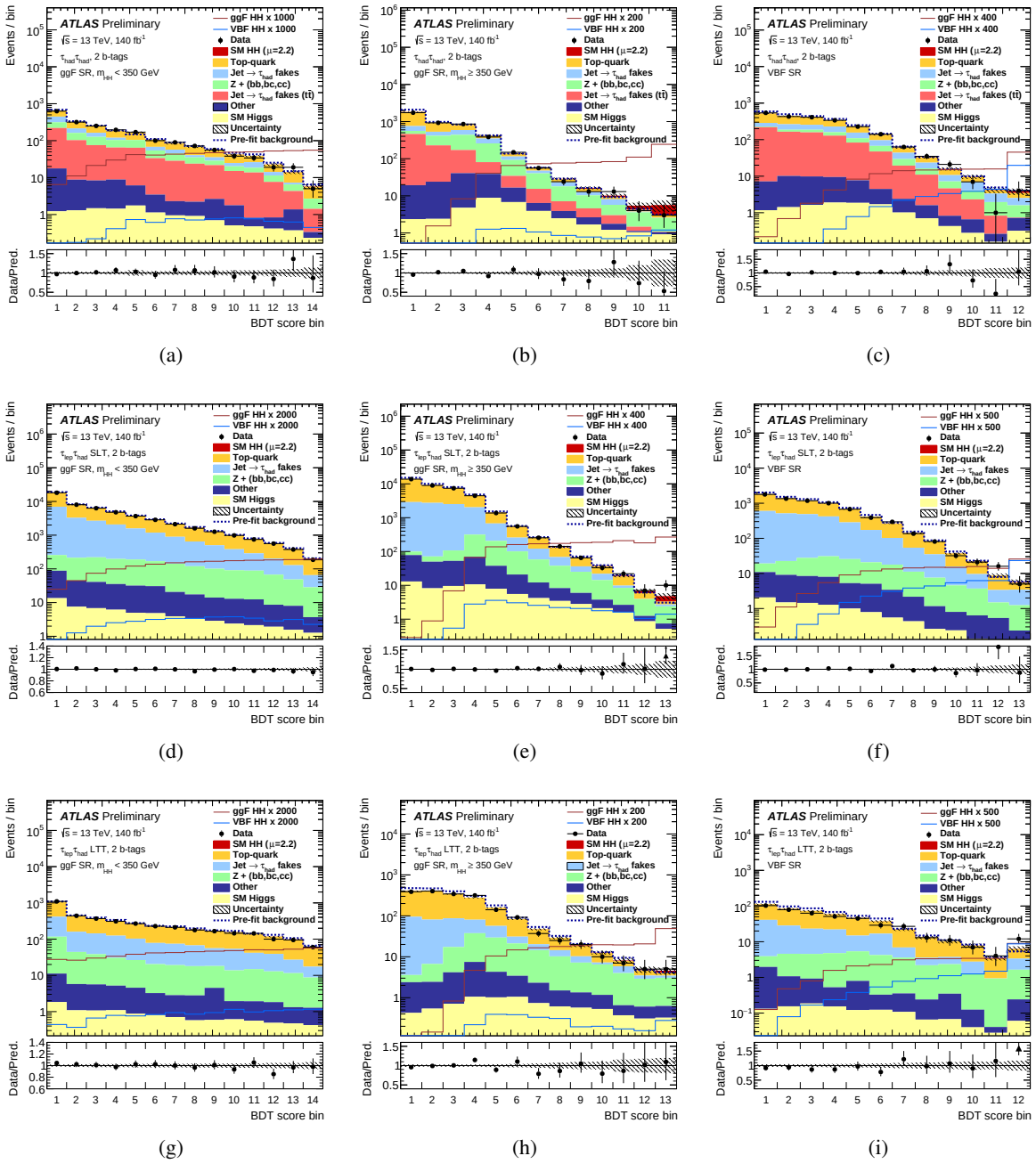


Figure 6: Predicted and observed distribution of the BDT score in the low-m_{HH} (left), high-m_{HH} (middle column) and VBF (right) categories of the $\tau_{\text{had}}\tau_{\text{had}}$ (top), $\tau_{\text{lep}}\tau_{\text{had}}$ SLT (middle row) and $\tau_{\text{lep}}\tau_{\text{had}}$ LTT (bottom) SRs. The signal and background distributions are shown at post-fit level as obtained from the combined likelihood fit to data described in Section 6. The "SM HH" signal contribution is scaled to the fitted signal strength μ_{HH} from the combined likelihood fit times the SM expectation. The ggF and VBF HH signal distributions are overlaid and scaled to the factor indicated in the legend times the SM expectation. The dashed histograms show the total pre-fit background. The lower panels show the ratio of data to the total post-fit sum of signal and background, where the hatched bands show the statistical and systematic uncertainties of this prediction. The BDT score distributions are shown with the binning used in the likelihood fit. For visualisation purposes each bin is displayed with uniform width and the x-axis indicates the bin number.

The maximum likelihood estimator for the total HH production signal strength is found to be $\hat{\mu}_{HH} = 2.2 \pm 1.7$ by the combined fit to data. The uncertainty on the fitted signal strength is obtained from the variation of log-likelihood-based test statistics Λ by one unit, and includes both statistical and systematic uncertainties. The maximum likelihood estimator for the unconstrained normalisation factor of the $t\bar{t}$ and Z + HF backgrounds are measured at 0.96 ± 0.03 and 1.34 ± 0.08, respectively, by the combined fit to data. An observed 95% CL upper limit of 5.9 is set on μ_{HH}, to be compared with an expected limit of 3.1 in the background-only hypothesis. From the simultaneous fit of μ_{ggF} and μ_{VBF}, observed (expected) 95% CL upper limits are $\mu_{\text{ggF}} < 5.8$ (3.2) and $\mu_{\text{VBF}} < 91$ (72), respectively, for each production mode, under the assumption that the signal strength parameters can vary independently for each production mode. If μ_{VBF} is fixed to the SM prediction, the observed (expected) 95% CL upper limit is $\mu_{\text{ggF}} < 5.9$ (3.2). Similarly, if μ_{ggF} is fixed to the SM prediction, the observed (expected) 95% upper limit is $\mu_{\text{VBF}} < 94$ (71).

Expected upper limits for the separate production mode signal strengths are derived with respect to the background-only hypothesis. The signal strength upper limits are summarised in Table 3 and Figure 7, in each SR individually along with the combined fit. The results for the individual SRs are obtained from the combined likelihood fit of the BDT score distributions in the categories of a single SR with the $m_{\ell\ell}$ distribution from the dedicated CR. The observed limit on μ_{HH} from the combined fit is looser than the expected one as a result of an excess in the $\tau_{\text{lep}}\tau_{\text{had}}$ SLT SR, in the high-m_{HH} category. The local significance of this excess with respect to the SM hypothesis ($\mu_{HH} = 1$) is 2.3σ, as obtained from the individual fit of the $\tau_{\text{lep}}\tau_{\text{had}}$ SLT SR.

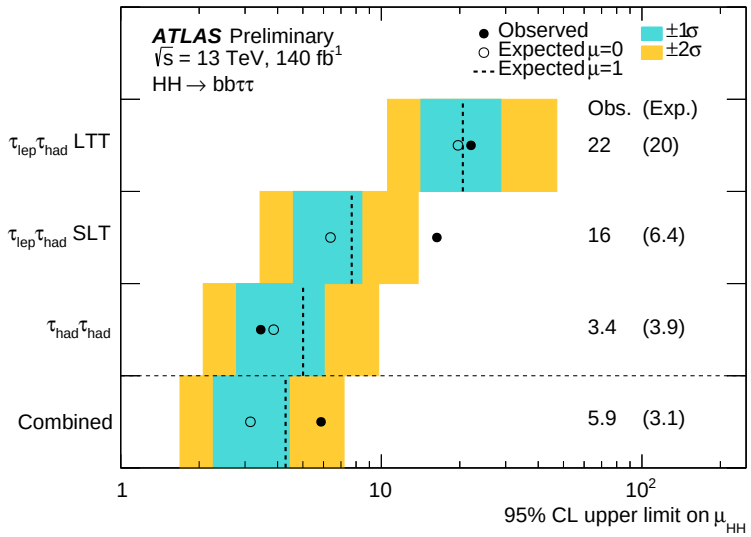


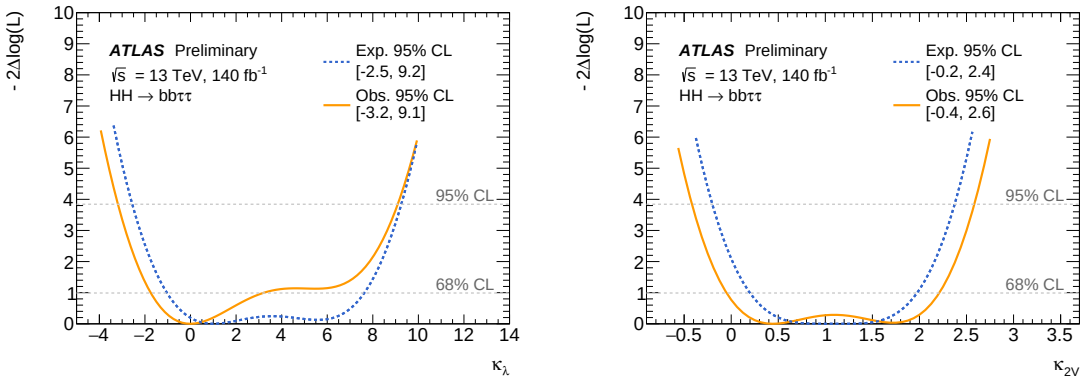
Figure 7: Summary of observed (filled circles) and expected (open circles) 95% CLs upper limits on μ_{HH} from the fit of each individual channel and the combined fit in the background-only ($\mu_{HH} = 0$) hypothesis. The dashed lines indicate the expected 95% CLs upper limits on μ_{HH} in the SM hypothesis ($\mu_{HH} = 1$). The cyan and yellow bands indicate the $\pm 1\sigma$ and $\pm 2\sigma$ variations on the expected limit with respect to the background-only hypothesis due to statistical and systematic uncertainties, respectively.

The observed and expected values of $-2\Delta\log(L)$ as a function of the coupling modifiers κ_λ and κ_{2V} are shown in Figure 8, under the hypothesis that all other coupling modifiers are equal to their SM predictions. The combined fit allows to set observed (expected) 95% CIs of $\kappa_\lambda \in [-3.2, 9.1]$ ($[-2.5, 9.2]$) (assuming $\kappa_{2V} = 1$) and $\kappa_{2V} \in [-0.4, 2.6]$ ($[-0.2, 2.4]$) (assuming $\kappa_\lambda = 1$). Additional constraints are set on κ_λ and κ_{2V} under the assumption that both coupling modifiers can vary simultaneously. The resulting observed

Table 3: Observed and expected 95% CL upper limits on μ_{HH}, μ_{ggF} and μ_{VBF} from the individual SR likelihood fits as well as the combined results. The μ_{ggF} and μ_{VBF} limits are quoted both from the results of the simultaneous fit of both signal strengths (central column), as well as from independent fits for the individual production modes, assuming the other to be SM-like. The uncertainties quoted on the combined expected upper limits correspond to the 1σ uncertainty band.

		μ_{HH}	μ_{ggF}	μ_{VBF}	μ_{ggF} (μ_{VBF}=1)	μ_{VBF} (μ_{ggF}=1)
$\tau_{\text{had}}\tau_{\text{had}}$	observed	3.4	3.6	87	3.5	80
	expected	3.9	4.0	103	3.9	101
$\tau_{\text{lep}}\tau_{\text{had}}$ SLT	observed	16	17	133	17	155
	expected	6.4	6.6	128	6.5	125
$\tau_{\text{lep}}\tau_{\text{had}}$ LTT	observed	22	18	767	21	731
	expected	20	21	323	20	317
Combined	observed	5.9	5.8	91	5.9	94
	expected	$3.1^{+1.3}_{-0.9}$	$3.2^{1.7}_{0.9}$	72^{+32}_{-20}	$3.2^{+1.7}_{-0.9}$	71^{+31}_{-20}

and expected two-dimensional 68% and 95% contours are shown in Figure 9. The observed and expected constraints on κ_λ are affected by the issue concerning the ggF HH prediction for BSM scenarios in Powheg reported in Ref. [78] and resolved in Ref. [79]. If the ggF HH signal yields in the analysis categories are scaled based on the ratio of the predicted differential ggF HH cross-sections with and without the change described in Ref. [79], the width of the 95% CI on κ_λ changes by less than 5%. The results reported in this note do not include the change to the BSM ggF HH prediction described in Ref. [79].



(a) (b)

Figure 8: Values of $-2\Delta\log(L)$ for different κ_λ (a) and κ_{2V} (b) hypotheses obtained from fits to the data (orange) and an Asimov dataset (dashed blue) constructed under the SM hypothesis. In each case, all coupling modifiers but the scanned parameter are fixed to their SM values.

As in Ref. [36], the analysis sensitivity is primarily limited by the statistical uncertainty of the data. The leading systematic uncertainty on the measurement of μ_{HH} is the uncertainty on the ggF HH production cross-section arising from variations of the QCD scales and the top-quark mass scheme. The next leading

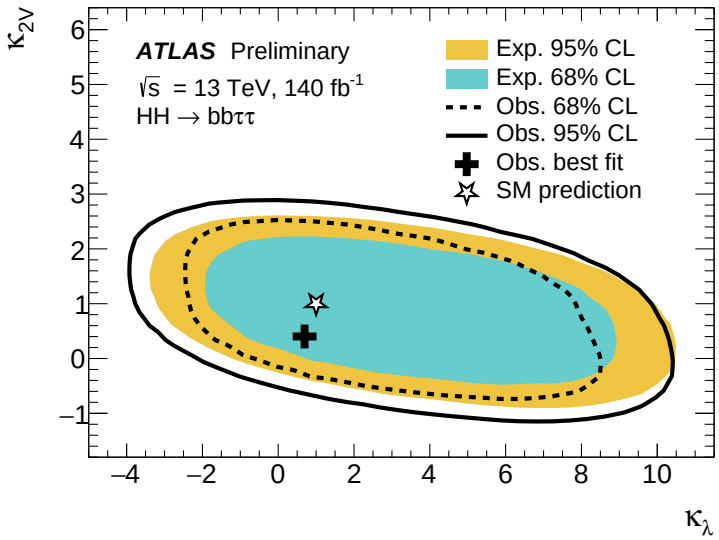


Figure 9: Likelihood contours at 68% (dashed line) and 95% (solid line) CL in the (κ_λ, κ_{2V}) parameter space, when all other coupling modifiers are fixed to one. The corresponding expected contours are shown by the cyan and yellow shaded regions. The SM prediction is indicated by the star, while the best-fit value is denoted by the black cross.

sources of uncertainty are the statistical precision of the background MC samples and the uncertainty related to the interference between the Wt and $t\bar{t}$ processes. The combined impact of all sources of systematic uncertainties leads to an increase in the expected upper limits on the signal strength μ_{HH} by 24% and to a widening of the expected 95% CI for κ_λ and κ_{2V} by 9% and 2%, respectively, with respect to the case in which systematic uncertainties are neglected (excluding the $t\bar{t}$ and Z + HF floating normalisation and MC statistical uncertainties).

Based on a consistent statistical procedure for evaluating the 95% CIs and CLs as described at the beginning of this section, these results can be compared with the previous analysis of Ref. [36]. The approach presented in this note leads to an increase (reduction) of the observed (expected) upper limit on the signal strength μ_{HH} by 25% (20%), along with new results for the upper limits on the separate ggF and VBF HH production mode signal strengths. The width of the observed (expected) CI for κ_λ is reduced by $< 1\%$ (10%) and the width of the observed (expected) CI for κ_{2V} is reduced by 2% (19%) compared to the previous analysis.

The compatibility, considering statistical and systematic uncertainties, between the upper limit at 95% CL on the signal strength μ_{HH} from this study and that of Ref. [36] is evaluated using a bootstrap technique [80] separately for the independent SR fits and for the combined fit. The compatibility between the two results is at the level of 1σ for the individual fit of the $\tau_{\text{had}}\tau_{\text{had}}$ SR, of 2.5σ for the individual fit of the $\tau_{\text{lep}}\tau_{\text{had}}$ SLT SR, and of 0.1σ for the individual fit of the $\tau_{\text{lep}}\tau_{\text{had}}$ LTT SR. The compatibility is at the level of 0.8σ for the combined fit.

7 Effective field theory interpretation

Effective field theory approaches can be employed to parametrise effects of new particles and anomalous couplings on the HH production rates as well as Higgs boson decay rates, assuming that the energy scale

of the underlying BSM processes is far too high to be probed directly. Interpretations of the results in the context of the Higgs Effective Field Theory (HEFT) are presented in this section, following closely the procedure in Refs. [37, 81].

Cross-section limits are placed on the seven m_{HH} shape benchmark models defined in Ref. [82]. These are constructed with a clustering algorithm in order to represent distinct shapes of the m_{HH} spectrum that can be obtained by varying five Wilson coefficients in the HEFT framework [83]. The Wilson coefficient of the only considered anomalous coupling that affects VBF HH production at NLO precision is c_{hhh}, which is equivalent to κ_λ. Due to the much smaller overall expected yield compared to ggF HH production, as well as the decorrelation of the two production modes through the event categorisation, the impact of VBF HH contribution on the m_{HH} shape benchmark limits is expected to be negligible, even when taking c_{hhh} variations into account. Hence, it is ignored in the following, i.e. μ_{VBF} is set to 0.

Signal predictions for ggF HH production assuming different effective coupling strengths are obtained from the SM sample. For this purpose, a reweighting method is used that assigns each event a weight corresponding to the ratio between the differential NLO cross-section (in m_{HH}) predicted by the SM and the hypothesis of interest. These weights are taken from Ref. [82]. Possible effects of the anomalous couplings on the Higgs boson branching fractions are neglected. Uncertainties on the ggF HH event yield related to the reweighting procedure are adopted from Ref. [84]. Additionally, the signal modelling uncertainties are evaluated for each signal configuration, as described in Section 5. The 95% CL upper limits on the ggF HH cross-section limits assuming the different benchmark m_{HH} shapes are summarised in Figure 10. The following changes are made in the fit model with respect to the measurement of signal strength limits described in Section 6. The ggF HH prediction is replaced depending on the benchmark scenario, while the contribution from VBF HH production is neglected. Moreover, as this is a cross-section measurement, no uncertainties on the ggF HH production cross-section are taken into account. In general, it is observed that lower average signal m_{HH} values, as represented by benchmarks 1 and 2, lead to weaker constraints.

The observed and expected constraints on the ggF HH cross-section limits assuming the HEFT shape benchmarks shown in Figure 10 are affected by the issue concerning the ggF HH prediction for BSM scenarios in Powheg reported in Ref. [78] and resolved in Ref. [79]. If the ggF HH signal yields in the analysis categories are scaled based on the ratio of the predicted differential ggF HH cross-sections with and without the change described in Ref. [79], the expected limits changes by less than 10%. The results reported in this note do not include the change to the BSM ggF HH prediction described in Ref. [79].

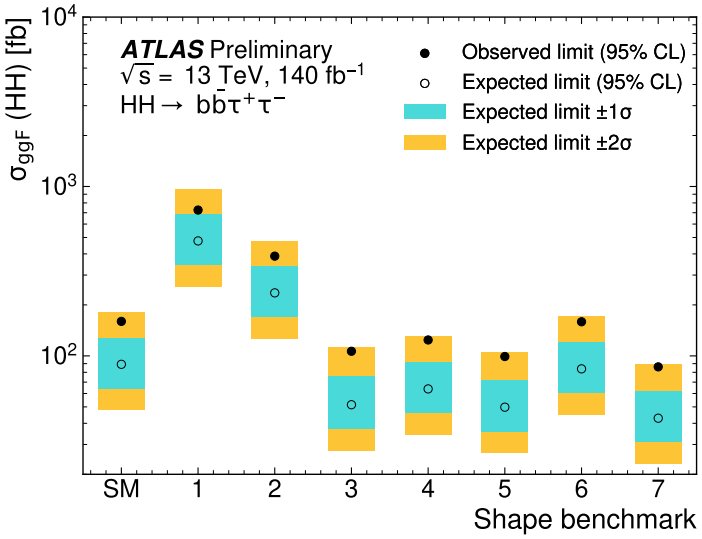


Figure 10: Observed (filled circles) and expected (open circles) 95% CL upper limits on the ggF HH production cross-section with respect to the background-only hypothesis in the SM and each of the seven HEFT shape benchmarks obtained from the combined fit. The cyan and yellow bands show the $\pm 1\sigma$ and $\pm 2\sigma$ variations on the expected upper limit. The contribution from VBF HH production is neglected for this result.

8 Conclusion

An updated search for non-resonant Higgs boson pair production in the $b\bar{b}\tau^+\tau^-$ final state has been performed using the full Run 2 ATLAS dataset, corresponding to 140 fb^{-1} of proton-proton collisions at a centre-of-mass energy of 13 TeV. The results supersede those of a previous search based on the same dataset described in Ref. [36]. Compared with the previous publication, the event classification has been reoptimised to enhance the sensitivity to κ_λ as well as to the VBF production mode. Improved multivariate classifiers are used to build the final discriminants, increasing the sensitivity to SM HH production and to anomalous values of the coupling modifiers κ_λ and κ_{2V}. The analysis sensitivity is improved by 10% to 20%, depending on the parameter of interest. Results interpreted in terms of ggF and VBF production modes have been added, compared to the results of Ref. [36]. The statistical procedure for the interpretation of the observed yields in terms of the signal coupling modifiers has also been updated.

No evidence of HH signal is found. An observed 95% CL upper limit of 5.9 is set on the HH production signal strength μ_{HH}, to be compared with an expected limit of 3.1 in the background-only ($\mu_{HH} = 0$) hypothesis. The observed limit on μ_{HH} is looser than the expected one as a result of a mild excess in the $\tau_{\text{lep}}\tau_{\text{had}}$ SLT SR, in the high-m_{HH} category. The corresponding observed (expected) 95% confidence intervals for the self-coupling modifier κ_λ and the quartic coupling modifier κ_{2V} are $-3.2 < \kappa_\lambda < 9.1$ ($-2.5 < \kappa_\lambda < 9.2$) and $-0.4 < \kappa_{2V} < 2.6$ ($-0.2 < \kappa_{2V} < 2.4$), respectively.

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, *The ATLAS Experiment at the CERN Large Hadron Collider*, JINST **3** (2008) S08003 (cit. on p. 2).

[4] CMS Collaboration, *The CMS Experiment at the CERN LHC*, JINST **3** (2008) S08004 (cit. on p. 2).

[5] ATLAS Collaboration, *Study of the spin and parity of the Higgs boson in diboson decays with the ATLAS detector*, Eur. Phys. J. C **75** (2015) 476, arXiv: 1506.05669 [hep-ex], Erratum: Eur. Phys. J. C **76** (2016) 152 (cit. on p. 2).

[6] ATLAS Collaboration, *CP Properties of Higgs Boson Interactions with Top Quarks in the $t\bar{t}H$ and tH Processes Using $H \to \gamma\gamma$ with the ATLAS Detector*, Phys. Rev. Lett. **125** (2020) 061802, arXiv: 2004.04545 [hep-ex] (cit. on p. 2).

[7] CMS Collaboration, *Constraints on anomalous Higgs boson couplings to vector bosons and fermions in its production and decay using the four-lepton final state*, Phys. Rev. D **104** (2021) 052004, arXiv: 2104.12152 [hep-ex] (cit. on p. 2).

[8] ATLAS Collaboration, *Evidence of off-shell Higgs boson production from ZZ leptonic decay channels and constraints on its total width with the ATLAS detector*, (2023), arXiv: 2304.01532 [hep-ex] (cit. on p. 2).

[9] CMS Collaboration, *Measurement of the Higgs boson width and evidence of its off-shell contributions to ZZ production*, Nature Phys. **18** (2022) 1329, arXiv: 2202.06923 [hep-ex] (cit. on p. 2).

[10] ATLAS Collaboration, *A detailed map of Higgs boson interactions by the ATLAS experiment ten years after the discovery*, Nature **607** (2022) 52, arXiv: 2207.00092 [hep-ex] (cit. on p. 2).

[11] CMS Collaboration, *A portrait of the Higgs boson by the CMS experiment ten years after the discovery*, Nature **607** (2022) 60, arXiv: 2207.00043 [hep-ex] (cit. on pp. 2, 3).

[12] F. Englert and R. Brout, *Broken Symmetry and the Mass of Gauge Vector Mesons*, Phys. Rev. Lett. **13** (1964) 321 (cit. on p. 2).

[13] P. Higgs, *Broken symmetries, massless particles and gauge fields*, Phys. Lett. **12** (1964) 132 (cit. on p. 2).

[14] P. W. Higgs, *Broken Symmetries and the Masses of Gauge Bosons*, Phys. Rev. Lett. **13** (1964) 508 (cit. on p. 2).

[15] G. S. Guralnik, C. R. Hagen and T. W. B. Kibble, *Global Conservation Laws and Massless Particles*, Phys. Rev. Lett. **13** (1964) 585 (cit. on p. 2).

[16] P. W. Higgs, *Spontaneous Symmetry Breakdown without Massless Bosons*, Phys. Rev. **145** (1966) 1156 (cit. on p. 2).

[17] T. W. B. Kibble, *Symmetry Breaking in Non-Abelian Gauge Theories*, Phys. Rev. **155** (1967) 1554 (cit. on p. 2).

[18] ATLAS and CMS Collaborations, *Combined Measurement of the Higgs Boson Mass in pp Collisions at $\sqrt{s} = 7$ and 8 TeV with the ATLAS and CMS Experiments*, Phys. Rev. Lett. **114** (2015) 191803 (cit. on pp. 2, 6).

[19] ATLAS Collaboration, *Combined measurement of the Higgs boson mass from the $H \to \gamma\gamma$ and $H \to ZZ^* \to 4\ell$ decay channels with the ATLAS detector using $\sqrt{s} = 7, 8$ and 13 TeV pp collision data*, ATLAS-CONF-2023-037, 2023, URL: https://cds.cern.ch/record/2865480 (cit. on pp. 2, 6).

[20] P. Agrawal, D. Saha, L.-X. Xu, J.-H. Yu and C.-P. Yuan, *Determining the shape of the Higgs potential at future colliders*, Phys. Rev. D **101** (2020) 075023 (cit. on p. 2).

[21] S. Dawson, S. Dittmaier and M. Spira, *Neutral Higgs boson pair production at hadron colliders: QCD corrections*, Phys. Rev. D **58** (1998) 115012, arXiv: hep-ph/9805244 (cit. on p. 2).

[22] S. Borowka et al., *Higgs Boson Pair Production in Gluon Fusion at Next-to-Leading Order with Full Top-Quark Mass Dependence*, Phys. Rev. Lett. **117** (2016) 012001, arXiv: 1604.06447 [hep-ph], Erratum: Phys. Rev. Lett. **117** (2016) 079901 (cit. on p. 2).

[23] J. Baglio et al., *Gluon fusion into Higgs pairs at NLO QCD and the top mass scheme*, Eur. Phys. J. C **79** (2019) 459, arXiv: 1811.05692 [hep-ph] (cit. on p. 2).

[24] D. de Florian and J. Mazzitelli, *Higgs Boson Pair Production at Next-to-Next-to-Leading Order in QCD*, Phys. Rev. Lett. **111** (2013) 201801, arXiv: 1309.6594 [hep-ph] (cit. on p. 2).

[25] D. Y. Shao, C. S. Li, H. T. Li and J. Wang, *Threshold resummation effects in Higgs boson pair production at the LHC*, JHEP **07** (2013) 169, arXiv: 1301.1245 [hep-ph] (cit. on p. 2).

[26] D. de Florian and J. Mazzitelli, *Higgs pair production at next-to-next-to-leading logarithmic accuracy at the LHC*, JHEP **09** (2015) 053, arXiv: 1505.07122 [hep-ph] (cit. on p. 2).

[27] 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 p. 2).

[28] J. Baglio et al., *$gg \to HH$: Combined uncertainties*, Phys. Rev. D **103** (2021) 056002, arXiv: 2008.11626 [hep-ph] (cit. on pp. 2, 15).

[29] 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. 3).

[30] R. Frederix et al., *Higgs pair production at the LHC with NLO and parton-shower effects*, Phys. Lett. B **732** (2014) 142, arXiv: 1401.7340 [hep-ph] (cit. on p. 3).

[31] 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. 3).

[32] 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. 3).

[33] 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 p. 3).

[34] ATLAS Collaboration, *Constraining the Higgs boson self-coupling from single- and double-Higgs production with the ATLAS detector using pp collisions at $\sqrt{s} = 13$ TeV*,
Phys. Lett. B **843** (2022) 137745, arXiv: 2211.01216 [hep-ex] (cit. on pp. 3, 4, 6, 11, 16).

[35] 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*,
Phys. Rev. D **106** (2021) 052001, arXiv: 2112.11876 [hep-ex] (cit. on p. 3).

[36] 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*,
JHEP **07** (2023) 040, arXiv: 2209.10910 [hep-ex] (cit. on pp. 3–8, 13–16, 19, 20, 22).

[37] ATLAS Collaboration, *Search for nonresonant 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*, Phys. Rev. D **108** (2023) 052003,
arXiv: 2301.03212 [hep-ex] (cit. on pp. 3, 21).

[38] CMS Collaboration, *Search for nonresonant Higgs boson pair production in final state with two bottom quarks and two tau leptons in proton–proton collisions at $\sqrt{s} = 13$ TeV*,
Phys. Lett. B **842** (2022) 137531, arXiv: 2206.09401 [hep-ex] (cit. on p. 4).

[39] ATLAS Collaboration, *The ATLAS Collaboration Software and Firmware*,
ATL-SOFT-PUB-2021-001, 2021, URL: https://cds.cern.ch/record/2767187 (cit. on p. 5).

[40] 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. 5).

[41] ATLAS Collaboration,
Luminosity determination in pp collisions at $\sqrt{s} = 13$ TeV using the ATLAS detector at the LHC,
(2022), arXiv: 2212.09379 [hep-ex] (cit. on pp. 5, 13).

[42] 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 pp. 5, 7).

[43] The NNPDF Collaboration, R. D. Ball et al., *Parton distributions for the LHC run II*,
JHEP **04** (2015) 040, arXiv: 1410.8849 [hep-ph] (cit. on pp. 5, 7).

[44] T. Sjöstrand, S. Mrenna and P. Skands, *A brief introduction to PYTHIA 8.1*,
Comput. Phys. Commun. **178** (2008) 852, arXiv: 0710.3820 [hep-ph] (cit. on pp. 5, 7).

[45] 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 pp. 5, 7, 14).

[46] ATLAS Collaboration, *Summary of ATLAS Pythia 8 tunes*, ATL-PHYS-PUB-2012-003, 2012,
URL: https://cds.cern.ch/record/1474107 (cit. on pp. 5, 7).

[47] E. Bothmann et al., *Event generation with Sherpa 2.2*, SciPost Phys. **7** (2019) 034,
arXiv: 1905.09127 [hep-ph] (cit. on pp. 5, 7).

[48] F. Buccioni et al., *OpenLoops 2*, Eur. Phys. J. C **79** (2019) 866, arXiv: 1907.13071 [hep-ph]
(cit. on p. 5).

[49] F. Cascioli, P. Maierhöfer and S. Pozzorini, *Scattering Amplitudes with Open Loops*, Phys. Rev. Lett. **108** (2012) 111601, arXiv: 1111.5206 [hep-ph] (cit. on p. 5).

[50] F. Buccioni, S. Pozzorini and M. Zoller, *On-the-fly reduction of open loops*, Eur. Phys. J. C **78** (2018) 70, arXiv: 1710.11452 [hep-ph] (cit. on p. 5).

[51] A. Denner, S. Dittmaier and L. Hofer, *COLLIER: A fortran-based complex one-loop library in extended regularizations*, Comput. Phys. Commun. **212** (2017) 220, arXiv: 1604.06792 [hep-ph] (cit. on p. 5).

[52] T. Gleisberg and S. Höche, *Comix, a new matrix element generator*, JHEP **12** (2008) 039, arXiv: 0808.3674 [hep-ph] (cit. on p. 5).

[53] S. Höche, F. Krauss, M. Schönherr and F. Siegert, *QCD matrix elements + parton showers. The NLO case*, JHEP **04** (2013) 027, arXiv: 1207.5030 [hep-ph] (cit. on p. 5).

[54] ATLAS Collaboration, *Measurement of the Z/γ^* boson transverse momentum distribution in pp collisions at $\sqrt{s} = 7$ TeV with the ATLAS detector*, JHEP **09** (2014) 145, arXiv: 1406.3660 [hep-ex] (cit. on pp. 6, 7).

[55] G. Heinrich, J. Lang and L. Scyboz, *SMEFT predictions for $gg \to hh$ at full NLO QCD and truncation uncertainties*, JHEP **08** (2022) 079, arXiv: 2204.13045 [hep-ph] (cit. on p. 6).

[56] 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 pp. 6, 7).

[57] 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. 6).

[58] ATLAS Collaboration, *The ATLAS Simulation Infrastructure*, Eur. Phys. J. C **70** (2010) 823, arXiv: 1005.4568 [physics.ins-det] (cit. on p. 6).

[59] S. Agostinelli et al., *GEANT4 – a simulation toolkit*, Nucl. Instrum. Meth. A **506** (2003) 250 (cit. on p. 6).

[60] J. Butterworth et al., *PDF4LHC recommendations for LHC Run II*, J. Phys. G **43** (2016) 023001, arXiv: 1510.03865 [hep-ph] (cit. on p. 7).

[61] ATLAS Collaboration, *Search for non-resonant Higgs boson pair production in the $2b + 2\ell + E_T^{miss}$ final state in pp collisions at $\sqrt{s} = 13$ TeV with the ATLAS detector*, ATLAS-CONF-2023-064, 2023, URL: https://cds.cern.ch/record/2873518 (cit. on p. 6).

[62] A. Elagin, P. Murat, A. Pranko and A. Safonov, *A new mass reconstruction technique for resonances decaying to $\tau\tau$*, Nucl. Instrum. Meth. A **654** (2011) 481, arXiv: 1012.4686 [hep-ex] (cit. on p. 8).

[63] ATLAS Collaboration, *ATLAS flavour-tagging algorithms for the LHC Run 2 pp collision dataset*, Eur. Phys. J. C **83** (2022) 681, arXiv: 2211.16345 [physics.data-an] (cit. on p. 8).

[64] A. Hoecker et al., *TMVA - Toolkit for Multivariate Data Analysis*, 2007, arXiv: physics/0703039 [physics.data-an] (cit. on p. 9).

[65] 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 pp. 10, 16),
Erratum: Eur. Phys. J. C **73** (2013) 2501.

[66] C. Bernaciak, M. S. A. Buschmann, A. Butter and T. Plehn, *Fox-Wolfram moments in Higgs physics*, Phys. Rev. D **87** (2013) 073014 (cit. on p. 10).

[67] L. A. Spiller, *Modification of Fox-Wolfram moments for hadron colliders*, JHEP **03** (2016) 27 (cit. on p. 10).

[68] C. G. Lester, *The stransverse mass, M_{T2}, in special cases*, JHEP **05** (2011) 76 (cit. on p. 13).

[69] J. H. Kim, K. Kong, K. T. Matchev and M. Park,
Probing the Triple Higgs Self-Interaction at the Large Hadron Collider,
Phys. Rev. Lett. **122** (9 2019) 091801 (cit. on p. 13).

[70] R. Barlow and C. Beeston, *Fitting using finite Monte Carlo samples*,
Comput. Phys. Commun. **77** (1993) 219 (cit. on p. 13).

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

[72] R. Frederix and S. Frixione, *Merging meets matching in MC@NLO*, JHEP **12** (2012) 061,
arXiv: 1209.6215 [hep-ph] (cit. on p. 14).

[73] S. Frixione, E. Laenen, P. Motylinski, C. White and B. R. Webber,
Single-top hadroproduction in association with a W boson, JHEP **07** (2008) 029,
arXiv: 0805.3067 [hep-ph] (cit. on p. 15).

[74] ATLAS Collaboration,
Measurements of inclusive and differential fiducial cross-sections of $t\bar{t}$ production with additional heavy-flavour jets in proton–proton collisions at $\sqrt{s}$ = 13 TeV with the ATLAS detector,
JHEP **04** (2019) 046, arXiv: 1811.12113 [hep-ex] (cit. on p. 15).

[75] ATLAS Collaboration, *Measurement of the cross-section for W boson production in association with b-jets in pp collisions at $\sqrt{s}$ = 7 TeV with the ATLAS detector*, JHEP **06** (2013) 084,
arXiv: 1302.2929 [hep-ex] (cit. on p. 15).

[76] 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 p. 15).

[77] A. L. Read, *Presentation of search results: the CL_S technique*, J. Phys. G **28** (2002) 2693
(cit. on p. 16).

[78] G. Heinrich, J. Lang and L. Scyboz,
Erratum to: SMEFT predictions for $gg \to hh$ at full NLO QCD and truncation uncertainties,
JHEP **10** (2023) 086 (cit. on pp. 19, 21).

[79] E. Bagnaschi, G. Degrassi and R. Gröber, *Higgs boson pair production at NLO in the POWHEG approach and the top quark mass uncertainties*, (2023), arXiv: 2309.10525 [hep-ph]
(cit. on pp. 19, 21).

[80] K. G. Hayes, M. L. Perl and B. Efron,
Application of the bootstrap statistical method to the tau-decay-mode problem,
Phys. Rev. D **39** (1989) 274 (cit. on p. 20).

[81] ATLAS Collaboration, *Studies of new Higgs boson interactions through nonresonant HH production in the $b\bar{b}\gamma\gamma$ final state in pp collisions at $\sqrt{s}$ = 13 TeV with the ATLAS detector*, ATLAS-CONF-2023-050, 2023, URL: https://cds.cern.ch/record/2870214 (cit. on p. 21).

[82] L. Alasfar et al., *Effective Field Theory descriptions of Higgs boson pair production*, LHCHWG-2022-004, 2023, URL: https://cds.cern.ch/record/2843280 (cit. on p. 21).

[83] M. Capozi and G. Heinrich, *Exploring anomalous couplings in Higgs boson pair production through shape analysis*, JHEP **03** (2020) 091 (cit. on p. 21).

[84] ATLAS Collaboration, *HEFT interpretations of Higgs boson pair searches in $b\bar{b}\gamma\gamma$ and $b\bar{b}\tau^+\tau^-$ final states and of their combination in ATLAS*, ATL-PHYS-PUB-2022-019, 2022, URL: https://cds.cern.ch/record/2806411 (cit. on p. 21).