CMS Physics Analysis Summary

Contact: cms-pag-conveners-higgs@cern.ch 2026/02/04

Search for non-resonant Higgs boson pair production in the $b\bar{b}WW$ decay channel with two leptons in the final state using proton-proton collision data at $\sqrt{s} = 13.6$ TeV

The CMS Collaboration

Abstract

A search for non-resonant Higgs boson pair production is presented, targeting final states where one Higgs boson decays to a pair of bottom quarks and the other Higgs boson decays to two W bosons, both of which decay leptonically (to an electron or a muon and a neutrino). For the first time, the search is conducted with proton-proton collision data from the LHC at $\sqrt{s} = 13.6$ TeV. The data have been recorded by the CMS detector in 2022 and 2023, and correspond to an integrated luminosity of 62 fb^{-1}. The data are consistent with standard model predictions. An upper limit is set on the Higgs boson pair production cross section of 12.7 times the standard model prediction at 95% confidence level, with an expected limit of 18.6. The results are also used to constrain the strength of the trilinear coupling of the Higgs boson as well as of the quartic coupling between two Higgs bosons and two vector bosons.

1 Introduction

The measured properties of the Higgs boson discovered [1–3] by the ATLAS and CMS Collaborations at the CERN LHC are consistent within the current experimental precision [4, 5] with the expectations of the standard model (SM) Higgs mechanism. One aspect of the Higgs mechanism that is experimentally largely unknown is the exact shape of the Higgs potential, which, in the SM, is related to the trilinear coupling of the Higgs boson and can be directly probed using Higgs boson pair (HH) production. In the SM, the dominant HH production mode at the LHC is gluon-gluon fusion (ggF), followed by vector boson fusion (VBF) production, with cross sections of $34.1^{+2.2}_{-7.9}$ fb [6–8] and 1.87 ± 0.05 fb [9, 10], respectively, at 13.6 TeV for a Higgs boson mass of 125 GeV. The VBF production mode is furthermore sensitive to the quartic coupling between two Higgs bosons and two vector bosons.

The HH production process has not yet been observed. Previous searches for HH production have been performed by the ATLAS and CMS Collaborations with LHC Run 2 data at 13 TeV in various final states and their combinations [11–13]. The CMS Collaboration has also conducted searches for HH production with LHC Run 3 data at 13.6 TeV in final states with two bottom quarks and two photons ($b\bar{b}\gamma\gamma$) [14] and with Run 2 and Run 3 data combined in final states with four bottom quarks [15], while the ATLAS Collaboration published a search in the $b\bar{b}\gamma\gamma$ final state using Run 2 and Run 3 data [16].

This note presents a search for non-resonant HH production, targeting final states where one Higgs boson decays to $b\bar{b}$ and the other Higgs boson decays to two W bosons, both of which decay leptonically to an electron or a muon and a neutrino. Previous searches in the $b\bar{b}WW$ channel with the Run 2 dataset were performed by the CMS Collaboration in final states with one or two leptons, setting an observed (expected) upper limit on the inclusive HH production cross section of 14 (18) times the SM prediction at 95% confidence level (CL) [17], and by the ATLAS Collaboration in the two-lepton final state with an upper limit of 9.7 (16.2) times the SM prediction at 95% CL [18].

In the presented search, events are selected with exactly two opposite-sign leptons ($\mu^+\mu^-$, e^+e^-, $e^\pm\mu^\mp$). The selected events are categorised based on the output of a multiclassification neural network (NN) into several regions targeting HH signal events from ggF or VBF production, or background events from different processes, and further based on the b-tagged jets multiplicity as well as the presence of a boosted $H \to b\bar{b}$ candidate jet. The events in the signal-enriched regions are further classified by dedicated binary NNs designed to separate ggF or VBF signal events from background events. The output distributions of the binary NNs together with the event yield in the background-enriched regions enter the final fit as sensitive observables. An upper limit is set on the inclusive HH production cross section, and the result is used to constrain the strength of the trilinear as well as the quartic couplings. Throughout this note, the values of the Higgs boson couplings are expressed in terms of coupling modifiers κ, denoting the coupling strength relative to the SM prediction [19]. For example, the modifier of the trilinear coupling is κ_λ and of the quartic coupling κ_{2V}.

Several refinements have been implemented relative to the previous analysis [17]. A new classification strategy, utilizing a staged multiclassification-plus-binary NN approach, resulted in an approximately 50% improvement in the exclusion limits. Furthermore, the adoption of a more efficient b tagging algorithm and an improved trigger strategy increased the signal yield by 7% and 5%, respectively, up to a 17% increase in the most signal-like bins of the final discriminant. Finally, the analysis benefits from the higher center-of-mass energy of Run 3, which provides a 5% increase in number of signal to square root of number of background events.

This document is structured as follows. The object and event reconstruction is described in Section 2, followed by a description of the data and the simulated event samples in Section 3. The event selection and the analysis strategy are detailed in Section 4, and systematic uncertainties are discussed in Section 5. The statistical analysis and the obtained results are presented in Section 6. Finally, a summary is provided in Section 7.

2 Event reconstruction

Individual particles in an event are reconstructed using a global "particle-flow" (PF) algorithm [20] that combines information provided by different detector systems of the CMS apparatus [21, 22].

The reconstruction of muons relies on a combination of measurements in the tracker and in the muon detectors [23]. The muons are identified based on the quality of the combined track fit and on the number of hits in the different tracking detectors, with an efficiency of about 90%, including the isolation requirements described below. The transverse momentum (p_T) resolution in the barrel is better than 1% (7%) for muons with p_T up to 100 GeV (1 TeV) [23]. Muons are required to lie within the acceptance of the muon system, covering pseudorapidity (η) values of $|\eta| < 2.4$, and to have $p_T > 25$ and 15 GeV for the p_T-leading and subleading muons, respectively.

Electrons are reconstructed by combining the momentum measurement in the tracker with the energy measurement of the corresponding cluster in the electromagnetic calorimeter (ECAL) and with the energy sum of all bremsstrahlung photons, obtained from the ECAL, spatially compatible with originating from the electron track [24, 25]. The electrons are identified using criteria on the cluster shape in the ECAL, the track quality, and the compatibility between the tracker and ECAL measurements, corresponding to an identification efficiency of approximately 80%, including the isolation requirements described below. The momentum resolution for electrons with $p_T \approx 45$ GeV ranges from 1.6 to 5% [24, 25]. Electrons are required to have $|\eta| < 2.4$ as well as $p_T > 25$ and 15 GeV for the p_T-leading and subleading electrons, respectively. Electrons reconstructed in the transition region $1.44 < |\eta| < 1.56$ between the barrel and the endcap calorimeters are discarded.

Isolation requirements are imposed on the muons and electrons with the aim to suppress leptons originating from decays of heavy hadrons inside jets. The isolation is based on the scalar p_T sum of all PF objects in a cone of radius 0.3 in the η-ϕ plane around the track direction of the lepton, where ϕ denotes the azimuthal angle, and it is corrected by removing contributions arising from additional pp interactions in the same bunch crossing (pileup) [26].

Hadronic jets are clustered from the PF objects using the anti-k_T algorithm [27, 28] with a distance parameter of 0.4 (jets) and of 0.8 (large-radius jets). The pileup per particle identification algorithm (PUPPI) [26, 29] is used to mitigate the effect of pileup at the reconstructed particle level. Jet energy corrections are applied to bring the measured response of jets to that of particle level jets on average [30]. The jet energy resolution amounts typically to 15–20% at 30 GeV, 10% at 100 GeV, and 5% at 1 TeV [30]. Jets (large-radius jets) overlapping within a cone of radius 0.4 (0.8) in the η-ϕ plane with a muon or an electron passing the criteria described above are discarded. For the analysis, jets and large-radius jets with $|\eta| < 2.4$ as well as with $p_T > 25$ GeV and $p_T > 200$ GeV, respectively, are selected.

In addition, jets with $|\eta| > 2.4$ (forward jets) are used in the identification of HH production via the VBF process, which is characterised by the presence of jets with large $|\eta|$ values. Forward

jets with $p_T > 50\,\text{GeV}$ ($p_T > 30\,\text{GeV}$) and $2.4 < |\eta| < 3$ ($|\eta| > 3$) are selected. Out of all jets and forward jets that are not used for the $H \to b\bar{b}$ reconstruction all possible pairs are selected, and the pair with the largest $\Delta\eta$ is referred to as VBF jet pair. An event is considered as VBF tagged if the VBF jet pair has $\Delta\eta > 3$ and an invariant dijet mass larger than 500 GeV.

Jets arising from the hadronisation of b quarks (b jets) are identified using the PARTICLENET tagging algorithm [31, 32]. Jets are considered as b-tagged if they fulfil a requirement on the PARTICLENET discriminant that corresponds to a b tagging efficiency of approximately 80% and a 1% misidentification rate for light-flavour quark and gluon-induced jets. The PARTICLENET algorithm is further used to identify Higgs bosons with large p_T decaying to $b\bar{b}$ [31, 33]. Large-radius jets are considered as boosted $H \to b\bar{b}$ candidates if they fulfil a requirement on the PARTICLENET discriminant that corresponds to an efficiency of approximately 26% (22%) of identifying $H \to b\bar{b}$ jets for ggF (VBF) production, with a background misidentification rate of 2%. The b jet (mis)identification probabilities in simulated events are corrected to match the efficiencies measured in data [34, 35], and uncertainties are assigned to account for residual mismodelling of the $H \to b\bar{b}$ tagging efficiency.

3 Data and simulated event samples

Data from pp collisions at $\sqrt{s} = 13.6\,\text{TeV}$ collected in 2022 and 2023, corresponding to an integrated luminosity of $62\,\text{fb}^{-1}$, are analysed. The data are reconstructed separately in four different data-taking periods with different detector conditions—$8\,\text{fb}^{-1}$ and $27\,\text{fb}^{-1}$ of the data collected in 2022 as well as $18\,\text{fb}^{-1}$ and $10\,\text{fb}^{-1}$ of the data collected in 2023—and appropriate per-period calibrations are applied, before the data are combined for the analysis. The simulated events described in the following are generated accordingly in four different samples, reflecting the different data-taking conditions.

Signal and background events are modelled using the Monte Carlo (MC) event generators POWHEG v. 2 [36–39] or MADGRAPH5_aMC@NLO v. 2.9.13 [40], interfaced with a detailed detector simulation based on GEANT4 [41]. The proton structure is described by the parton distribution function (PDF) set NNPDF3.1 [42]. Parton showering (PS) and hadronisation are simulated with PYTHIA v. 8.306 [43], where the parameters for the underlying event description correspond to the CP5 tune [44]. For comparison with the observed distributions, the events in the simulated samples are normalised to the same integrated luminosity as the data sample, according to their predicted cross sections. Pileup effects are modelled by adding simulated minimum bias events to all simulated events with a frequency distribution adjusted to match that observed in the data.

The HH signal process in the ggF production channel is simulated at next-to-leading order (NLO) precision in QCD perturbation theory with POWHEG [45]. Samples with four different values of κ_λ are generated, allowing modelling kinematic distributions for a large range of κ_λ values using a weighted sum of the samples [46]. Signal events in the VBF production channel are generated with MADGRAPH5_aMC@NLO at leading order (LO) precision. Ten samples with varying values of κ_λ and κ_{2V} are produced, which allows modelling the kinematic distributions for different values of κ_λ and κ_{2V}, following the same principle as for the ggF samples. The resulting inclusive HH cross section for each value of the coupling modifiers is normalised to a calculation at next-to-NLO (NNLO) precision in QCD [6–8]. All relevant HH decay modes are considered, with relevant contributions arising from $b\bar{b}WW$, $b\bar{b}\tau\tau$, and $b\bar{b}ZZ$.

The final state targeted by this search is dominated by background processes from top quark-antiquark ($t\bar{t}$) and single top quark (single-$t/\bar{t}$) production as well as $Z/\gamma^* + \text{jets}$, referred to as

Drell–Yan (DY), production. The overall normalisation of the $t\bar{t}$ and DY processes are measured in-situ in the final fit to data described in Section 6.

The $t\bar{t}$ background process is simulated at NLO precision in QCD using POWHEG [47]. The top quark p_T spectrum in the POWHEG simulation has been found to be harder than the one observed in data, attributed to missing higher-order calculations [48, 49], and therefore, the simulated events are reweighted to match the prediction at NNLO precision in QCD and NLO precision in electroweak corrections [50]. The corrections have been computed for 13 TeV [50], and an additional small correction is applied that extrapolates the corrections to 13.6 TeV, following Ref. [51]. The inclusive cross section is normalised to a calculation at NNLO precision in QCD including resummation of next-to-next-to-leading logarithmic soft-gluon terms [52–58].

The single-$t/\bar{t}$ background encompasses single top quark production in the t- and s channels as well as tW production. The processes are simulated at NLO precision in QCD, in the t and tW channels with POWHEG [59, 60], and in the s channel with MADGRAPH5_aMC@NLO. The inclusive cross sections of the t and s channel processes are normalised to the predictions at NLO precision [61, 62] and the tW channel to calculations at approximate NNLO precision [63].

The DY background process is simulated with up to two partons at the matrix element (ME) level at NLO precision in QCD with MADGRAPH5_aMC@NLO, where the merging of the jets from ME calculation and the PS is done using the FXFX [64] prescription. Dedicated corrections, described later in Section 4.4, are derived from data to improve the modelling of relevant observables such as the jet multiplicity spectrum in DY events. The DY events are separated into two mutually exclusive processes, based on the flavour of the jets at the particle-level, where these jets fulfil the acceptance requirements of $p_T > 20\,\text{GeV}$ and of $|\eta| < 2.4$. Events with at least one particle-level jet that contains one or more b- or c-flavoured hadrons are classified as DY + heavy-flavour jets (DY + HF) events, and all other events are classified as DY + light-flavour and gluon jets (DY + LF) events.

Minor background contributions arise from the resonant production of single Higgs bosons (H production). The processes are simulated at NLO precision in the ggF, VBF, VH, ggZH, and $t\bar{t}H$ production channels using POWHEG and in the $t\bar{t}VH$ production channel using MADGRAPH5_aMC@NLO, as well as at LO precision in the tHq and tHW channels using MADGRAPH5_aMC@NLO. Their cross sections are normalised to the latest available calculations, taken from Ref. [7].

Further minor background contributions stem from W + jets production, the production of two (VV) or three (VVV) Z or W bosons, $t\bar{t}$ production in association with one ($t\bar{t}V$) or two ($t\bar{t}VV$) Z or a W bosons, and from the production of four top quarks ($t\bar{t}t\bar{t}$). They are simulated using POWHEG (VV, VVV) and MADGRAPH5_aMC@NLO (W + jets, $t\bar{t}V$).

4 Event selection and analysis strategy

Events are selected with exactly two opposite-sign leptons ($\mu^+\mu^-$, e^+e^-, $e^\pm\mu^\mp$). The selected events are categorised into several signal regions (SRs), targeting ggF and VBF production, or into background control regions (CRs), based on the output of a multiclassification NN as well as the b-tagged jets multiplicity and the presence of a boosted $H \to b\bar{b}$ candidate jet. The events in the SRs are classified by additional NNs designed to separate ggF or VBF signal events from background events. The output distributions of the binary NNs in the SRs together with the event yield in the CRs enter the final fit as sensitive observables.

Contributions from SM background processes are modelled using the simulated events de-

scribed in Section 3. Dedicated corrections are derived for the DY events to improve the modelling of the data, and the normalisation of the dominant $\mathrm{t\bar{t}}$ and DY background processes are determined in the final fit to data.

The event selection, the analysis strategy, the details of the NN architecture and training, as well as the background modelling are detailed in the following, and the analysis strategy is illustrated in Fig. 1.

4.1 Event selection

The data are selected using a combination of triggers. The triggers require the presence of either one or two muon candidates for the $\mu^+\mu^-$ events, one or two electron candidates for the $\mathrm{e}^+\mathrm{e}^-$ events, as well as one muon or one electron candidate or one muon and one electron candidate for the $\mathrm{e}^\pm\mu^\mp$ events. The minimal thresholds on the trigger objects are $p_\mathrm{T} > 8\,\mathrm{GeV}$ for the muons and $p_\mathrm{T} > 12\,\mathrm{GeV}$ for the electrons. In addition, to maximize the signal selection efficiency for $\mathrm{e}^+\mathrm{e}^-$ and $\mathrm{e}^\pm\mu^\mp$ events, the analysis uses triggers requiring the presence of one electron and one jet with $p_\mathrm{T} > 165\,\mathrm{GeV}$. The selection efficiency with respect to the offline reconstruction achieved for signal events after the event selection described below is 98% in the $\mu^+\mu^-$ channel, 95% in the $\mathrm{e}^+\mathrm{e}^-$ channel, and 96% in the $\mathrm{e}^\pm\mu^\mp$ channel. The single-lepton triggers and the electron+jet triggers contribute up to 10% to the signal selection efficiency, depending on the channel.

A further offline selection, referred to as baseline selection, requires events to have exactly two muons or two electrons or one muon and one electron of opposite charge, and at least one b-tagged jet. Events are rejected if further muons or electrons are present with $p_\mathrm{T} > 5\,\mathrm{GeV}$ or $p_\mathrm{T} > 7\,\mathrm{GeV}$, respectively. Events must get accepted by a trigger based on online leptons with a flavour content consistent with that of the two leptons selected offline; for example, $\mu^+\mu^-$ events are required to pass a dimuon or single-muon trigger. The invariant mass of the selected lepton pair, $m_{\ell\ell}$, is required to be larger than 20 GeV to suppress events from heavy-flavour resonance decays and low-mass DY processes. The baseline event selection criteria are summarised in Table 1.

Table 1: Baseline event selection criteria.

Number of leptons	2
Sign and flavour of leptons	$\mu^+\mu^-$, $\mathrm{e}^+\mathrm{e}^-$, $\mathrm{e}^\pm\mu^\mp$
p_T of leading lepton	$>25\,\mathrm{GeV}$
p_T of sub-leading lepton	$>15\,\mathrm{GeV}$
$\lvert\eta\rvert$ of electrons	<2.5
$\lvert\eta\rvert$ of muons	<2.4
$m_{\ell\ell}$	$>20\,\mathrm{GeV}$
Number of jets	≥ 1
Number of b-tagged jets	≥ 1
p_T of jets	$>25\,\mathrm{GeV}$
$\lvert\eta\rvert$ of jets	<2.4

The selected events are then categorised based on the lepton flavour into three mutually exclusive channels, $\mu^+\mu^-$, $\mathrm{e}^+\mathrm{e}^-$, and $\mathrm{e}^\pm\mu^\mp$. The events are further separated into an analysis region with $20 < m_{\ell\ell} < 70\,\mathrm{GeV}$ and $p_\mathrm{T}^\mathrm{miss} > 40\,\mathrm{GeV}$ as well as into validation regions with $70 < m_{\ell\ell} < 110\,\mathrm{GeV}$, enriched in DY events, and with $m_{\ell\ell} > 110\,\mathrm{GeV}$, enriched in $\mathrm{t\bar{t}}$ events.

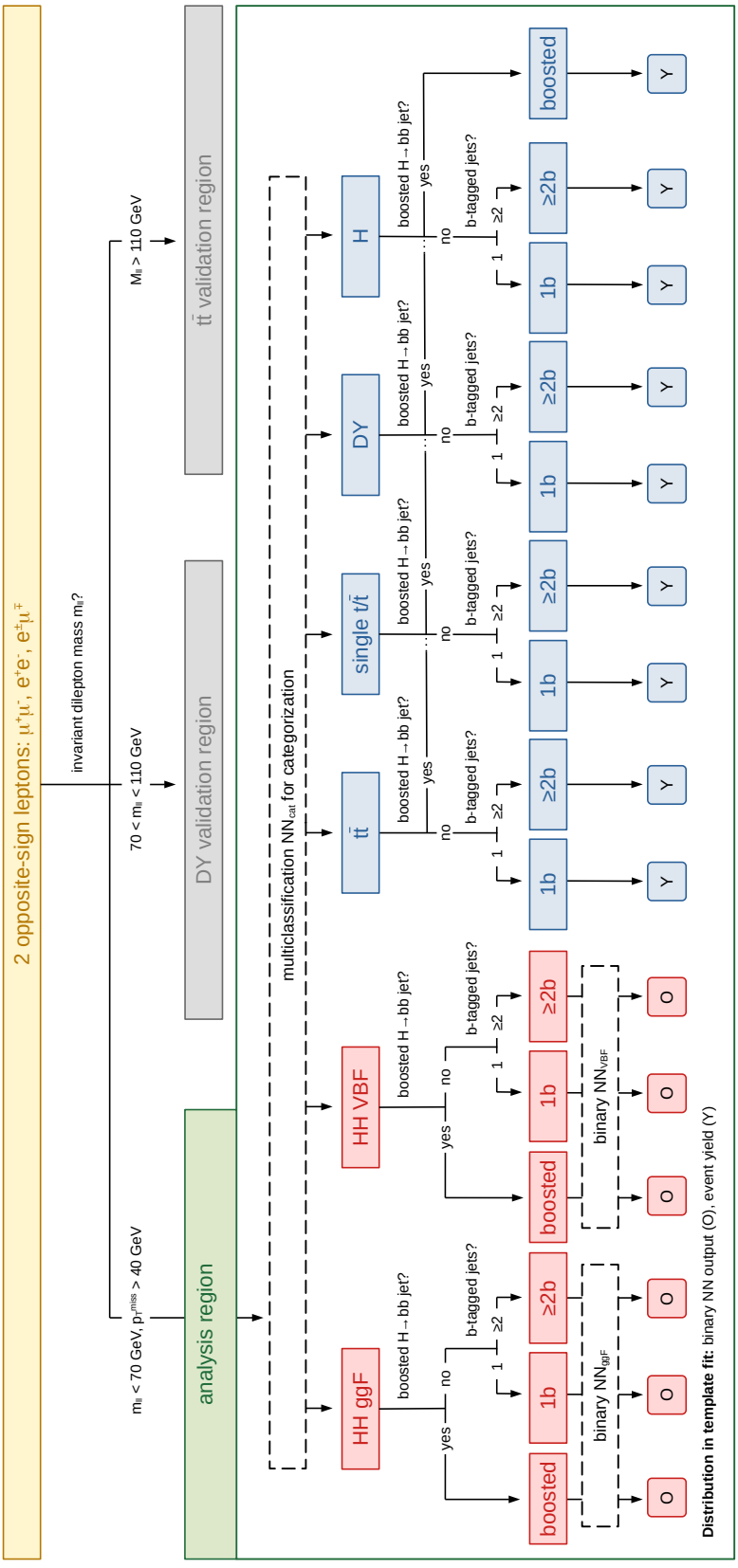

Figure 1: Illustration of the event categorisation: SRs are depicted in red, background CRs in blue. Details of the NNs are described in the text. The binary NN output distributions (O) and the event yields (Y) in the CRs enter the final fit as sensitive observables.

The analysis region events are used in the final analysis to extract the signal, and the validation region events are used to verify the DY and $t\bar{t}$ background modelling and derive corrections for the DY background.

4.2 Analysis strategy

The events in the analysis region are categorised in several steps, as illustrated in Fig. 1. First, the events are categorised by a multiclassification NN (NN_{cat}) designed to separate between different signal and background processes. The values obtained in the output nodes of the NN_{cat} are normalised to unity using a "soft-max" function, and, as a result, the output values can be interpreted as probabilities describing the likelihood of the event being of a certain process. Events are assigned to a category corresponding to the most probable process according to this NN multiclassification. The events are separated into one of two SRs targeting ggF HH production (SR_{ggF}) or VBF HH production (SR_{VBF}), or into one of four background CRs targeting the main background processes: $t\bar{t}$, single-t/$\bar{t}$, DY, and H production.

The events in the SR_{ggF} and SR_{VBF} are further sorted into one of three categories: if a boosted $H \to b\bar{b}$ candidate jet is present, into the "boosted" categories $SR^{boosted}_{ggF}$ and $SR^{boosted}_{VBF}$; else, if there is exactly one b-tagged jet present in the events, into the "1b" categories SR^{1b}_{ggF} and SR^{1b}_{VBF}; and else, into the "$\geq$2b" categories $SR^{\geq 2b}_{ggF}$ and $SR^{\geq 2b}_{VBF}$. The events in the CRs are sorted as follows. If a boosted $H \to b\bar{b}$ candidate jet is present, the events enter a dedicated boosted category; there is one single such category for all CR events owing to the low number of events. Otherwise, the events in each CR are sorted into 1b and $\geq$2b categories if there is exactly 1 or $\geq$2 b-tagged jets present, respectively. This results in 15 categories in total: three categories (boosted, 1b, $\geq$2b) in each of the two SRs (SR_{ggF}, SR_{VBF}), and two categories (1b, $\geq$2b) in each of the four CRs plus a boosted CR.

The events in the three SR_{ggF} categories are further classified by a dedicated binary NN (NN_{ggF}), which is optimised to separate ggF production HH events from the inclusive background. The output values x of the NN_{ggF} are transformed with a "logit" function $\text{logit}(x) = \ln\left(\frac{x}{1-x}\right)$, and the transformed values are used as discriminating observable in the SR_{ggF} categories in the final fit. The logit transformation helps to resolve the distribution of NN_{ggF} output values for signal events, which strongly peaks close to 1. Likewise, the events in the three SR_{VBF} categories are classified by another binary NN (NN_{VBF}) optimised to separate VBF production HH events from the background, and the output values of the NN_{VBF}, after a logit transformation, are used as final discriminating observables. The binning in the SR_{ggF} (SR_{VBF}) categories has been chosen such that the NN_{ggF} (NN_{VBF}) output distributions are uniform for ggF (VBF) signal events, and such that the uncertainty due to the limited number of simulated events is sufficiently small. In the background CRs, the yield in each category enters the final fit. The resulting discriminant distribution is shown in Fig. 4.

The analysis strategy has been optimised using simulated data to achieve the best sensitivity. For example, the inclusion of the dedicated boosted categories improves the expected limit on the SM HH production cross section by 4% and reduces the confidence interval on κ_{2V} by 26%.

4.3 Neural network architecture and training

The hyperparameters [65, 66] of the multiclassification NN_{cat} and the binary NN_{ggF} and NN_{VBF} are listed in Table 2. They have been optimised based on the achieved expected limit to SM HH production, evaluated on simulated events.

Table 2: Hyperparameters of the neural networks. Where they differ for the multiclassification NN_{cat} and the binary NN_{ggF} and NN_{VBF}, they are listed as "$NN_{cat}/NN_{ggF}/NN_{VBF}$", otherwise they are the same for all networks.

Hidden layers	3
Nodes per hidden layer	512
Activation function	ReLu
Output activation function	Softmax
Loss function	categorical crossentropy
Pre-processing	`BatchNormalization`
Dropout rate	20%
Training epochs	100
Batch size	4032/10648/15972
Steps per epoch	52/23/15
Initial learning rate	$5 \cdot 10^{-4}$
Learning rate decay	0.8
Learning rate decay patience	3 epochs

The NNs are trained using simulated signal and background events from all four analysed data-taking periods together, thereby reducing statistical fluctuations in the trained NN parameter values. The events are weighted to reflect the different integrated luminosities. It has been validated that the sensitivity of the analysis does not degrade compared to the case of training NNs separately for each data-taking period or when using information about the period of data taking as an input feature to the NNs.

The events of the different simulated events samples are effectively weighted in the training, essentially to avoid that the NN classification decision is impacted by the relative frequency of the different processes in a suboptimal way. For signal events, the weighting is chosen such that the different simulated κ_λ and κ_{2V} points have the same effective number of events in the training. For background events in the NN_{cat} training, the weighting is chosen such that each background process has the same number of effective events as the total effective number of signal events. In the binary NN trainings, the background events are weighted relative to each other by the cross section of the processes. However, a larger relative weight is assigned to the events of some of the minor background processes in order to avoid large statistical fluctuations from few events in the analysis bins with largest signal purity. Finally, the total effective number of background events is normalised to the total effective number of signal events. This scheme has been chosen in order to find a compromise between sensitivity to the SM HH signal, to κ_λ, and to κ_{2V}, evaluated by the expected limits, as well as to minimize the size of MC statistical uncertainties.

A five-fold cross-validation technique is used for the training and evaluation of the NNs. In each fold, the 80% of events used for training are further split into three independent subsamples used for the actual training (75%), for the optimisation of the hyperparameters (10%), and for validating the performance of the NNs (15%). The training is terminated after 100 epochs.

The input variables of the NNs are listed in Table 3. The same variables are used in the multiclassification and in the binary NNs, with the exception of the VBF tag information, which only enters the NN_{VBF}. The input variables have been chosen out of a larger set of possible observables as the ones that are most important for the performance of the NNs, based on sev-

eral ranking procedures including a "data Shapley" metric [67] and the expected limit on the SM HH production cross section. As an example, the four input variables with the highest importance based on the data Shapley metric are shown in Fig. 2: the approximate invariant mass and the p_T of the H boson candidate decaying to $b\bar{b}$, reconstructed as the invariant mass and p_T, respectively, of the two jets with the highest b tagging score; the approximate invariant mass of the HH system, reconstructed as the invariant mass of the two jets with the highest b tagging score, the two leptons, and p_T^{miss}; and the p_T of the jet with the highest b tagging score.

Table 3: Observables used as input variables to the NNs.

invariant mass of $b\bar{b}$ system† (m_{bb}); proxy for mass of $H \to b\bar{b}$ candidate
invariant mass of $\ell\ell$ system ($m_{\ell\ell}$)
invariant mass of the leptons and p_T^{miss}; proxy for mass of $H \to WW$ candidate
invariant mass of the b jets†, leptons, and p_T^{miss}; proxy for mass of HH system
number of jets
scalar sum of jet p_T (H_T)
p_T of jet with highest b tagging score
p_T of jet with second-highest b tagging score
p_T of leading jet that is not considered as part of the $b\bar{b}$ system†
p_T of the $b\bar{b}$ system†
sum of b tagging scores of all jets
p_T of leading lepton
p_T of subleading lepton
p_T of the $\ell\ell$ system
η of leading jet that is not considered as part of the $b\bar{b}$ system†
maximum ΔR between jets
ΔR between jets of $b\bar{b}$ system†
ΔR between leptons
ΔR between $b\bar{b}$ system† and $\ell\ell$ system
minimum ΔR between lepton and b-tagged jet
p_T^{miss}
binary tag of lepton flavour combinations: same flavour or mixed flavour
binary VBF tag based on the presence of a VBF jet pair ‡

†: two jets with highest b tagging scores; in events with only one jet, the observable is assigned a non-physical default value of -10

‡ only used in NN_{VBF}

4.4 Background modelling

All background processes are modelled using the simulated events samples described in Section 3, with the corrections and calibrations discussed in the previous sections. The overall

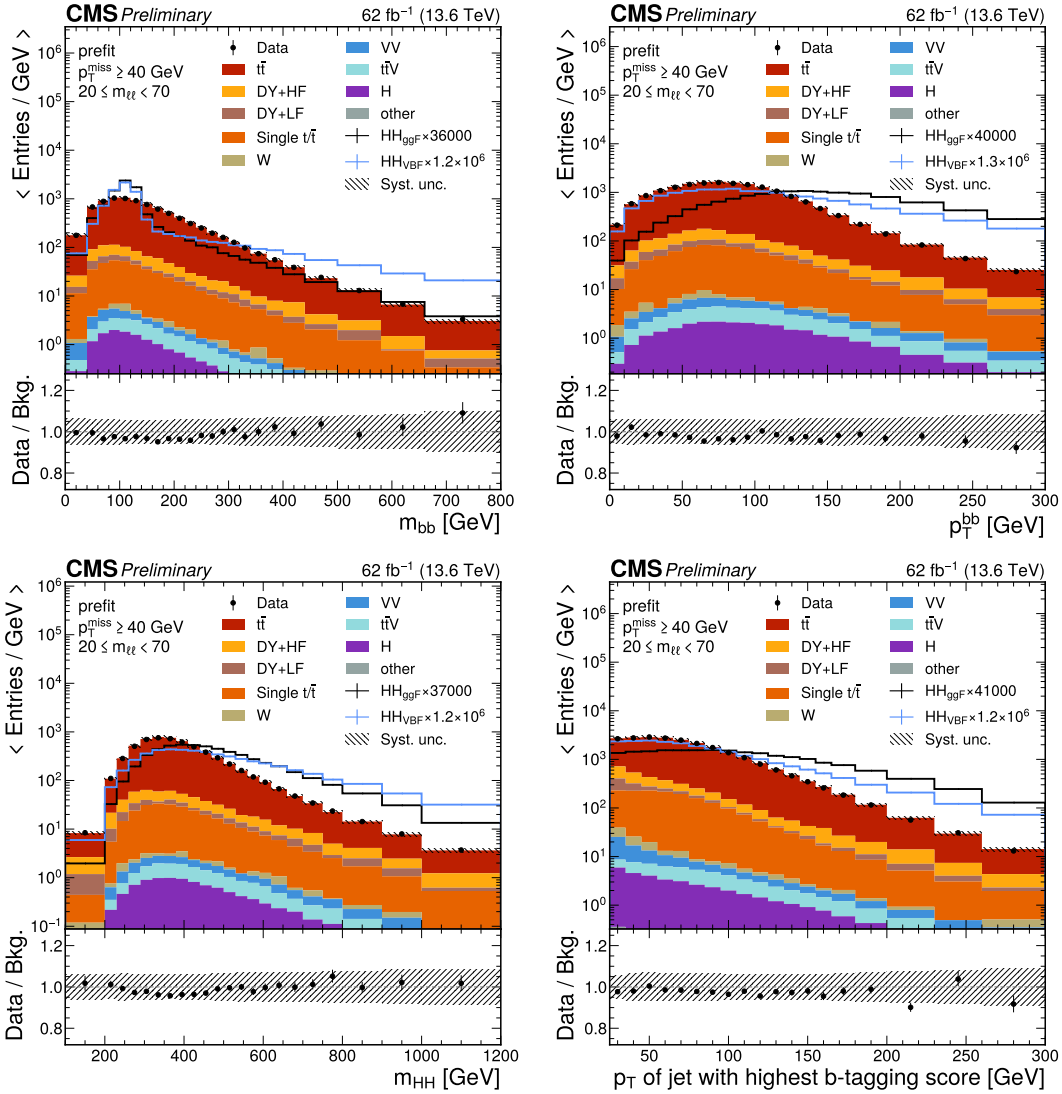

Figure 2: Approximate invariant mass (top left) and p_T (top right) of the H boson candidate decaying to $b\bar{b}$, reconstructed as the invariant mass and p_T, respectively, of the two jets with the highest b tagging score; approximate invariant mass of the HH system (bottom left), reconstructed as the invariant mass of the two jets with the highest b tagging score, the two leptons, and p_T^{miss}; and p_T of the jet with the highest b tagging score (bottom right), for events in the analysis region observed in data (markers) and predicted by the background model (stacked histograms) prior to the fit to data. The HH signal distributions in the ggF and VBF production channels as predicted in the SM, scaled to the total background yield for better visibility, are overlayed (solid lines). The uncertainty band represents the total (statistical and systematic) uncertainty. The last bin includes the overflow events.

normalisation of the $t\bar{t}$ and the DY background processes are free parameters in the final fit to data. Two separate $t\bar{t}$ normalisation parameters are assigned, one for the 1b and $\geq$2b categories and one for the boosted categories, in order to account for potential modelling differences in the two very different p_T regimes. For the DY background, two separate normalisation parameters are assigned, one to the DY + HF and one to DY + LF process, to account for flavour-dependent effects in the hadronic part of the events.

The background modelling of the dominant $t\bar{t}$ and DY background processes is verified in the validation regions with $m_{\ell\ell} > 110\,\text{GeV}$ and $70 < m_{\ell\ell} < 110\,\text{GeV}$, respectively. While the relevant observables are sufficiently well modelled for $t\bar{t}$ events, a mismodelling is observed for DY events in particular in the distribution of the jet multiplicity, N_{jets}, as well as in the p_T spectrum of the dilepton system, $p_T^{\ell\ell}$. Therefore, a dedicated correction of the DY events is derived as a function of $p_T^{\ell\ell}$ and N_{jets}, using the data in the DY validation region.

The DY correction is determined from the data in the $\mu^+\mu^-$ channel of the DY validation region, which has a purity of 94%, using the following procedure. The $p_T^{\ell\ell}$ distribution observed in data, subtracted by the expected contribution of non-DY events predicted from simulation, is divided by the $p_T^{\ell\ell}$ distribution predicted by the DY simulation, and the ratio is parametrised with an empiric function $f_{\text{DY}}^{\text{corr}}(p_T^{\ell\ell})$. Then, $f_{\text{DY}}^{\text{corr}}$ is applied as event weight to the simulated DY events in the analysis region as a function of the particle-level $p_T^{\ell\ell}$, correcting the $p_T^{\ell\ell}$ spectrum to match the data. Applying the correction as a function of the particle-level $p_T^{\ell\ell}$ avoids dilution due to the electron momentum resolution, while for the muons the effect of the p_T resolution is negligible. The correction $f_{\text{DY}}^{\text{corr}}$ is derived separately for events with $N_{\text{jets}} = 1$ and $N_{\text{jets}} \geq 2$, and for the latter, an additional normalisation factor is applied per N_{jets} bin to account for residual mismodelling of the jet multiplicity. The correction derived in the DY validation region is applicable also in the analysis region, since the relative fraction of DY + HF and DY + LF events is the same in both regions. The effect of the corrections is demonstrated in Fig. 3, which shows the $p_T^{\ell\ell}$ distribution in the DY validation region before and after application of the corrections.

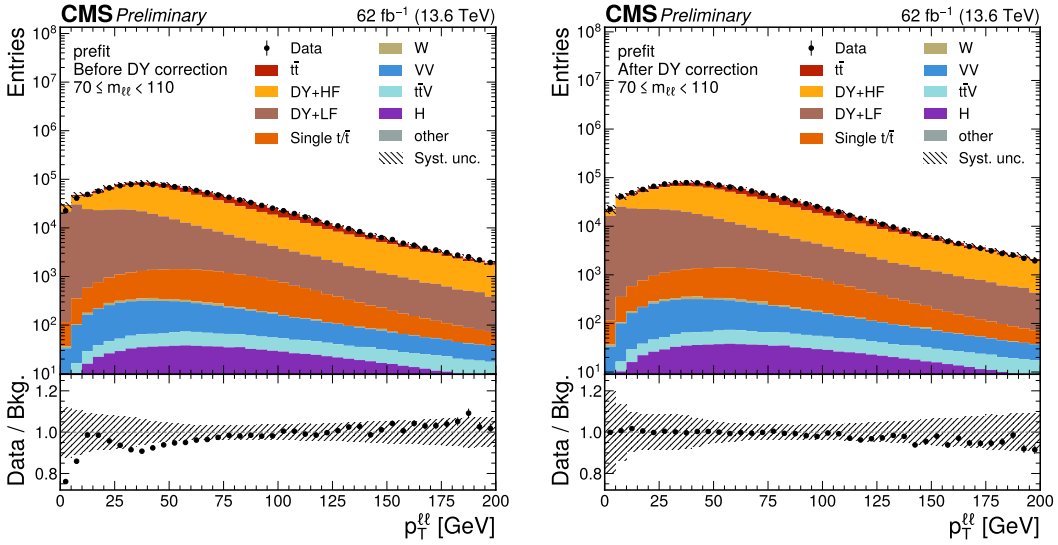

Figure 3: The p_T of the dilepton system in the DY validation region before (left) and after (right) application of the DY corrections. The error band shows the total uncertainty.

5 Systematic uncertainties

The sensitivity of the analysis is primarily limited by statistical uncertainties. However, several sources of systematic uncertainties also affect the result and have been considered as described in the following. The systematic uncertainties alter either the rate or both the rate and the shape of the discriminant distributions for the signal or background processes. Their impact on the measured signal and background yields is discussed in Section 6.

Theoretical uncertainties are treated as fully correlated among the different data-taking periods, while experimental uncertainties are in most cases treated as uncorrelated between the periods. The latter assumption is motivated by the fact that the experimental uncertainties are mainly of statistical origin related to the limited size of the data and simulation samples used in auxiliary measurements, which are independent across the data-taking periods, and that the detector conditions changed between the data-taking periods.

Uncertainties in the inclusive cross sections used to predict the rate of the signal and background processes arise primarily from variations of the renormalisation and factorisation scales and the PDFs, and are propagated to the yield estimates. The cross section uncertainties are each separated into their scale ($\sigma(\mu_R/\mu_F)$) and PDF ($\sigma(\text{PDF} + \alpha_S)$) components, and are correlated where appropriate among processes. Since the normalisation of the $t\bar{t}$ and DY background processes are measured in-situ, they are not affected by the uncertainties of the inclusive cross section predictions. For the signal, the uncertainties on the inclusive cross section prediction generally affect the result since the measured rate is normalised to the SM expectation.

Uncertainties in the simulation of the signal and background events are evaluated as follows. Uncertainties arising from missing higher-order terms at the ME level are evaluated from variations of the renormalisation and factorisation scales by factors of 2 and 0.5 with respect to the nominal values. All possible permutations of independent scale variations ($\mu_{R/F}$) are performed, excluding the $\mu_R \cdot \mu_F = 1$ cases, and the envelope of the six obtained variations in the fitted distributions is assigned as uncertainty [68]. The uncertainty originating from the PDF set (PDF shape) is determined from the PDF variations provided with the NNPDF3.1 set [42], following the procedure described in Ref. [69], correlating processes for which the same flavour scheme and order in the strong coupling constant α_S are used in the PDF set. The uncertainty in the PYTHIA PS is determined by varying the parameters controlling the amount of initial- and final-state radiation independently by factors of 2 and 0.5 with respect to their nominal value (PS scale ISR/FSR) [43]. The ISR uncertainties are treated as correlated among the processes, while the FSR uncertainties are treated as uncorrelated between the processes. Uncertainties in the corrections applied to the p_T spectrum of the t quarks in the $t\bar{t}$ simulation [48] are also considered; they amount up to 7% of the expected yield depending on the p_T of the top quarks.

The variations associated with the theory modelling uncertainties described in the previous paragraph generally affect the overall normalisation, the simulated acceptance, and the shapes of the fitted distributions. Since the normalisation of the signal as well as the $t\bar{t}$, DY + HF, and DY + LF background processes are determined in the fit, the impact of these variations on the overall normalisation before any selection is removed. In case the result is quoted relative to the SM prediction, the uncertainty on the inclusive signal cross section is considered for the prediction, though. Likewise, for the minor background processes, only the uncertainty in the inclusive cross section prediction is retained to avoid double counting of normalisation uncertainties.

Effects of the uncertainty in the distribution of the number of pileup interactions are evaluated

by varying the total inelastic cross section used to predict the number of pileup interactions in the simulated events by $\pm 4.6\%$ from its nominal value [26]. The uncertainties in the integrated luminosity used to normalize the predicted signal and background yields are 1.4% for 2022 [70] and 1.3% for 2023 [71].

The trigger efficiency is measured in an unbiased sample of events, mostly $\mathrm{t\bar{t}}$ leptonic decays, that contain two leptons and fulfil the offline event selection criteria but were collected with triggers based on $p_{\mathrm{T}}^{\mathrm{miss}}$, which feature a negligible correlation with the triggers used in this analysis. Corrections are derived to match the efficiency in simulation to that observed in data, with uncertainties of up to 5%, dominated by statistical fluctuations in the data samples used in the measurement.

Residual differences between the lepton reconstruction, identification, and isolation efficiencies in data and simulation are corrected based on efficiency measurements using a "tag-and-probe" method in data samples of $\mathrm{Z} \to \ell\ell$ events [23, 24]. The uncertainties (Lepton Reco/ID/Iso) are smaller than 1 and 2% for the muon and electron correction factors, respectively.

The uncertainties in the jet energy scale and resolution are propagated by varying the corrections applied to all jets in the signal and background simulations within their uncertainty [30]. The uncertainties in the corrections of the b jet (mis)identification probabilities in simulated events [34, 35], which can arise from the contamination of background processes in the data samples used for their derivation, are considered separately for light jets and b/c jets, and vary between 5 and 30%, depending on the jet p_{T}. An uncertainty of 30% is assigned to account for mismodelling in the boosted $\mathrm{H} \to \mathrm{b\bar{b}}$ candidate jet tagging efficiency, separately to signal and background contributions in the boosted ggF and VBF SRs and the boosted CR.

Uncertainties in the data-driven DY correction are evaluated as half of the difference between the nominal and the corrected distributions, and applied symmetrically to the corrected distributions.

The impact of statistical fluctuations in the signal and background prediction due to the limited number of simulated events (size of the MC samples) is accounted for using the Barlow–Beeston-lite approach [72].

6 Results

A simultaneous binned profile likelihood fit of the signal and background templates to the observed discriminant distributions in all channels and categories is performed, using the COMBINE program [73]. The rates of the $\mathrm{t\bar{t}}$ and DY background processes described in Section 4.4 are free parameters in the fit, and the systematic uncertainties described in Section 5 are taken into account via nuisance parameters in the likelihood function. Log-normal distributions are used to model the effects of nuisance parameters on the overall rate, while nuisance parameters that affect the shape of the distributions are parametrised using an interpolation between alternative templates [72, 73]. The observed yields in each bin of the final discriminant distributions in all channels and categories are shown in Fig. 4.

The goodness-of-fit is quantified using a p-value that takes into account the postfit uncertainty model [73] and amounts to $p = 0.20$, indicating good description of the data by the fit model. The best-fit value of the $\mathrm{t\bar{t}}$ background normalisation parameter is $0.95^{+0.03}_{-0.02}$ ($0.79^{+0.22}_{-0.17}$ in the boosted categories).

No significant excess of events over the background-only hypothesis is observed. Upper limits

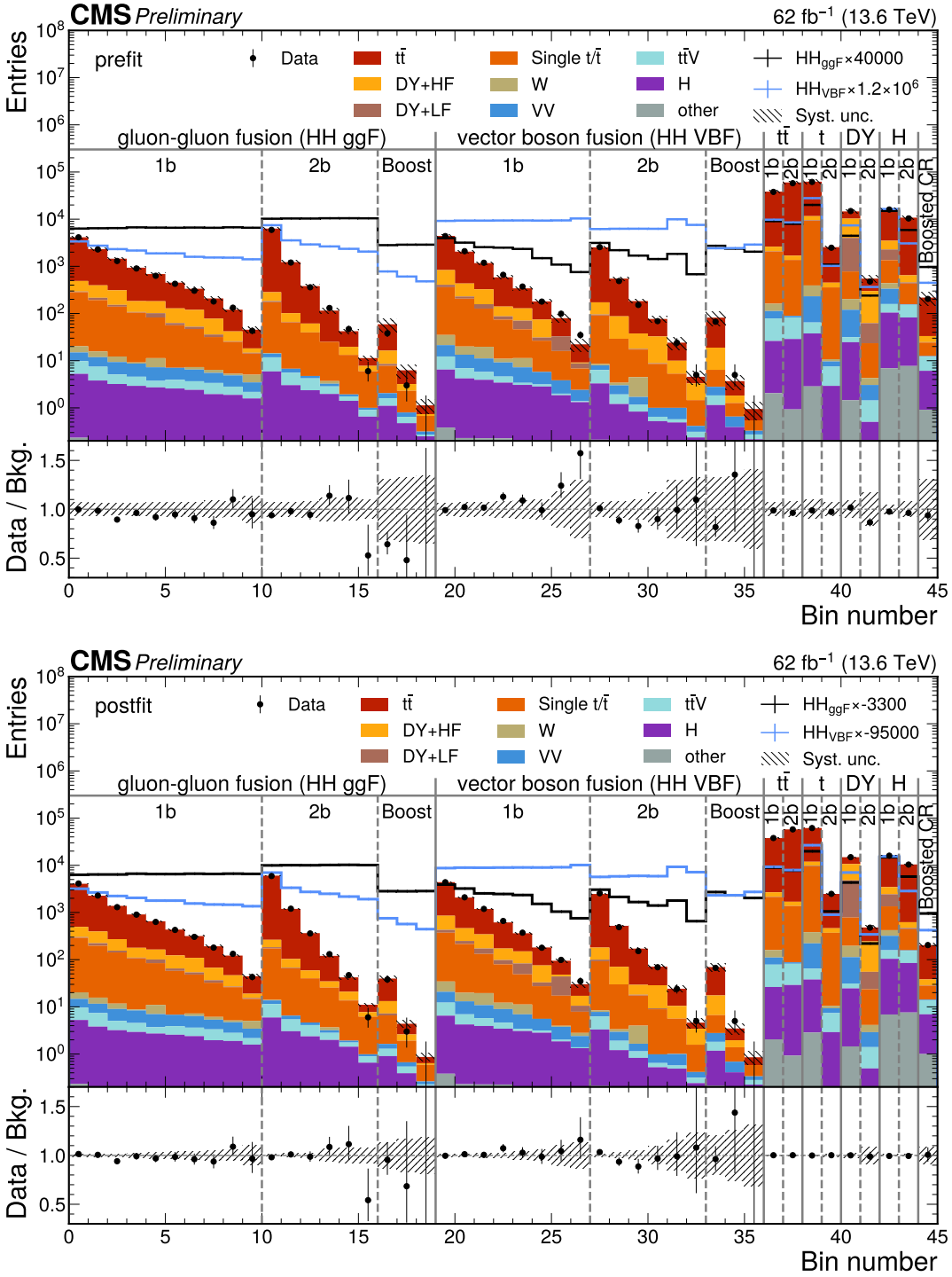

Figure 4: Observed (points) and expected (filled histograms) yields in each discriminant (NN score or category yield) bin before (top) and after (bottom) the fit to data. The HH signal distributions in the ggF and VBF production channels, scaled to the total background yield for better visibility, are overlayed (solid lines). The uncertainty bands include the total uncertainty of the fit model. The lower pads show the ratio of the data to the to the expected background yields.

on the cross section of non-resonant Higgs boson pair production are set at 95% CL, using the CL_s criterion [74, 75] and assuming asymptotic distributions [76] of the profile likelihood ratio test statistics defined in Ref. [73]. The observed (median expected) upper limit on the inclusive HH production cross section is 12.7 (18.6) times the value expected in the SM. The limit is also shown as a function of κ_λ and as a function of κ_{2V} in Fig. 5, where in each case all other couplings are assumed to conform to the SM prediction.

In Fig. 6, the observed and expected profile likelihood ratio values are presented as a function of κ_λ and κ_{2V}, where again all other couplings are assumed as in the SM. Values of κ_λ outside $[-9.7, 15.8]$ ($[-13.4, 19.9]$ expected) and of κ_{2V} outside $[-0.27, 2.32]$ ($[-0.58, 2.64]$ expected) are excluded at 95% CL. The likelihood contours from a simultaneous fit of κ_λ and κ_{2V} are shown in Fig. 7.

The by-far largest uncertainty source of the analysis is the limited statistical precision due to the small signal rate. The impact of the statistical uncertainties compared to the systematic effects is shown in Fig. 6.

The best-fit values of the nuisance parameters used to describe the systematic uncertainties are shown in Fig. 8; they are well within 1 standard deviation of the prior uncertainty. The impact of each normalisation and nuisance parameter on the result is evaluated as the difference of the nominal best-fit value of the HH signal strength modifier and the best-fit value obtained when fixing the parameter under scrutiny to its best-fit value plus/minus its postfit uncertainty. The best-fit values and impacts of the 25 parameters with the highest impact are shown in Fig. 8. They are related to the boosted $H \to b\bar{b}$ candidate tagging uncertainty, the $t\bar{t}$ normalisation in the boosted categories, and the uncertainty on the inclusive HH cross section prediction (which enters the measured signal strength modifier in the denominator).

7 Summary

A search has been presented for non-resonant Higgs boson pair production in the $b\bar{b}WW$ decay channel with two leptons in the final state, conducted with $62\,\text{fb}^{-1}$ of proton-proton collision data.

The data are consistent with standard model predictions. An upper limit is set on the Higgs boson pair production cross section of 12.7 times the standard model prediction at 95% confidence level, with an expectation of 18.6. Compared to a previous search by the CMS Collaboration with $138\,\text{fb}^{-1}$ of Run 2 data in the same channel, significant improvements in sensitivity have been achieved, owing to a refined classification strategy, employment of additional triggers as well as an improved b tagging algorithm, leading to 30% better expected sensitivity despite the smaller analysed dataset.

The cross section limit is further used to constrain the trilinear and quartic coupling of the Higgs boson between $[-9.7, 15.8]$ ($[-13.4, 19.9]$ expected) and $[-0.27, 2.32]$ ($[-0.58, 2.64]$ expected), respectively, at 95% CL

The presented search is the first result with $\sqrt{s} = 13.6\,\text{TeV}$ data in this channel.

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, doi:10.1016/j.physletb.2012.08.020, arXiv:1207.7214.

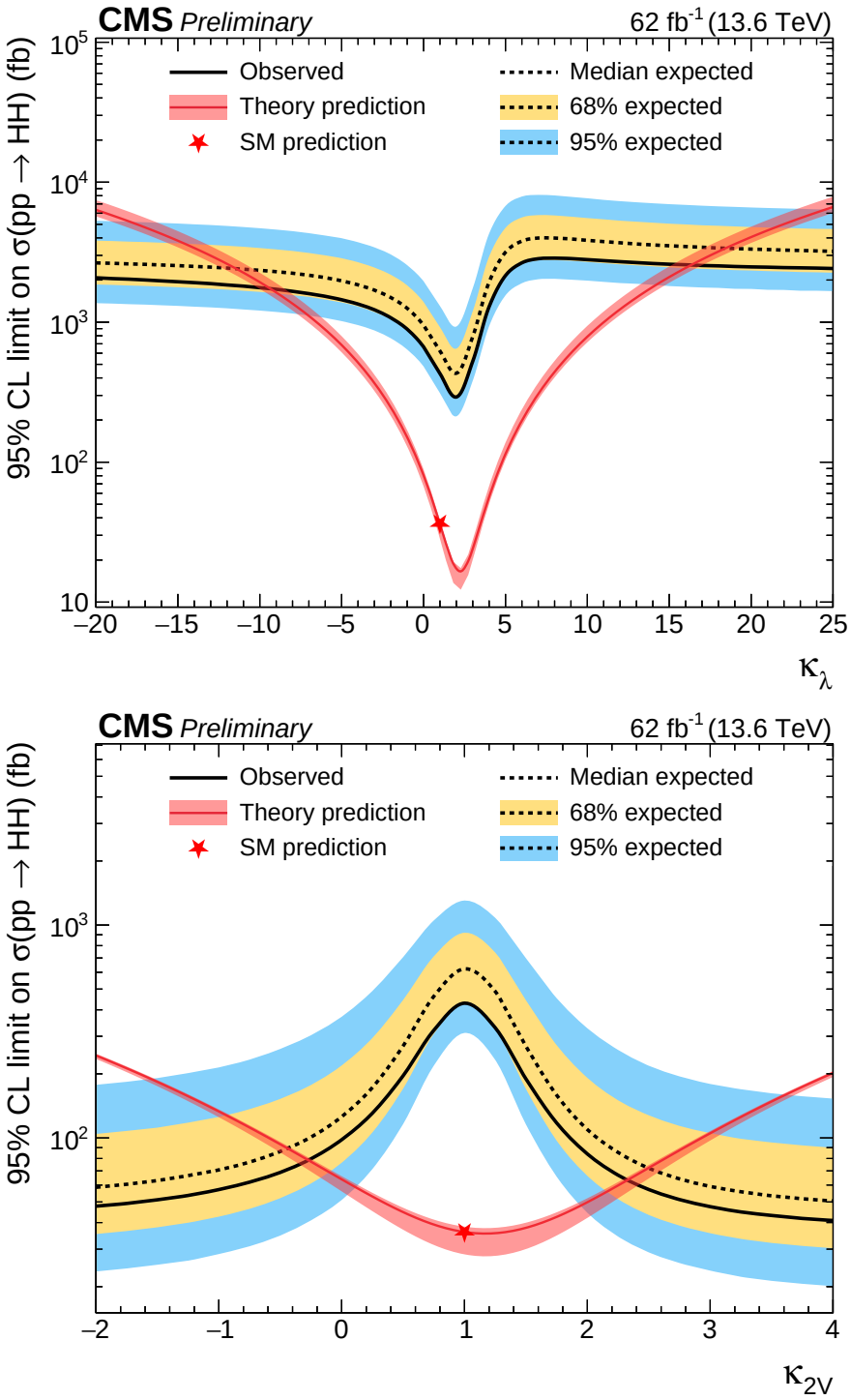

Figure 5: Observed (solid black line) and median expected (dashed black line) upper limits at 95% CL on the inclusive HH production cross section as a function of κ_λ (top) and as a function of κ_{2V} (bottom); in both cases, all respective other couplings are fixed to the SM prediction. The yellow (blue) bands show the 68% (95%) confidence level intervals of the expected limit. The predicted cross section is overlayed (red curve), and the SM prediction is indicated (red star).

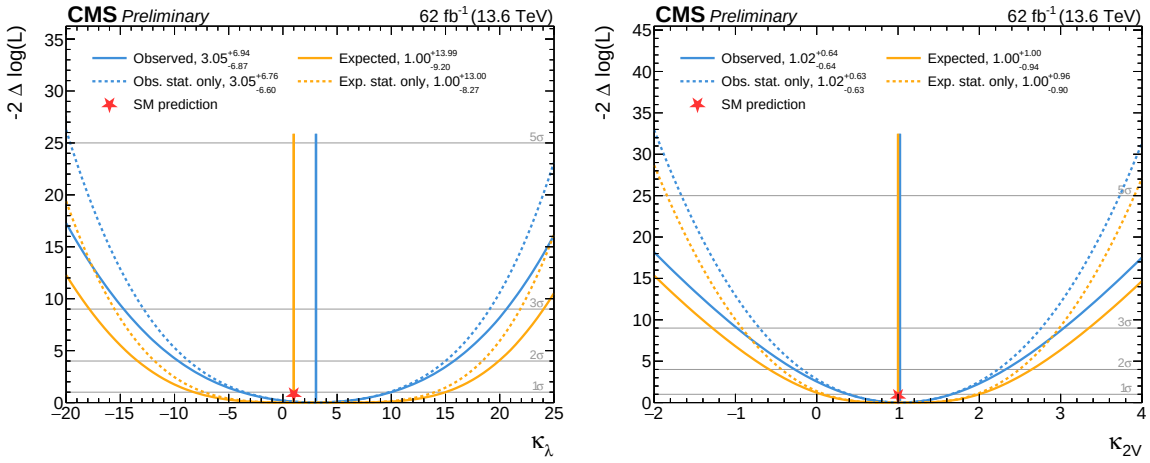

Figure 6: Observed (blue) and expected (orange) negative log-likelihood values as a function of κ_λ (left) and κ_{2V} (right), assuming all other couplings conform to the SM prediction. The solid line includes the full uncertainty model, while the dashed line only includes statistical uncertainties. The vertical lines indicate the best-fit values and the SM prediction (red star).

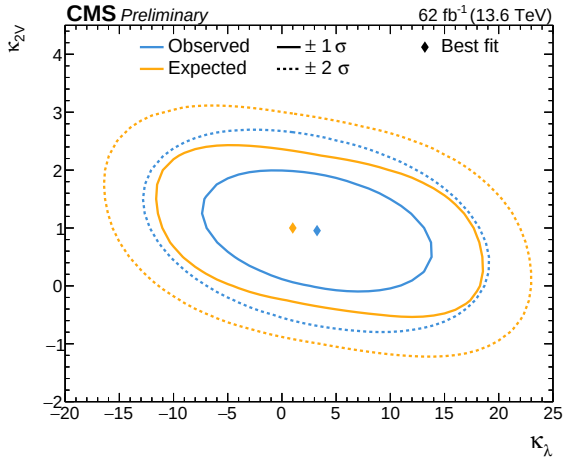

Figure 7: Observed (blue) and expected (orange) negative log-likelihood contours as a function of κ_λ and κ_{2V}, assuming all other couplings conform to the SM prediction. Shown are the best-fit points (marker) and the 68% (solide lines) and 95% (dashed lines) CL contours.

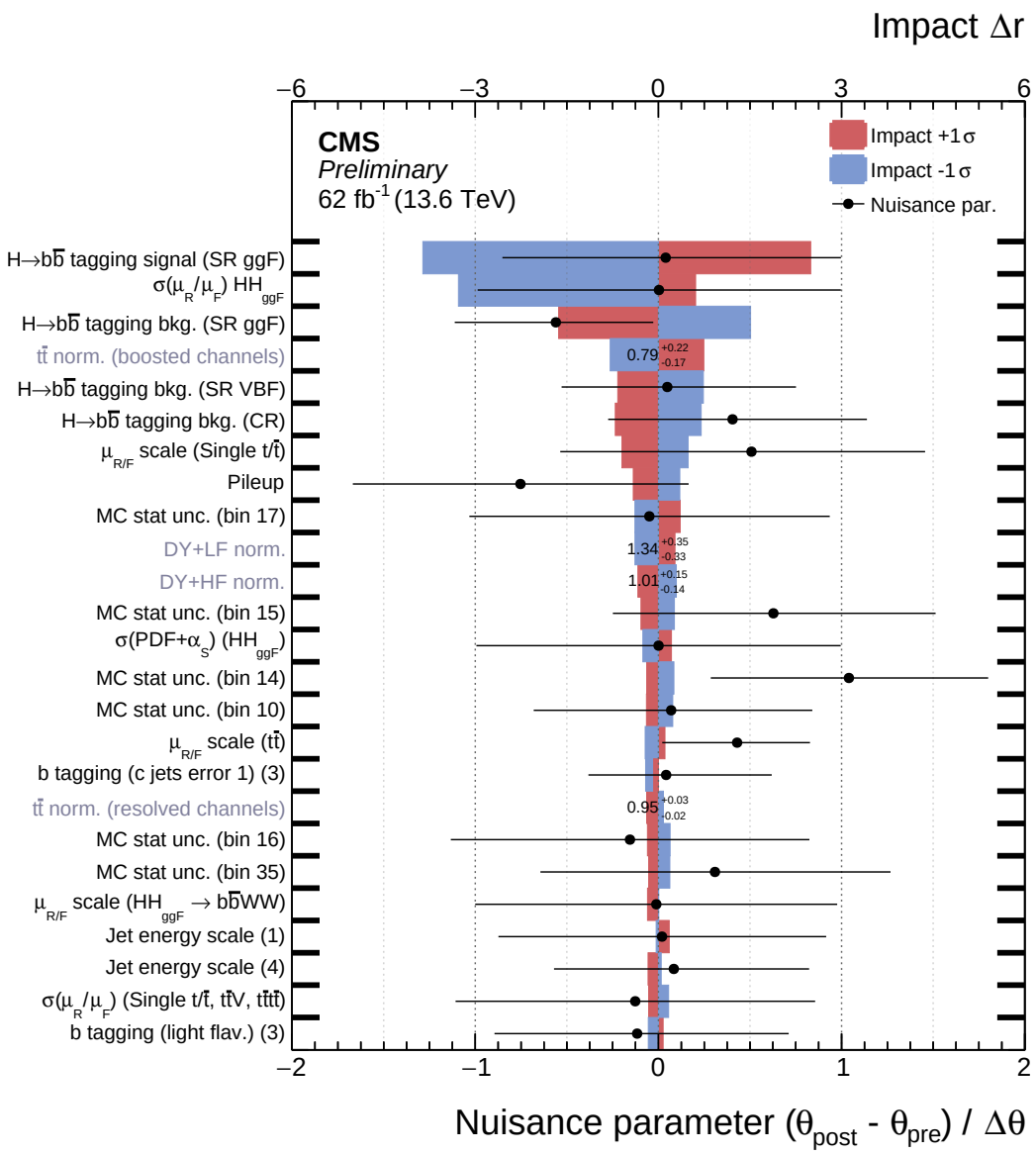

Figure 8: Best fit values of the background normalisation and nuisance parameters (black markers). The nuisance parameter values are shown as the difference of their best-fit values, θ_{post}, and prefit values, θ_{pre}, relative to the prefit uncertainties $\Delta\theta$. The impact (coloured areas) of the nuisance parameters on the HH signal strength is computed as the difference of the nominal best-fit value of the signal strength and the best-fit value obtained when fixing the nuisance parameter under scrutiny to its best-fit value θ_{post} plus/minus its postfit uncertainty. The nuisance parameters are ordered by their impact, and only the 25 highest ranked parameters are shown. The number in parentheses for the jet energy scale and b tagging uncertainties correspond to a numbering of the data-taking period to which they are associated. The MC stat unc. refers to the systematic uncertainty due to the limited number of simulated events; in this case, the number in parentheses refers to the bin numbers shown in Fig. 4.

[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, doi:10.1016/j.physletb.2012.08.021, arXiv:1207.7235.

[3] CMS Collaboration, "Observation of a new boson with mass near 125 GeV in pp collisions at $\sqrt{s} = 7$ and 8 TeV", *JHEP* **06** (2013) 081, doi:10.1007/JHEP06(2013)081, arXiv:1303.4571.

[4] ATLAS Collaboration, "A detailed map of Higgs boson interactions by the ATLAS experiment ten years after the discovery", *Nature* **607** (2022) 52, doi:10.1038/s41586-022-04893-w, arXiv:2207.00092. [Erratum: doi:10.1038/s41586-022-05581-5].

[5] CMS Collaboration, "A portrait of the Higgs boson by the CMS experiment ten years after the discovery", *Nature* **607** (2022) 60, doi:10.1038/s41586-022-04892-x, arXiv:2207.00043.

[6] M. Grazzini et al., "Higgs boson pair production at NNLO with top quark mass effects", *JHEP* **05** (2018) 059, doi:10.1007/JHEP05(2018)059, arXiv:1803.02463.

[7] A. Karlberg et al., "Ad interim recommendations for the Higgs boson production cross sections at $\sqrt{s} = 13.6$ TeV", 2024. arXiv:2402.09955.

[8] J. Baglio et al., "gg $\to$ HH: Combined uncertainties", *Phys. Rev. D* **103** (2021) 056002, doi:10.1103/PhysRevD.103.056002, arXiv:2008.11626.

[9] F. A. Dreyer, A. Karlberg, J.-N. Lang, and M. Pellen, "Precise predictions for double-Higgs production via vector-boson fusion", *Eur. Phys. J. C* **80** (2020) 1037, doi:10.1140/epjc/s10052-020-08610-7, arXiv:2005.13341.

[10] F. A. Dreyer and A. Karlberg, "Vector-boson fusion Higgs pair production at N^{3}LO", *Phys. Rev. D* **98** (2018) 114016, doi:10.1103/PhysRevD.98.114016, arXiv:1811.07906.

[11] ATLAS Collaboration, "Combination of searches for Higgs boson pair production in pp collisions at $\sqrt{s} = 13$ TeV with the ATLAS detector", *Phys. Rev. Lett.* **133** (2024) 101801, doi:10.1103/PhysRevLett.133.101801, arXiv:2406.09971.

[12] CMS Collaboration, "Combination of searches for nonresonant Higgs boson pair production in proton-proton collisions at $\sqrt{s} = 13$ TeV", 2025. arXiv:2510.07527. Submitted to Reports on Progress in Physics.

[13] ATLAS and CMS Collaborations, "Combination of ATLAS and CMS searches for Higgs boson pair production at 13 TeV", ATLAS Conference Note ATLAS-CONF-2025-012, CMS Physics Analysis Summary CMS-PAS-HIG-25-014, 2025.

[14] CMS Collaboration, "Search for HH production decaying into two b quarks and two photons in pp collisions at 13.6 TeV with a partial CMS Run 3 dataset", CMS Physics Analysis Summary CMS-PAS-HIG-25-007, 2025.

[15] CMS Collaboration, "Improved results on Higgs boson pair production in the 4b final state", CMS Physics Analysis Summary CMS-PAS-HIG-24-010, 2025.

[16] ATLAS Collaboration, "Study of Higgs boson pair production in the $HH \to b\bar{b}\gamma\gamma$ final state with 308 fb^{-1} of data collected at $\sqrt{s} = 13$ TeV and 13.6 TeV by the ATLAS experiment", 2025. arXiv:2507.03495. Submitted to Phys. Lett. B.

[17] CMS Collaboration, "Search for Higgs boson pair production in the $\text{b}\bar{\text{b}}\text{W}^+\text{W}^-$ decay mode in proton-proton collisions at $\sqrt{s} = 13$ TeV", *JHEP* **07** (2024) 293, doi:10.1007/JHEP07(2024)293, arXiv:2403.09430.

[18] ATLAS Collaboration, "Search for non-resonant Higgs boson pair production in the $2b + 2\ell + E_{\text{T}}^{\text{miss}}$ final state in pp collisions at $\sqrt{s} = 13$ TeV with the ATLAS detector", *JHEP* **02** (2024) 037, doi:10.1007/JHEP02(2024)037, arXiv:2310.11286.

[19] LHC Higgs Cross Section Working Group, "Handbook of LHC Higgs cross sections: 3. Higgs properties", 2013. arXiv:1307.1347. doi:10.5170/CERN-2013-004.

[20] CMS Collaboration, "Particle-flow reconstruction and global event description with the CMS detector", *JINST* **12** (2017) P10003, doi:10.1088/1748-0221/12/10/P10003, arXiv:1706.04965.

[21] CMS Collaboration, "The CMS experiment at the CERN LHC", *JINST* **3** (2008) S08004, doi:10.1088/1748-0221/3/08/S08004.

[22] CMS Collaboration, "Development of the CMS detector for the CERN LHC Run 3", *JINST* **19** (2024) P05064, doi:10.1088/1748-0221/19/05/P05064.

[23] CMS Collaboration, "Performance of the CMS muon detector and muon reconstruction with proton-proton collisions at $\sqrt{s} = 13$ TeV", *JINST* **13** (2018) P06015, doi:10.1088/1748-0221/13/06/P06015, arXiv:1804.04528.

[24] CMS Collaboration, "Electron and photon reconstruction and identification with the CMS experiment at the CERN LHC", *JINST* **16** (2021) P05014, doi:10.1088/1748-0221/16/05/P05014, arXiv:2012.06888.

[25] CMS Collaboration, "ECAL 2016 refined calibration and Run 2 summary plots", CMS Detector Performance Summary CMS-DP-2020-021, 2020.

[26] CMS Collaboration, "Pileup mitigation at CMS in 13 TeV data", *JINST* **15** (2020) P09018, doi:10.1088/1748-0221/15/09/P09018, arXiv:2003.00503.

[27] M. Cacciari, G. P. Salam, and G. Soyez, "The anti-k_{T} jet clustering algorithm", *JHEP* **04** (2008) 063, doi:10.1088/1126-6708/2008/04/063, arXiv:0802.1189.

[28] M. Cacciari, G. P. Salam, and G. Soyez, "FastJet user manual", *Eur. Phys. J. C* **72** (2012) 1896, doi:10.1140/epjc/s10052-012-1896-2, arXiv:1111.6097.

[29] D. Bertolini, P. Harris, M. Low, and N. Tran, "Pileup per particle identification", *JHEP* **10** (2014) 059, doi:10.1007/JHEP10(2014)059, arXiv:1407.6013.

[30] CMS Collaboration, "Jet energy scale and resolution in the CMS experiment in pp collisions at 8 TeV", *JINST* **12** (2017) P02014, doi:10.1088/1748-0221/12/02/P02014, arXiv:1607.03663.

[31] H. Qu and L. Gouskos, "ParticleNet: Jet tagging via particle clouds", *Phys. Rev. D* **101** (2020) 056019, doi:10.1103/PhysRevD.101.056019, arXiv:1902.08570.

[32] CMS Collaboration, "Run 3 commissioning results of heavy-flavor jet tagging at $\sqrt{s} = 13.6$ TeV with CMS data using a modern framework for data processing", CMS Detector Performance Summary CMS-DP-2024-024, 2024.

[33] CMS Collaboration, "Performance of heavy-flavour jet identification in boosted topologies in proton-proton collisions at 13 TeV", CMS Physics Analysis Summary CMS-PAS-BTV-22-001, 2023.

[34] CMS Collaboration, "Identification of heavy-flavour jets with the CMS detector in pp collisions at 13 TeV", *JINST* **13** (2018) P05011, doi:10.1088/1748-0221/13/05/P05011, arXiv:1712.07158.

[35] CMS Collaboration, "Performance summary of AK4 jet b tagging with data from 2022 proton-proton collisions at $\sqrt{s} = 13.6$ TeV with the CMS detector", CMS Detector Performance Summary CMS-DP-2024-025, 2024.

[36] P. Nason, "A new method for combining NLO QCD with shower Monte Carlo algorithms", *JHEP* **11** (2004) 040, doi:10.1088/1126-6708/2004/11/040, arXiv:hep-ph/0409146.

[37] S. Frixione, P. Nason, and C. Oleari, "Matching NLO QCD computations with parton shower simulations: The POWHEG method", *JHEP* **11** (2007) 070, doi:10.1088/1126-6708/2007/11/070, arXiv:0709.2092.

[38] 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, doi:10.1007/JHEP06(2010)043, arXiv:1002.2581.

[39] T. Ježo and P. Nason, "On the treatment of resonances in next-to-leading order calculations matched to a parton shower", *JHEP* **12** (2015) 065, doi:10.1007/JHEP12(2015)065, arXiv:1509.09071.

[40] 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, doi:10.1007/JHEP07(2014)079, arXiv:1405.0301.

[41] GEANT4 Collaboration, "GEANT4—a simulation toolkit", *Nucl. Instrum. Meth. A* **506** (2003) 250, doi:10.1016/S0168-9002(03)01368-8.

[42] NNPDF Collaboration, "Parton distributions from high-precision collider data", *Eur. Phys. J. C* **77** (2017) 663, doi:10.1140/epjc/s10052-017-5199-5, arXiv:1706.00428.

[43] C. Bierlich et al., "A comprehensive guide to the physics and usage of PYTHIA 8.3", *SciPost Phys. Codeb.* **2022** (2022) 8, doi:10.21468/SciPostPhysCodeb.8, arXiv:2203.11601.

[44] CMS Collaboration, "Extraction and validation of a new set of CMS PYTHIA 8 tunes from underlying-event measurements", *Eur. Phys. J. C* **80** (2020) 4, doi:10.1140/epjc/s10052-019-7499-4, arXiv:1903.12179.

[45] G. Heinrich et al., "NLO predictions for Higgs boson pair production with full top quark mass dependence matched to parton showers", *JHEP* **08** (2017) 088, doi:10.1007/JHEP08(2017)088, arXiv:1703.09252.

[46] G. Heinrich et al., "Probing the trilinear Higgs boson coupling in di-Higgs production at NLO QCD including parton shower effects", *JHEP* **06** (2019) 066, doi:10.1007/JHEP06(2019)066, arXiv:1903.08137.

[47] S. Frixione, P. Nason, and G. Ridolfi, "A positive-weight next-to-leading-order Monte Carlo for heavy flavour hadroproduction", *JHEP* **09** (2007) 126, doi:10.1088/1126-6708/2007/09/126, arXiv:0707.3088.

[48] CMS Collaboration, "Measurement of differential cross sections for the production of top quark pairs and of additional jets in lepton+jets events from pp collisions at $\sqrt{s} = 13\,\text{TeV}$", *Phys. Rev. D* **97** (2018) 112003, doi:10.1103/PhysRevD.97.112003, arXiv:1803.08856.

[49] CMS Collaboration, "Differential cross section measurements for the production of top quark pairs and of additional jets using dilepton events from pp collisions at $\sqrt{s} = 13\,\text{TeV}$", *JHEP* **02** (2025) 064, doi:10.1007/JHEP02(2025)064, arXiv:2402.08486.

[50] M. Czakon et al., "Top-pair production at the LHC through NNLO QCD and NLO EW", *JHEP* **10** (2017) 186, doi:10.1007/JHEP10(2017)186, arXiv:1705.04105.

[51] CMS Collaboration, "Measurement of inclusive and differential cross sections of single top quark production in association with a W boson in proton-proton collisions at $\sqrt{s} = 13.6\,\text{TeV}$", *JHEP* **01** (2025) 107, doi:10.1007/JHEP01(2025)107, arXiv:2409.06444.

[52] M. Beneke, P. Falgari, S. Klein, and C. Schwinn, "Hadronic top-quark pair production with NNLL threshold resummation", *Nucl. Phys. B* **855** (2012) 695, doi:10.1016/j.nuclphysb.2011.10.021, arXiv:1109.1536.

[53] M. Cacciari et al., "Top-pair production at hadron colliders with next-to-next-to-leading logarithmic soft-gluon resummation", *Phys. Lett. B* **710** (2012) 612, doi:10.1016/j.physletb.2012.03.013, arXiv:1111.5869.

[54] P. Bärnreuther, M. Czakon, and A. Mitov, "Percent-level-precision physics at the Tevatron: Next-to-next-to-leading order QCD corrections to $q\bar{q} \to t\bar{t}$+X", *Phys. Rev. Lett.* **109** (2012) 132001, doi:10.1103/PhysRevLett.109.132001, arXiv:1204.5201.

[55] M. Czakon and A. Mitov, "NNLO corrections to top-pair production at hadron colliders: The all-fermionic scattering channels", *JHEP* **12** (2012) 054, doi:10.1007/JHEP12(2012)054, arXiv:1207.0236.

[56] M. Czakon and A. Mitov, "NNLO corrections to top pair production at hadron colliders: The quark-gluon reaction", *JHEP* **01** (2013) 080, doi:10.1007/JHEP01(2013)080, arXiv:1210.6832.

[57] M. Czakon, P. Fiedler, and A. Mitov, "Total top-quark pair-production cross section at hadron colliders through $\mathcal{O}(\alpha_s^4)$", *Phys. Rev. Lett.* **110** (2013) 252004, doi:10.1103/PhysRevLett.110.252004, arXiv:1303.6254.

[58] 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, doi:10.1016/j.cpc.2014.06.021, arXiv:1112.5675.

[59] S. Alioli, P. Nason, C. Oleari, and E. Re, "NLO single-top production matched with shower in POWHEG: s- and t-channel contributions", *JHEP* **09** (2009) 111, doi:10.1088/1126-6708/2009/09/111, arXiv:0907.4076. [Erratum: doi:10.1007/JHEP02(2010)011].

[60] E. Re, "Single-top Wt-channel production matched with parton showers using the POWHEG method", *Eur. Phys. J. C* **71** (2011) 1547, doi:10.1140/epjc/s10052-011-1547-z, arXiv:1009.2450.

[61] M. Aliev et al., "HATHOR: HAdronic Top and Heavy quarks crOss section calculatoR", *Comput. Phys. Commun.* **182** (2011) 1034, doi:10.1016/j.cpc.2010.12.040, arXiv:1007.1327.

[62] P. Kant et al., "HatHor for single top-quark production: Updated predictions and uncertainty estimates for single top-quark production in hadronic collisions", *Comput. Phys. Commun.* **191** (2015) 74, doi:10.1016/j.cpc.2015.02.001, arXiv:1406.4403.

[63] N. Kidonakis, "Two-loop soft anomalous dimensions for single top quark associated production with W^- or H^-", *Phys. Rev. D* **82** (2010) 054018, doi:10.1103/PhysRevD.82.054018, arXiv:1005.4451.

[64] R. Frederix and S. Frixione, "Merging meets matching in MC@NLO", *JHEP* **12** (2012) 061, doi:10.1007/JHEP12(2012)061, arXiv:1209.6215.

[65] M. Erdmann, J. Glombitza, G. Kasieczka, and U. Klemradt, "Deep Learning for Physics Research". WORLD SCIENTIFIC, 2021. doi:10.1142/12294.

[66] F. Chollet et al., "KERAS". https://keras.io, 2015.

[67] A. Ghorbani and J. Zou, "Data Shapley: Equitable valuation of data for machine learning", 2019. arXiv:1904.02868. doi:10.48550/arXiv.1904.02868.

[68] M. Cacciari et al., "The t anti-t cross-section at 1.8-TeV and 1.96-TeV: A study of the systematics due to parton densities and scale dependence", *JHEP* **04** (2004) 068, doi:10.1088/1126-6708/2004/04/068, arXiv:hep-ph/0303085.

[69] J. Butterworth et al., "PDF4LHC recommendations for LHC Run II", *J. Phys. G* **43** (2016) 023001, doi:10.1088/0954-3899/43/2/023001, arXiv:1510.03865.

[70] CMS Collaboration, "Luminosity measurement in proton-proton collisions at 13.6 TeV in 2022 at CMS", CMS Physics Analysis Summary CMS-PAS-LUM-22-001, 2024.

[71] CMS Collaboration, "Measurement of the offline integrated luminosity for the CMS proton-proton collision dataset recorded in 2023", CMS Detector Performance Summary CMS-DP-2024-068, 2024.

[72] J. S. Conway, "Incorporating nuisance parameters in likelihoods for multisource spectra", in *PHYSTAT (2011): Workshop on statistical issues related to discovery claims in search experiments and unfolding*, p. 115. 2011. arXiv:1103.0354. doi:10.5170/CERN-2011-006.115.

[73] CMS Collaboration, "The CMS statistical analysis and combination tool: COMBINE", *Comput. Softw. Big Sci.* **8** (2024) 19, doi:10.1007/s41781-024-00121-4, arXiv:2404.06614.

[74] T. Junk, "Confidence level computation for combining searches with small statistics", *Nucl. Instrum. Meth. A* **434** (1999) 435, doi:10.1016/S0168-9002(99)00498-2, arXiv:hep-ex/9902006.

[75] A. L. Read, "Presentation of search results: The CL_s technique", *J. Phys. G* **28** (2002) 2693, doi:10.1088/0954-3899/28/10/313.

[76] 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, doi:10.1140/epjc/s10052-011-1554-0, arXiv:1007.1727. [Erratum: doi:10.1140/epjc/s10052-013-2501-z].