CMS Physics Analysis Summary

Contact: cms-pag-conveners-higgs@cern.ch 2020/07/29

Measurement of Higgs boson decay to a pair of muons in proton-proton collisions at $\sqrt{s} = 13\,\text{TeV}$

The CMS Collaboration

Abstract

A measurement of the Higgs boson decay to a pair of muons is presented. This result combines searches in four exclusive categories targeting the production of the Higgs boson via gluon fusion, via vector boson fusion, in association with a weak vector boson, and in association with a pair of top quarks. The measurement is performed using $\sqrt{s} = 13\,\text{TeV}$ proton-proton (pp) collision data, corresponding to an integrated luminosity of $137\,\text{fb}^{-1}$, recorded by the CMS experiment at the CERN LHC. An excess of events is observed in data with a significance of 3.0 standard deviations, where the expectation for the standard model (SM) Higgs boson with $m_{\text{H}} = 125.38\,\text{GeV}$ is 2.5. The measured signal strength, relative to the SM expectation, is $1.19^{+0.41}_{-0.39}(\text{stat})^{+0.17}_{-0.16}(\text{sys})$. The combination of this result with that from data recorded at centre-of-mass energies of 7 and 8 TeV improves both expected and observed sensitivity by 1%. This result constitutes the first evidence for the Higgs boson decay to fermions of the second generation.

1 Introduction

Since the discovery of the 125 GeV Higgs boson in 2012 at the CERN LHC [1–3], various measurements of its interactions with standard model (SM) particles have been performed. The interactions of the Higgs boson with the electroweak gauge bosons and charged fermions belonging to the third generation of the SM have now been observed, with coupling strengths consistent with SM predictions [4–17]. The Yukawa couplings of the Higgs boson with fermions of the first and second generation, however, have yet to be established experimentally. The SM predicts that the strengths of the couplings of the Higgs boson to fermions are proportional to the fermion masses [18–20]. Consequently, the branching fractions of the Higgs boson decay to fermions of the first and second generation are expected to be small and therefore their measurement is experimentally challenging. The study of $\mathrm{H} \to \mu\mu$ decays is of particular importance since it is the most accessible probe of the Higgs boson couplings to the second generation fermions at the LHC. The expected branching fraction for the Higgs boson decay to a pair of muons with mass 125 GeV is $\mathcal{B}(\mathrm{H} \to \mu\mu) = 2.18 \times 10^{-4}$ [21].

The CMS Collaboration performed a search for $\mathrm{H} \to \mu\mu$ decays using a combination of proton-proton (pp) collision data collected at centre-of-mass energies of 7, 8, and 13 TeV [22], corresponding to integrated luminosities of 5.0, 19.8, and 35.9 fb^{-1}, respectively. An observed (expected in absence of $\mathrm{H} \to \mu\mu$ decays) upper limit of 2.9 (2.2) times the SM prediction was set at 95% confidence level (CL) on the Higgs boson production cross section times $\mathcal{B}(\mathrm{H} \to \mu\mu)$. The corresponding signal strength, relative to the SM expectation, is $\mu = 1.0 \pm 1.0$. The ATLAS Collaboration has also performed a search for $\mathrm{H} \to \mu\mu$ decays using 13 TeV pp collision data corresponding to an integrated luminosity of 139 fb^{-1}, resulting in an excess of events with an observed (expected) significance of 2.0 (1.7) standard deviations and a best-fit signal strength of $\mu = 1.2 \pm 0.6$ [23].

This note describes a search for $\mathrm{H} \to \mu\mu$ decays performed using pp collision data collected by the CMS experiment at $\sqrt{s} = 13$ TeV. This data set corresponds to a total integrated luminosity of 137 fb^{-1}. Results are given for $m_{\mathrm{H}} = 125.38$ GeV, corresponding to the most precise measurement of the Higgs boson mass to date [24]. The final states considered contain two prompt, isolated, and oppositely charged muons from the Higgs boson decay, with a narrow resonant invariant mass peak around the Higgs boson mass. This feature serves as a powerful discriminant against SM background processes. Events are separated into mutually exclusive classes targeting the main production modes of the Higgs boson in pp colliders, namely gluon fusion (ggH), vector boson fusion (VBF), associated production with a weak vector boson (VH, where V = W or Z), and associated production with a pair of top quarks ($\mathrm{t\bar{t}H}$).

The ggH and VBF Higgs boson production modes have the largest cross sections at the LHC, therefore the event categories targeting these production modes are the most sensitive in this search. In the ggH category, the final state may contain, in addition to the pair of muons from the Higgs boson decay, hadronic jets produced by initial (ISR) or final (FSR) state radiation. The largest background in this category consists of Drell–Yan (DY) events in which an off-shell Z boson decays to a pair of muons. Smaller background contaminations arise from leptonic $\mathrm{t\bar{t}}$ decays and diboson (WW, WZ, ZZ) processes. In the VBF analysis, the final state contains two hadronic jets with a large pseudorapidity gap ($\Delta\eta_{\mathrm{jj}}$) and large dijet invariant mass (m_{jj}). These characteristic features enable a significant suppression of the DY background, providing an expected sensitivity to $\mathrm{H} \to \mu\mu$ decays better than in the ggH category despite the smaller VBF production cross section. The VH signal events targeted by this search contain leptonic decays ($\ell = \mu, \mathrm{e}$) of the W or Z boson. This results in a final state with three or more leptons, with the dominant background comprising WZ or ZZ events. Finally, the $\mathrm{t\bar{t}H}$ final state contains

the decays of a pair of top quarks. Therefore, events in this category are characterized by the presence of one or more b quark jets, and may contain additional leptons. The dominant backgrounds in the $t\bar{t}H$ search are $t\bar{t}$ and $t\bar{t}Z$ processes.

2 The CMS detector

The central feature of the CMS apparatus is a superconducting solenoid of 6 m internal diameter, providing a magnetic field of 3.8 T. Within the solenoid volume are a silicon pixel and strip tracker, a lead tungstate crystal electromagnetic calorimeter (ECAL), and a brass and scintillator hadron calorimeter (HCAL), each composed of a barrel and two endcap sections. Forward calorimeters extend the pseudorapidity coverage provided by the barrel and endcap detectors. Muons are detected in gas-ionization chambers embedded in the steel flux-return yoke outside the solenoid. Events of interest are selected using a two-tiered trigger system [25]. The first level (L1) is composed of custom hardware processors, which use information from the calorimeters and muon detectors to select events at a rate of about 100 kHz. The second level, known as high-level trigger (HLT), is a software-based system which runs a version of the CMS full event reconstruction optimized for fast processing, reducing the event rate to about 1 kHz. A more detailed description of the CMS detector, together with a definition of the coordinate system used and the relevant kinematic variables, can be found in Ref. [26].

3 Event reconstruction

The particle-flow (PF) algorithm [27] aims to reconstruct and identify each individual particle in an event, with an optimized combination of information from the various elements of the CMS detector. The energy of photons is obtained from the ECAL measurement. The energy of electrons is determined from a combination of the electron momentum at the primary interaction vertex as determined by the tracker, the energy of the corresponding ECAL cluster, and the energy sum of all bremsstrahlung photons spatially compatible with originating from the electron track. The energy of charged hadrons is determined from a combination of their momentum measured in the tracker and the matching ECAL and HCAL energy deposits, corrected for the response function of the calorimeters to hadronic showers. The energy of neutral hadrons is obtained from the corresponding corrected ECAL and HCAL energies. Finally, the momentum of muons is obtained from the curvature of the corresponding track reconstructed in the inner silicon tracker as well as in the outer muon system.

For each event, hadronic jets are clustered from these reconstructed particles using the infrared and collinear safe anti-k_T algorithm [28, 29] with a distance parameter of $R = \sqrt{\Delta\eta^2 + \Delta\phi^2} = 0.4$. Jet momentum is determined as the vectorial sum of all particle momenta in the jet, and is found from simulation to be, on average, within 5 to 10% of the true momentum over the whole p_T spectrum and detector acceptance. Additional pp interactions within the same or nearby bunch crossings (pileup) can contribute additional tracks and calorimetric energy depositions to the jet momentum. To mitigate this effect, charged particles identified to be originating from pileup vertices are discarded and an offset correction is applied to subtract the remaining contributions from neutral particles. Jet energy corrections are derived from simulation to bring the measured response of jets to that of particle-level jets on average. In situ measurements of the momentum balance in dijet, γ+jets, Z+jets, and multijet events are used to account for any residual differences in jet energy scale in data and simulation [30]. The jet energy resolution in the central region of the detector is typically 15% at 10 GeV, 8% at 100 GeV, and 4% at 1 TeV. Additional selection criteria are applied to each jet to remove those potentially dominated by

anomalous contributions from various subdetector components or reconstruction failures.

The missing transverse momentum vector $\vec{p}_{\mathrm{T}}^{\mathrm{miss}}$ is computed as the negative vector sum of the transverse momenta of all the PF candidates in an event, and its magnitude is denoted as $p_{\mathrm{T}}^{\mathrm{miss}}$ [31]. The $\vec{p}_{\mathrm{T}}^{\mathrm{miss}}$ is modified to account for corrections to the energy scale of the reconstructed jets in the event. Anomalous high-$p_{\mathrm{T}}^{\mathrm{miss}}$ events can be due to a variety of reconstruction failures, detector malfunctions, or noncollision backgrounds. Such events are rejected by event filters that are designed to identify more than 85–90% of the spurious high-$p_{\mathrm{T}}^{\mathrm{miss}}$ events with a mistagging rate smaller than 0.1% [31].

Primary vertices are reconstructed from charged particle tracks in the event. The candidate vertex with the largest value of summed physics-object p_{T}^2 is taken to be the primary pp interaction vertex. The physics objects are the jets, clustered using the jet finding algorithm [28, 29] with the tracks assigned to candidate vertices as inputs, and the associated missing transverse momentum, taken as the negative vector sum of the $\vec{p}_{\mathrm{T}}$ of those jets. Jets are identified to originate from b quarks using a deep neural network discriminant (DeepCSV) that takes as input tracks displaced from the primary interaction vertex, identified secondary vertices, jet kinematic variables, and information related to the presence of soft leptons in the jet [32]. Working points (WP) that yield either a 1% (medium WP) or a 10% (loose WP) probability of misidentifying a light-flavour jet with $p_{\mathrm{T}} > 30\,\mathrm{GeV}$ as a b quark jet are used. The corresponding average efficiencies for the identification of the hadronization products of a b quark as a b quark jet are about 70% and 85%, respectively.

Muon candidates, within the geometrical acceptance of the muon detectors ($|\eta| < 2.4$), are reconstructed by combining the information from the tracker and the muon chambers [33]. These candidates are required to satisfy a set of quality criteria based on the number of hits measured in the tracker and in the muon system, the properties of the fitted muon track, and the impact parameters of the track with respect to the primary vertex of the event. Electron candidates within $|\eta| < 2.5$ are reconstructed using an algorithm that associates fitted tracks in the silicon tracker with electromagnetic energy clusters in the ECAL [34]. Because of non-optimal reconstruction performance, electron candidates in the transition region between the ECAL barrel and endcaps, $1.44 < |\eta| < 1.57$, are discarded. To reduce the misidentification rate, these candidates are required to satisfy identification criteria based on the shower shape of the energy deposit, the matching of the electron track to the ECAL energy cluster, the relative amount of energy deposited in the HCAL detector, and the consistency of the electron track with the primary vertex. Electron candidates identified as coming from photon conversions in the detector are rejected. Identified electrons and muons are required to be isolated from hadronic activity in the event. The isolation sum is defined by summing the p_{T} of all the PF candidates in a cone of radius $R = 0.4$ (0.3) around the muon (electron) track, and is corrected for the contribution of neutral hadrons from pileup interactions [33].

4 Event simulation

Simulated Monte Carlo (MC) events for the signal and dominant background processes are used to optimize the search strategy, evaluate the acceptance, and assess systematic uncertainties. The interactions of the generated final-state particles with the CMS detector are simulated using GEANT4 [35] and are reconstructed with the same algorithms that are used for data. The effect of pileup interactions are modelled by overlaying simulated inelastic pp collisions over the hard-scattering event. The MC events are weighted to reproduce the distribution of the number of interactions per bunch crossing observed in data.

The ggH signal process is simulated at next-to-leading order (NLO) accuracy in perturbative quantum chromodynamics (QCD), using both the MADGRAPH5_aMC@NLO v2.4.2 [36] and POWHEG v2.0 [37–40] Monte Carlo event generators. In the MADGRAPH5_aMC@NLO event generation, up to two additional partons in the final state are included in the matrix element (ME) calculation. The VBF, $qq \to VH$, and $t\bar{t}H$ processes are simulated with POWHEG v2.0 [41, 42] at NLO precision in QCD theory. In addition to the four main production modes, the contributions due to Higgs boson production in association with a pair of b quarks ($b\bar{b}H$), with a Z boson through gluon fusion ($gg \to ZH$), and with a single top quark and either a W boson (tHW) or a quark (tHq) are also considered. The $b\bar{b}H$ process is simulated at NLO precision in QCD with POWHEG, while tHq and tHW ($gg \to ZH$) events are generated at leading order (LO) with the MADGRAPH5_aMC@NLO (POWHEG) generator. Expected signal events are normalized to the production cross sections and $\mathcal{B}(H \to \mu\mu)$ values taken from the recommendations in Ref. [21]. The ggH production cross section is computed at next-to-next-to-NLO (N3LO) precision in QCD, and at NLO in electroweak (EW) theory [43]. The cross section for Higgs boson production in the VBF [44] and $qq \to VH$ [45] modes is calculated at next-to-NLO (NNLO) in QCD, including NLO EW corrections, while the $t\bar{t}H$ cross section is computed at NLO in QCD and EW theory [46, 47]. The $b\bar{b}H$, tHq, and tHW cross sections are computed at NLO in QCD without including higher-order EW corrections. The $H \to \mu\mu$ partial width is computed with HDECAY [48, 49] at NLO in QCD and EW theory. The p_T distribution of the Higgs boson produced via gluon fusion with both POWHEG and MADGRAPH5_aMC@NLO generators is reweighted to match the POWHEG NNLOPS predictions [50, 51]. Simulated signal events are generated, for each production mode, at m_H values of 120, 125, and 130 GeV. Signal predictions for intermediate values of m_H are obtained using different methods depending on the signal extraction strategy designed in each analysis category.

The DY process, which represents the main background in the ggH and VBF categories, is simulated at NLO in QCD using the MADGRAPH5_aMC@NLO generator with up to two partons in the final state at ME level. The corresponding cross section is calculated with FEWZ v3.1 [52] at NNLO in QCD and NLO accuracy in EW theory. The EW production of a Z boson in association with two jets (Zjj-EW) is an important background in the VBF category. This process is simulated at LO using the MADGRAPH5_aMC@NLO v2.6.5 generator. The WZ, $q\bar{q} \to ZZ$, and WW processes, which constitute the main backgrounds in the VH category, are simulated at NLO in QCD using either the POWHEG or MADGRAPH5_aMC@NLO generators. Their production cross sections are corrected with NNLO/NLO K factors taken from Refs. [53], [54], and [55]. The gluon-initiated loop-induced ZZ process ($gg \to ZZ$) is simulated with the MCFM generator [56] at LO and the corresponding production cross section is corrected to match higher-order QCD predictions following the same strategy detailed in Ref. [9]. Minor contributions from triboson processes (WWW, WWZ, WZZ, and ZZZ) are also taken into account and are simulated at NLO in QCD using the MADGRAPH5_aMC@NLO generator. The main backgrounds in the $t\bar{t}H$ category involve the production of top quarks. The $t\bar{t}$ background is simulated with NLO precision in QCD using the POWHEG generator, while single top quark processes are simulated via either POWHEG or MADGRAPH5_aMC@NLO. Cross sections for these background processes are taken from the TOP++ v2.0 [57] and HATOR [58] predictions, derived at NNLO and NLO in QCD, respectively. Finally, contributions from the $t\bar{t}Z$, $t\bar{t}W$, $t\bar{t}WW$, $tt\bar{t}t$, and tZq processes are also considered and are simulated using the MADGRAPH5_aMC@NLO generator at NLO precision in perturbative QCD. The 2016 (2017, 2018) simulations use the NNPDF v3.0 (v3.1) parton distribution functions [59, 60].

The MC simulated events at ME level for both signal and background processes are interfaced with PYTHIA v8.2.2 or higher [61] to simulate the fragmentation, parton shower (PS), and had-

ronization of partons in the initial and final states along with the underlying event description. The CUETP8M1 tune [62] is used for simulated samples corresponding to the 2016 data-taking period, while the CP5 tune [63] is used for the 2017 and 2018 simulated data. For processes simulated at NLO (LO) in QCD with the MADGRAPH5_aMC@NLO generator, jets from the ME calculation are matched to the PS following the FxFx (MLM) prescription [64, 65]. A different prescription is used for VBF signal and Zjj-EW events. The fragmentation, PS, hadronization, and simulation of the underlying event for the Zjj-EW process is performed with the HERWIG++ (2016 simulation) and HERWIG 7 (2017 and 2018) programs [66], as they have been shown to better match the observed data compared to PYTHIA-based predictions in the description of the additional hadronic activity in the rapidity gap between the two leading jets [67]. The EE5C [62] and CH3 tunes [68] are used in the HERWIG++ and HERWIG 7 simulated samples, respectively. Simulated VBF signal events are interfaced with PYTHIA but, rather than the standard p_T-ordered parton shower, the dipole shower is chosen to model ISR and FSR [69]. The dipole shower correctly takes into account the structure of the colour flow between incoming and outgoing quark lines, and its predictions are found to be in good agreement with NNLO QCD calculations, as reported in Ref. [70]. For simulated signal events, the exclusive decay of the Higgs boson to a pair of muons is performed by the chosen parton shower program.

5 Event selection

Events considered in this search are expected to contain two prompt and isolated muons, regardless of the targeted Higgs boson production mode. Events are initially selected by the L1 trigger, requiring at least one muon candidate reconstructed in the muon chambers with p_T larger than 22 GeV. During the 2016 and 2017 data-taking periods, a gradual shift in the timing of the inputs of the ECAL L1 trigger in the forward forward endcap region ($|\eta| > 2.4$) led to a specific trigger inefficiency. A correction for this effect was determined using an unbiased data sample and is found to be relevant only for the VBF category, in events with high-p_T jets with $2.4 < |\eta| < 3.0$. This correction is about 2% (3%) at m_{jj} of 400 GeV in the 2016 (2017) data-taking period and it increases to about 6% (9%) for m_{jj} larger than 2 TeV. At the HLT level, events of interest are collected using single muon triggers that have a p_T threshold of 27 (24) GeV for data recorded in 2017 (2016, 2018).

After passing the trigger selections, every event is required to contain at least two oppositely charged muons with $p_T > 20$ GeV and $|\eta| < 2.4$ passing certain selection requirements on the number of measurements in the tracker and in the muon systems, as well as on the quality of the fitted muon track. Each muon is also required to be isolated in order to reject events with nonprompt or misidentified muon candidates. The muon isolation, as defined in Section 3, is required to be less than 25% of the muon p_T. Muons are selected with an average efficiency of about 95%. In addition, at least one of the two muons is required to have $p_T > 29$ (26) GeV for data collected in 2017 (2016, 2018), to be consistent with the trigger selections.

In about 9% of signal events, a muon from the Higgs boson decay radiates a photon that carries away a significant fraction of the muon momentum. If not taken into account, this worsens the resolution of the dimuon invariant mass ($m_{\mu\mu}$) peak in signal events. Furthermore, if the FSR photon falls in the isolation cone of the corresponding muon candidate, it can significantly increase the value of the isolation sum, thereby creating an inefficiency in selecting signal events. Therefore, a procedure is implemented to identify and recover the contribution of FSR photons. PF photons with $p_T > 2$ GeV and $|\eta| < 2.5$ that are not associated with reconstructed electrons are considered as FSR photon candidates if they lie inside a cone of $R = 0.5$ around a muon track. These candidates are then required to be loosely isolated

and collinear with the muon such that $I_\gamma/p_{\mathrm{T}}(\gamma) = (\Sigma_i^{PF} p_{\mathrm{T}}^i(\Delta R(\gamma, i) < 0.3))/p_{\mathrm{T}}(\gamma) < 1.8$ and $\Delta R(\mu, \gamma) < 0.012 \times p_{\mathrm{T}}(\gamma)^2$, where $p_{\mathrm{T}}(\gamma)$ is the p_{T} of the FSR photon candidate and the index i refers to the PF candidates other than the muon within a cone of $R = 0.3$ around the photon. In order to suppress the possible contaminations from $\mathrm{H} \to \mathrm{Z}(\mu\mu)\gamma$ decays, the ratio between the transverse momentum of the FSR photon and the associated muon is required to be smaller than 0.4. In the case of multiple FSR candidates associated with a muon, the candidate with the smallest value of $\Delta R(\mu, \gamma)/p_{\mathrm{T}}(\gamma)^2$ is chosen. The momentum of the photon is added to that of the muon and its contribution to the muon isolation sum is ignored. This FSR recovery procedure increases the signal efficiency by about 3% and improves the $m_{\mu\mu}$ resolution by around 2%.

The sensitivity of this search depends primarily on the resolution of the $m_{\mu\mu}$ peak in the signal events. The $m_{\mu\mu}$ resolution depends on the precision with which the p_{T} of the muons is measured, which decreases with increasing muon $|\eta|$. The p_{T} resolution of muons passing through the central barrel region of the detector ($|\eta| < 0.9$) is around 1–2%, whereas the p_{T} resolution of muons passing through the endcaps of the muon system ($|\eta| > 1.2$) ranges from 2 to 3.5%. The muon momentum scale and resolution are calibrated in bins of p_{T} and η using the decay products of known dilepton resonances, following the method described in Ref. [71]. In signal events, the Higgs boson decays into a muon pair at the interaction point. Therefore, the precision of the muon p_{T} measurement can be improved by including that position as an additional hit of the muon track. The corresponding adjustment, derived in simulated $\mathrm{Z} \to \mu\mu$ events, is proportional (on average) to the product of the muon p_{T}^2 and the minimum distance in the transverse plane between the muon track and the beam position. The resulting improvement in the expected $m_{\mu\mu}$ resolution in signal events ranges from 3% to 10%, depending on muon p_{T}, η, and the data-taking period.

In order to maximize the analysis sensitivity, dimuon event candidates selected with the requirements described above are separated into orthogonal classes based on the features of the final state expected from each production mode. Events with one or two b-tagged jets are assigned to the $\mathrm{t\bar{t}H}$ production category, which is further split into the $\mathrm{t\bar{t}H}$ hadronic and $\mathrm{t\bar{t}H}$ leptonic subclasses by the number of additional leptons (μ or e) in the final state. Dimuon events with one (two) additional lepton(s) and no b-tagged jets are assigned to the WH (ZH) category. Events with neither additional leptons nor b-tagged jets belong to the VBF category if a pair of jets is present with large m_{jj} and $\Delta\eta_{\mathrm{jj}}$. The remaining untagged events, which constitute about 96% of the total dimuon candidate events, belong to the ggH-enriched category. In each production category, multivariate techniques are used to enhance the discrimination between the expected signal and background contributions by further dividing events into several subcategories with different signal-to-background ratios. The measured $\mathrm{H} \to \mu\mu$ signal is then extracted via a simultaneous maximum-likelihood fit across all event categories to observables chosen for each category to maximize the overall measurement precision.

6 Event categories for VBF production

A dimuon event passing the baseline event selection detailed in Section 5 is considered in the category targeting the VBF production mode if it contains two or more jets, with the p_{T} of the leading jet ($p_{\mathrm{T}}(\mathrm{j}_1)$) larger than 35 GeV, the p_{T} of the second highest-p_{T} jet ($p_{\mathrm{T}}(\mathrm{j}_2)$) greater than 25 GeV, and the $|\eta|$ of both jets less than 4.7. Jet candidates are required to be spatially separated from both of the two muons, with $\Delta R(\mu, j) > 0.4$. In addition, the two highest-p_{T} jets in the event are required to have m_{jj} larger than 400 GeV and $|\Delta\eta_{\mathrm{jj}}|$ greater than 2.5. An event is rejected in the VBF category if it contains one (two) jet(s) inside the tracker fiducial volume

($|\eta| < 2.5$) with $p_T > 25$ GeV and identified as a b quark jet by the medium (loose) WP of the DeepCSV algorithm. These requirements suppress the $t\bar{t}$ and single top backgrounds and ensure mutual exclusivity between the VBF and $t\bar{t}H$ categories. Moreover, events containing an additional muon (electron) with $p_T > 20$ GeV and $|\eta| < 2.4$ (2.5) passing the selection criteria described in Section 8 are discarded. This requirement ensures mutual exclusivity between the analyses targeting VBF and VH production. Selected events are further grouped into two independent categories. Events in which the two muons form an invariant mass between 115 and 135 GeV belong to the signal region (VBF-SR), which is enriched in signal-like events. Events with $110 < m_{\mu\mu} < 115$ GeV or $135 < m_{\mu\mu} < 150$ GeV belong to the mass sideband region (VBF-SB), which is used as a control region to estimate the background.

A deep neural network (DNN) multivariate discriminant is trained to distinguish the expected signal from background events using kinematic input variables that characterize the signal and the main background processes in the VBF-SR. The DNN is implemented using KERAS [72] with TENSORFLOW [73] as backend. The DNN inputs include six variables associated with the production and decay of the dimuon system, namely the $m_{\mu\mu}$, the per-event uncertainty in the measured dimuon mass $\sigma(m_{\mu\mu})$, the dimuon transverse momentum ($p_T(\mu\mu)$), the dimuon rapidity ($y_{\mu\mu}$), and the azimuthal angle (ϕ_{CS}) and the cosine of the polar angle ($\cos\theta_{CS}$) computed in the dimuon Collins–Soper rest frame [74]. The DNN also takes as input a set of variables describing the properties of the dijet system, namely the full momentum vector of the two highest-p_T jets in the event ($p_T(j_1)$, $p_T(j_2)$, $\eta(j_1)$, $\eta(j_2)$, $\phi(j_1)$, and $\phi(j_2)$), m_{jj}, and $\Delta\eta_{jj}$. Furthermore, observables sensitive to angular and p_T correlations between muons and jets are also included, namely the minimum $\Delta\eta$ between the muon pair and the two leading jets, the Zeppenfeld variable (z^*) [75] constructed from $y_{\mu\mu}$ and the rapidities of the two jets as

$$z^* = \frac{y_{\mu\mu} - (y_{j_1} + y_{j_2})/2}{|y_{j_1} - y_{j_2}|}, \tag{1}$$

and the p_T-balance ratio as

$$R(p_T) = \frac{|\vec{p}_T^{\,\mu\mu} + \vec{p}_T^{\,jj}|}{p_T(\mu\mu) + p_T(j_1) + p_T(j_2)}. \tag{2}$$

VBF signal events are expected to have suppressed hadronic activity in the rapidity gap between the two leading jets. This feature is exploited by considering "soft jets" in the event that are defined by clustering, via the anti-k_T algorithm with a distance parameter of 0.4, charged particles from the primary interaction vertex excluding the two identified muons and those associated with the two VBF jets. The soft jets are required to have $p_T > 5$ GeV. The number of soft jets in an event, as well as the scalar sum of their transverse momenta, are used as additional input variables. Finally, since jets in signal events are expected to originate from quarks, whereas in the DY process they can also be initiated by gluons, the quark-gluon likelihood (QGL) [76, 77] of the two leading jets is also used as input to the DNN.

The DNN is trained using simulated events from signal (VBF) and background (DY, Zjj-EW, $t\bar{t}$, and diboson) processes selected in the VBF-SR. Signal events generated with $m_H = 125$ GeV are used in the DNN training. Four independent networks are first optimized to accomplish different goals and starting from different input features. Two networks exploit the full set of variables described above in order to optimize the separation between the VBF signal and the Zjj-EW or DY background, while the other two networks optimize the separation between the VBF signal and the total expected background. The first of the two networks discriminating

against the total background uses all the input features except for the $m_{\mu\mu}$, while the second uses only the dimuon mass and its resolution. Finally, the outputs of these four networks are combined into a final network that classifies events. Every network contains three or four hidden layers, each with a few tens of nodes. After each hidden layer, a 20% dropout [78] is used in order to regularize the model. In each network, the minimized loss function is the binary cross-entropy. All trainings are performed using a four-fold strategy, where 50% of the events are used for training, 25% for validation, and 25% for testing. The validation sample is used to optimize the DNN hyper-parameters, while the test sample is used to evaluate the DNN performance. The selected training epoch minimizes the difference in the expected Asimov significance [79] between the training and the validation samples.

Events belonging to the VBF-SR are divided into nonoverlapping bins based on the DNN value, independently for each data-taking period. These bins are defined to achieve optimal sensitivity while minimizing the total number of bins. Given the negligible correlation between the dimuon mass and other input variables, in the VBF-SB region a mass-decorrelated DNN is evaluated by replacing the dimuon mass with a fixed value of 125 GeV. The $m_{\mu\mu}$ variable is therefore marginalized from the network, producing an output score that resembles the main features of the DNN distribution in the VBF-SR.

The signal is extracted from a binned maximum-likelihood fit to the output of the DNN discriminator simultaneously in the VBF-SR and the VBF-SB regions. Because of significant variations in the detector response for forward jets during different data-taking periods, the fit is performed separately for data collected in 2016, 2017, and 2018. In order to further improve the sensitivity to $\mathrm{H} \to \mu\mu$ decays, the contributions of the various background processes are estimated from simulation. This strategy follows that employed by the CMS Collaboration in the measurement of the Zjj-EW cross section with 13 TeV data [67] and provides a better performance with respect to the strategy used in previous results [22, 80]. In the VBF selected regions with high signal purity, the background prediction from the previous data-driven approach is strongly limited by the number of observed events in the mass sidebands. The simulation-based method therefore better constrains the background prediction compared to a data-driven approach because of a better overall precision in the prediction from simulated events, including systematic uncertainties. This results in an improved sensitivity of about 20% in the VBF category.

Theoretical uncertainties affect both the expected rate and the shape of signal and background histograms (templates) used in the fit. The Higgs boson production cross section for the various modes, and their corresponding uncertainties, are taken from Ref. [21]. These include uncertainties in the choice of the PDF as well as the QCD renormalization (μ_{R}) and factorization (μ_{F}) scales. The uncertainty in the prediction of $\mathcal{B}(\mathrm{H} \to \mu\mu)$ is also considered. For the ggH process, seven independent additional sources are included to account for the uncertainty in the modelling of the $p_{\mathrm{T}}(\mathrm{H})$ distribution, the number of jets in the event, and its contamination in the VBF selected region as described in Ref. [21]. The magnitude of these uncertainties for ggH events in the VBF category varies from around 15% to 25%. Similarly, for the VBF process, uncertainties in the modelling of the $p_{\mathrm{T}}(\mathrm{H})$, $p_{\mathrm{T}}(\mathrm{Hjj})$, jet multiplicity, and m_{jj} distributions are considered. The total uncertainty from these sources is around 2–4%. For each background process, template variations are built by changing the values of μ_{R} and μ_{F} by factors of 2 and 0.5 from the default values used in the ME calculation, as well as by comparing the nominal distributions with those obtained using the alternative PDFs of the NNPDF set. These theoretical uncertainties are correlated across years and regions but are uncorrelated between processes. The shape uncertainty arising from the PS model is assessed by varying several parameters that control the properties of the ISR and FSR produced by PYTHIA. The Zjj-EW and

VBF signal simulations are very sensitive to the PS model, as shown in Refs. [67, 70]. A conservative PS uncertainty is assigned to the Zjj-EW background and VBF signal, defined as the full symmetrized difference between PYTHIA (dipole shower) and HERWIG (angular-ordered) predictions in each DNN bin, which is larger than that obtained by varying the parton shower ISR and FSR parameters.

Several experimental sources of uncertainty are taken into account for both signal and background processes. These include the uncertainty in the measurement of the integrated luminosity, in the modelling of the pileup conditions during data taking, in the measurement of the muon selection and trigger efficiencies, in the muon energy scale and resolution, in the efficiency of vetoing b quark jets, and in the jet energy scale and resolution. Most of the uncertainty sources affecting the jet energy scale are correlated across processes and years, while those affecting the jet energy resolution are only correlated across processes but uncorrelated among data-taking periods. For data collected in 2016 and 2017, an inefficiency in the L1 trigger was observed as detailed in Section 5. A correction based on dedicated measurements performed on data is applied to simulated events, and an uncertainty corresponding to 20% of this correction is considered. Lastly, a significant fraction (about 30–35%) of the DY background populating the DNN bins at low score comprises events in which either the leading or the subleading jet are in the forward region of the detector ($|\eta| > 3.0$) and are not matched with a jet at the generator level. These jets originate from either the soft emissions produced by the parton shower or from pileup interactions, and are promoted above the jet p_{T} thresholds used in the analysis by the detector response. The normalization of this component of the DY background is assigned a flat prior and is directly constrained by the low DNN score events in the observed data belonging to the VBF-SR and VBF-SB regions. Because of significant variations in the detector response in the forward region over time, this normalization parameter is uncorrelated across years. The uncertainty arising from the limited size of simulated samples is also taken into account by allowing each bin of the total background template to vary within the corresponding statistical uncertainty using the Beeston–Barlow technique [81]. These uncertainties are uncorrelated across the bins of the DNN templates used in the fit.

Figure 1 shows the observed and the predicted distributions of the DNN discriminant in the VBF-SB after applying the VBF event selection. The background prediction is obtained from a simultaneous signal-plus-background fit performed across the VBF-SR and VBF-SB regions as well as data-taking periods. Similarly, Fig. 2 shows the distributions of the DNN discriminator in the VBF-SR, obtained after performing the same signal-plus-background fit. The expected distributions for the Higgs boson signal produced via ggH and VBF production, assuming the SM production cross sections and branching fraction to a pair of muons, are overlaid. Figure 3 shows the observed and predicted DNN score distributions in the VBF-SB (left) and VBF-SR (right) regions for the combination of 2016, 2017, and 2018 data. The lower panel shows the ratio between the data and the post-fit background prediction, with the best-fit signal contribution indicated by the blue line in the VBF-SR. Finally, Table 1 reports for each bin of the DNN output in the VBF-SR the expected number of VBF and ggH signal events, the observed number of events in data, the total background prediction and its uncertainty, and the $\mathrm{S}/(\mathrm{S}+\mathrm{B})$ and $\mathrm{S}/\sqrt{\mathrm{B}}$ ratios obtained by summing the post-fit estimates from each of the three data-taking periods.

7 Event categories for ggH production

An event is included in the ggH category if it contains exactly two muons passing the baseline selection requirements detailed in Section 5. Events with additional electrons or muons

Table 1: Event yields in each bin or in group of bins defined along the DNN output in the VBF-SR for various processes. The background yields and the corresponding uncertainties are obtained after performing a combined signal-plus-background fit across analysis regions and data-taking periods. The observed event yields and the expected signal contribution at $m_{\mathrm{H}} = 125.38\,\mathrm{GeV}$, produced via VBF and ggH modes and assuming SM cross sections and $\mathcal{B}(\mathrm{H} \to \mu\mu)$, are also reported.

DNN bin	Signal	VBF (%)	ggH (%)	Bkg. $\pm$ ΔB	S/(S + B) (%)	S/$\sqrt{\mathrm{B}}$	Data
1–3	19.5	30	70	8894 $\pm$ 67	0.22	0.21	8815
4–6	11.6	57	43	394 $\pm$ 8	2.90	0.59	388
7–9	8.43	73	27	103 $\pm$ 4	7.66	0.84	121
10	2.30	85	15	15.1 $\pm$ 1.4	13.2	0.60	18
11	2.15	88	12	9.1 $\pm$ 1.2	19.2	0.72	10
12	2.10	87	13	5.8 $\pm$ 1.1	26.7	0.88	6
13	1.87	94	6	2.6 $\pm$ 0.9	41.8	1.18	7

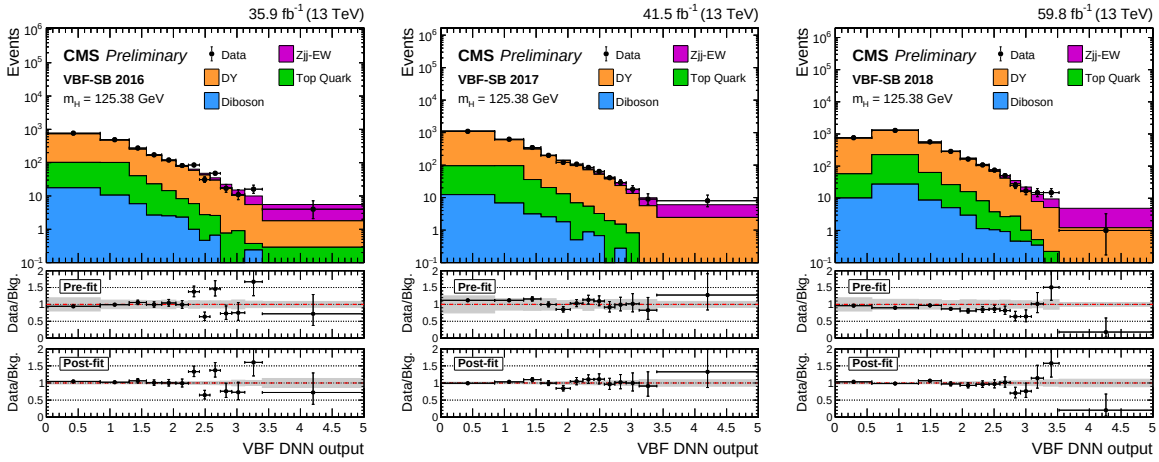

Figure 1: The observed DNN output distribution for data collected in 2016 (left), 2017 (middle), and 2018 (right) in the VBF-SB region compared to the post-fit background estimate from SM processes. The predicted backgrounds are obtained from a signal-plus-background fit performed across analysis regions and years. In the second panel, the ratio between data and the pre-fit background prediction is shown. The grey band indicates the total pre-fit uncertainty obtained from the systematic sources previously described. The third panel shows the ratio between data and the post-fit background prediction from the signal-plus-background fit. The grey band indicates the total background uncertainty after performing the fit.

are rejected. An event may contain zero or more jets that are spatially separated ($\Delta R > 0.4$) from either of the two muons. In order to ensure mutual exclusivity with the VBF category, events containing two or more jets with $p_{\mathrm{T}} > 25\,\mathrm{GeV}$ are only considered if the leading jet has $p_{\mathrm{T}} < 35\,\mathrm{GeV}$, the invariant mass of the two highest-p_{T} jets is smaller than 400 GeV, or the $|\Delta\eta_{\mathrm{jj}}|$ is smaller than 2.5. Lastly, events containing at least two b-tagged jets passing the loose WP of the DeepCSV algorithm or at least one jet passing the medium WP are also ignored.

A multivariate discriminant based on boosted decision trees (BDTs) is employed to discriminate between signal and background events. To account for the evolution in the detector response during data-taking periods, the BDT discriminant is trained separately for the 2016, 2017, and 2018 years using the TMVA package [82], resulting in three independent BDT outputs. The input variables are carefully chosen such that the BDT discriminants are effectively

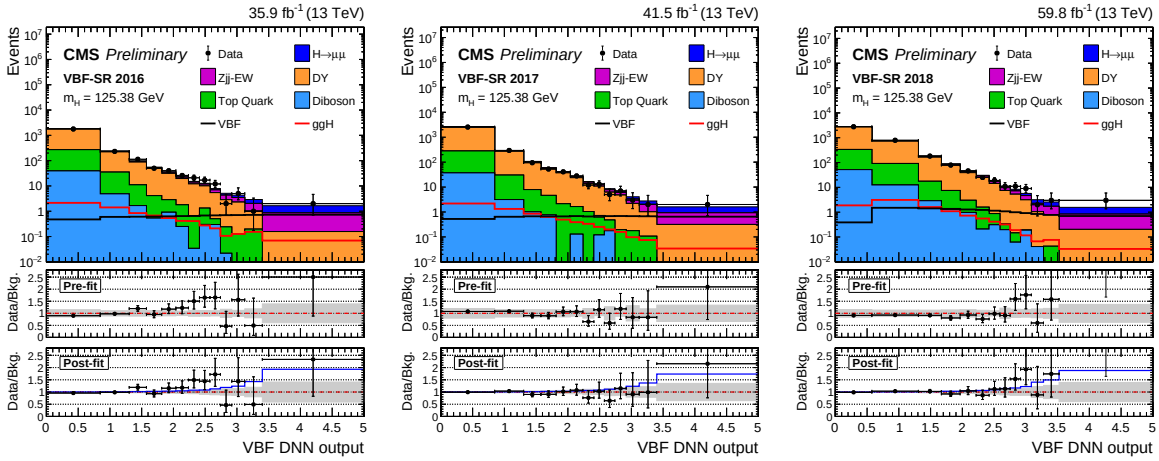

Figure 2: The observed DNN output distribution in the VBF-SR region compared to the post-fit background estimate for the contributing SM processes. The post-fit distributions for the Higgs boson signal produced via ggH and VBF modes with $m_{\mathrm{H}} = 125.38\,\mathrm{GeV}$ are overlaid. The predicted backgrounds are obtained from a signal-plus-background fit performed across analysis regions and years. The description of the three panels is the same as in Fig. 1. The blue histogram (first panel) and solid line (third panel) indicate the total signal extracted from the fit with $m_{\mathrm{H}} = 125.38\,\mathrm{GeV}$.

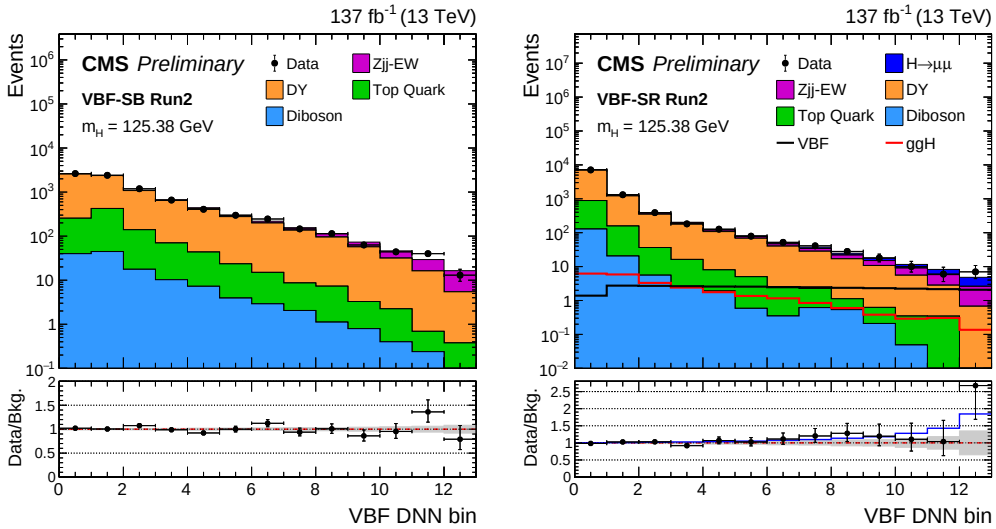

Figure 3: The observed DNN output distribution in the VBF-SB (left) and VBF-SR (right) regions for the combination of 2016, 2017, and 2018 data, compared to the post-fit prediction from SM processes. The lower panel shows the ratio between data and the post-fit background prediction from the signal-plus-background fit. The best-fit $\mathrm{H} \to \mu\mu$ signal contribution is indicated by the blue line, and the grey band indicates the total background uncertainty.

uncorrelated with $m_{\mu\mu}$. This is required by the chosen analysis strategy, in which events are first divided into independent subcategories based on the BDT output, then the presence of a potential signal is extracted from each subcategory by searching for a narrow peak over a smoothly falling background in the $m_{\mu\mu}$ distribution. Given the prior knowledge of the expected DY background shape and the large amount of data events in the mass sideband around the peak that can be used to constrain the background, this strategy maximizes the analysis sensitivity by estimating the total background directly from data.

The BDT discriminants include input variables that describe the production and decay of the dimuon system, namely $p_{\mathrm{T}}(\mu\mu)$, $y_{\mu\mu}$, ϕ_{CS}, and $\cos\theta_{\mathrm{CS}}$. In addition, the η of the two muons and the muon p_{T} relative to $m_{\mu\mu}$ are also included. In order to increase the signal-to-background separation for events in which the ggH signal is produced in association with jets, the BDT discriminants also take into account the p_{T} and η of the leading jet in the event with $p_{\mathrm{T}} > 25\,\mathrm{GeV}$ and the absolute distance in η and ϕ between the jet and the muon pair. For events with two or more jets in the final state with $p_{\mathrm{T}} > 25\,\mathrm{GeV}$, additional inputs are included: the m_{jj}, $\Delta\eta_{\mathrm{jj}}$, and $\Delta\phi_{\mathrm{jj}}$ of the two highest-p_{T} jets. The m_{jj} in particular is sensitive to the residual contamination from VBF and VH modes, in which the weak vector boson decays hadronically. Furthermore, the Zeppenfeld variable defined in Eq.(1), min-$\Delta\eta(\mu\mu, \mathrm{j}_1, \mathrm{j}_2)$, and min-$\Delta\phi(\mu\mu, \mathrm{j}_1, \mathrm{j}_2)$ are also included, which target the presence of VBF signal events in the ggH selected region. Lastly, the total number of jets in the event with $p_{\mathrm{T}} > 25\,\mathrm{GeV}$ and $|\eta| < 4.7$ is used as an input to the BDT discriminants.

The signal simulation considered in the training of the multivariate discriminators includes ggH, VBF, VH, and $\mathrm{t\bar{t}H}$ processes. The ggH sample used in the training is generated via POWHEG since it provides positively weighted events at NLO in QCD. In later stages of the analysis, the prediction from MADGRAPH5_AMC@NLO is used instead since it provides a more accurate description of gluon fusion events accompanied by more than one jets, as detailed in Section 4. The background simulation consists of DY, $\mathrm{t\bar{t}}$, single top, diboson, and Zjj-EW processes. Only events with $m_{\mu\mu}$ in the range 115–135 GeV are included in the training. Signal and background events both contain two prompt muons in the final state, and the corresponding dimuon mass resolution ($\sigma_{\mu\mu}/m_{\mu\mu}$) carries no discrimination power between them. For this reason, $\sigma_{\mu\mu}/m_{\mu\mu}$ is not added as an input to the BDT. Instead, signal events in the BDT training are assigned a weight proportional to the expected mass resolution, derived from the uncertainties in the p_{T} measurement from the individual muon tracks. This weighting improves the average signal $\sigma_{\mu\mu}/m_{\mu\mu}$ in the high score BDT region by assigning increased importance to the high-resolution signal events.

Apart from $m_{\mu\mu}$, the $p_{\mathrm{T}}(\mu\mu)$ is one of the most discriminating observables in the ggH category. Discrepancies between data and simulation in the $p_{\mathrm{T}}(\mu\mu)$ spectrum similar to those reported in Ref. [83] are also observed in this search. In order to correctly model the $p_{\mathrm{T}}(\mu\mu)$ spectrum of the DY background during the training of the BDT discriminants, corrections are derived for each data-taking period by reweighting the $p_{\mathrm{T}}(\mu\mu)$ distribution of the DY simulation to reproduce the observation in data for dimuon events with $70 < m_{\mu\mu} < 110\,\mathrm{GeV}$. These corrections are obtained separately for events containing zero, one, and two or more jets with $p_{\mathrm{T}} > 25\,\mathrm{GeV}$ and $|\eta| < 4.7$.

Figure 4 (left) shows the BDT score distribution, comparing data to the prediction from simulation in events with $110 < m_{\mu\mu} < 150\,\mathrm{GeV}$, where the outputs of the individual BDTs obtained in each year are summed together. The distributions for various signal processes (ggH, VBF, and VH + $\mathrm{t\bar{t}H}$) are also shown. Five event categories are defined based on the score of these BDT discriminants. The category boundaries are determined via an iterative process that aims to maximize the expected sensitivity of this analysis to $\mathrm{H} \to \mu\mu$ decays of the SM Higgs boson. The expected sensitivity is estimated from signal-plus-background fits to the $m_{\mu\mu}$ distribution in simulated events with $110 < m_{\mu\mu} < 150\,\mathrm{GeV}$. In these fits, the Higgs boson signal is modelled using a parametric shape, the double-sided Crystal Ball function (DCB)

$$\mathrm{DCB}(m_{\mu\mu}) = \begin{cases} e^{-(m_{\mu\mu}-\hat{m})^2/2\sigma^2} & -\alpha_{\mathrm{L}} < \frac{m_{\mu\mu}-\hat{m}}{\sigma} < \alpha_{\mathrm{R}} \\ \left(\frac{n_{\mathrm{L}}}{|\alpha_{\mathrm{L}}|}\right)^{n_{\mathrm{L}}} e^{-\alpha_{\mathrm{L}}^2/2}\left(\frac{n_{\mathrm{L}}}{|\alpha_{\mathrm{L}}|} - |\alpha_{\mathrm{L}}| - \frac{m_{\mu\mu}-\hat{m}}{\sigma}\right)^{-n_{\mathrm{L}}} & \frac{m_{\mu\mu}-\hat{m}}{\sigma} \leq -\alpha_{\mathrm{L}} \\ \left(\frac{n_{\mathrm{R}}}{|\alpha_{\mathrm{R}}|}\right)^{n_{\mathrm{R}}} e^{-\alpha_{\mathrm{R}}^2/2}\left(\frac{n_{\mathrm{R}}}{|\alpha_{\mathrm{R}}|} - |\alpha_{\mathrm{R}}| - \frac{m_{\mu\mu}-\hat{m}}{\sigma}\right)^{-n_{\mathrm{R}}} & \frac{m_{\mu\mu}-\hat{m}}{\sigma} \geq \alpha_{\mathrm{R}} \end{cases} . \quad (3)$$

The core of the DCB function consists of a Gaussian distribution of mean $\hat{m}$ and standard deviation σ, while the tails on either side are modelled by a power-law function with parameters α_{L} and n_{L} (low-mass tail), and α_{R} and n_{R} (high-mass tail). The total expected background is modelled with a modified form of the Breit–Wigner function (mBW),

$$\mathrm{mBW}(m_{\mu\mu}; m_{\mathrm{Z}}, \Gamma_{\mathrm{Z}}, a_1, a_2, a_3) = \frac{e^{a_2 m_{\mu\mu} + a_3 m_{\mu\mu}^2}}{(m_{\mu\mu} - m_{\mathrm{Z}})^{a_1} + (\Gamma_{\mathrm{Z}}/2)^{a_1}}, \quad (4)$$

where the parameters m_{Z} and Γ_{Z} refer to the measured Z boson mass of 91.19 GeV and width 2.49 GeV [84], and the parameters a_1, a_2, and a_3 have flat priors. A first category boundary is selected by optimizing the total expected significance against all possible boundaries defined in quantiles of signal efficiency. This strategy accounts for the slight differences in the BDT shapes among data-taking periods for both signal and background processes. This process is repeated recursively to define additional category boundaries until the further gain in the expected significance is less than 1%. The optimized event categories are labelled as "ggH-cat1″, "ggH-cat2″, "ggH-cat3″, "ggH-cat4″, and "ggH-cat5″ corresponding to signal efficiency intervals of 0–30%, 30–60%, 60–80%, 80–95%, and >95%, respectively. The grey vertical boxes in Figure 4 (left) indicate the range of variation of the BDT boundaries for the optimized event categories described above.

A simultaneous binned maximum-likelihood fit to the observed $m_{\mu\mu}$ distributions is performed over the mass range 110–150 GeV to extract the $\mathrm{H} \to \mu\mu$ signal. A bin size of 50 MeV is chosen for the $m_{\mu\mu}$ distributions, which is about one order of magnitude smaller than the expected resolution of the signal peak. In each event category, simulated signal distributions from the different production modes (ggH, VBF, WH, ZH, and $\mathrm{t\bar{t}H}$) are modelled independently with DCB functions, and the best-fit values of the DCB tail parameters are treated as constants in the final fit to the data. The $\hat{m}$ and σ parameters of the DCB function represent the peak position and resolution of the Higgs boson resonance, respectively. These are the only signal shape parameters allowed to vary, within Gaussian constraints, with widths corresponding to the muon momentum scale (up to 0.2%) and resolution uncertainties (up to 10%) in each event category. Figure 4 (right) shows the total signal model for $m_{\mathrm{H}} = 125$ GeV obtained by summing the contributions from the different production modes in the best and the worst resolution categories of the ggH category, ggH-cat4 and ggH-cat1.

The theoretical and experimental sources of systematic uncertainty affecting the expected signal rate in each event category are similar to those described in the VBF analysis. Experimental uncertainties in the measurement of the muon selection efficiencies (0.5–1% per category), jet energy scale (1–4% per category) and resolution (1–6% per category), the modelling of the pileup conditions (0.3–0.8% per category), the integrated luminosity (about 2.5% per year), and the efficiency of vetoing b quark jets (0.1–0.5% per category) are considered. Theoretical uncertainties in the prediction of the Higgs boson production cross section, decay rate, and acceptance are also included, corresponding to a total uncertainty in the ggH process yield ranging from 6–12% depending on category. Rate uncertainties are modelled in the signal extraction as nuisance parameters acting on the relative signal yield with log-normal priors.

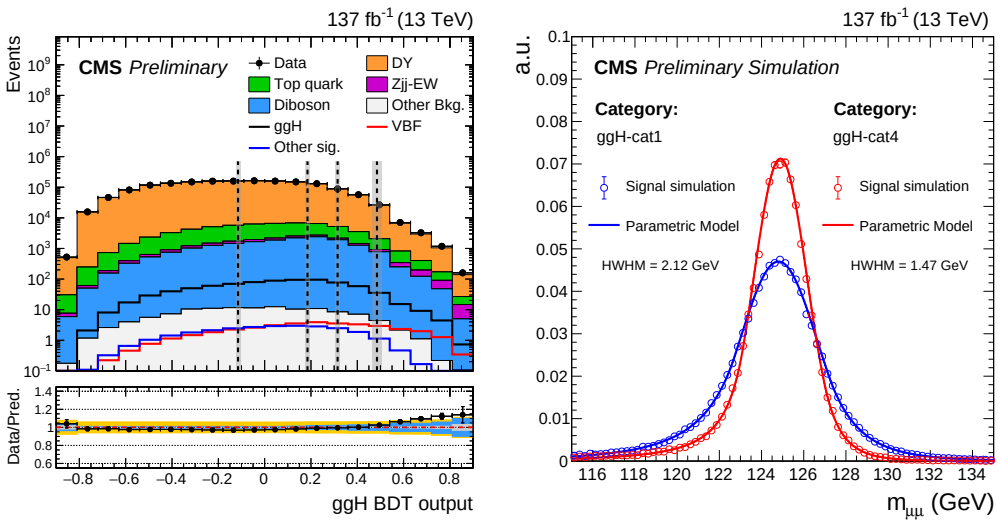

Figure 4: Left: the observed BDT output distribution compared to the prediction from the simulation of various SM background processes. Dimuon events passing the event selection requirements of the ggH category, with $m_{\mu\mu}$ between 110–150 GeV, are considered. The expected distributions for ggH, VBF, and other signal processes are overlaid. The gray vertical boxes indicate the range of variation of the BDT boundaries for the optimized event categories defined in each data-taking period. In the lower panel, the ratio between data and the expected background is shown. The grey band indicates the uncertainty due to the limited size of the simulated samples. The azure band corresponds to the sum in quadrature between the statistical and experimental systematic uncertainties, while the orange band additionally includes the theoretical uncertainties affecting the background prediction. Right: the signal shape model for the simulated $\mathrm{H} \to \mu\mu$ sample with $m_{\mathrm{H}} = 125$ GeV in the best (red) and the worst (blue) resolution categories.

The background contribution in each category is modelled with analytical functions. No prior knowledge of the parameters of these functions or the yield of the total background is assumed. These parameters are therefore constrained directly by the observed data in the signal-plus-background fit. Since the background composition expected from simulation is very similar across categories and largely dominated by the DY process, the background shape in $m_{\mu\mu}$ is similar in all categories. There are, however, variations in the overall slope of the $m_{\mu\mu}$ spectrum across the BDT score categories. The function describing the background in each event category is therefore defined as the product of a "core" shape that is common among all event categories, with parameters correlated across categories, and a polynomial term (shape modifier) specific to each event category that modulates the core shape. This background modelling approach is referred to as the "core-pdf method". The core background shape is obtained from an envelope of three distinct functions: the modified Breit–Wigner (mBW) defined in Eq.(4), a sum of two exponential functions, and the product of a non-analytical shape derived from the FEWZ v3.1 generator [52] and a third-order Bernstein polynomial. Each of these functions contains three freely floating shape parameters. The non-analytical shape derived from the FEWZ generator is obtained by simulating DY events at NNLO precision in QCD corrections and NLO accuracy in EW theory and smoothing out the resulting $m_{\mu\mu}$ distribution using a spline function [85, 86]. In a given category, each of the three core functions is modulated by either a third- (ggH-cat1 and ggH-cat2) or a second-order polynomial, with parameters uncorrelated across categories. A discrete profiling method [87] is employed, which treats the choice of the core function used to model the background as a discrete nuisance parameter in the signal extraction.

The following strategy is adopted to estimate the uncertainty in the measured signal due to the choice of parametric function for the background model. In each event category, background-only fits to the data are performed using different types of functions: the modified Breit–Wigner (mBW), a sum of two exponentials, a sum of two power laws, a Bernstein polynomial, the product between the FEWZ spline and a Bernstein polynomial, the product between the "BWZ" function, defined as

$$\mathrm{BWZ}(m_{\mu\mu}; a, m_{\mathrm{Z}}, \Gamma_{\mathrm{Z}}) = \frac{\Gamma_{\mathrm{Z}} e^{a m_{\mu\mu}}}{(m_{\mu\mu} - m_{\mathrm{Z}})^2 + (\Gamma_{\mathrm{Z}}/2)^2}, \tag{5}$$

and a Bernstein polynomial, and the "BWZGamma" function

$$\mathrm{BWZGamma}(m_{\mu\mu}; a, f, m_{\mathrm{Z}}, \Gamma_{\mathrm{Z}}) = f \times \mathrm{BWZ}(m_{\mu\mu}; a, m_{\mathrm{Z}}, \Gamma_{\mathrm{Z}}) + (1 - f) \times \frac{e^{a m_{\mu\mu}}}{m_{\mu\mu}^2}. \tag{6}$$

The BWZGamma function is the sum of a Breit–Wigner function and a $1/m_{\mu\mu}^2$ term, which are used to model the Z boson and the photon contributions to the $m_{\mu\mu}$ spectrum in DY events, respectively. Both terms are multiplied by an exponential function to approximate the effect of the PDF. The BWZ function is a Breit–Wigner distribution with an exponential tail. For the functions including Bernstein polynomials, a Fisher test [88] is used to determine the maximum order of the polynomials to be considered in the fit. The chosen functional forms are able to fit the data with a χ^2 probability larger than 5% in all categories.

Pseudodata sets are generated across all event categories from the post-fit background shapes obtained for each type of function in each category, taking into account the uncertainties in the fit parameters as well as their correlations, and injecting a given number of signal events. Signal-plus-background fits are performed on the pseudodata sets using the core-pdf method. The median difference between the measured and injected signal yields, relative to the post-fit uncertainty on the signal yields, gives an estimate of the bias due to the choice of the background model. The bias measured in each BDT category, as well from pseudodata sets in which the signal injected simultaneously in all event categories, is smaller than 20%. Including these observed deviations as spurious signals leads to a change in the overall uncertainty in measured signal rate of less than 1% and is therefore neglected.

Figure 5 shows the $m_{\mu\mu}$ distributions in each of the ggH categories, in which the signal is extracted by performing a binned maximum-likelihood fit using a DCB function to model the signal contribution, while the background is estimated with the core-pdf method. Table 2 reports the signal composition in each ggH category as well as the HWHM of the expected signal shape. In addition, the estimated number of background events, the observation in data, the $\mathrm{S}/(\mathrm{S}+\mathrm{B})$, and the $\mathrm{S}/\sqrt{\mathrm{B}}$ ratios computed within the HWHM range around the signal peak are also listed.

8 Event categories for VH production

Events considered in the VH category contain at least two muons passing the selection requirements listed in Section 5. In order to ensure mutual exclusivity with the $\mathrm{t\bar{t}H}$ category, events containing at least two b-tagged jets with $p_{\mathrm{T}} > 25\,\mathrm{GeV}$ and $|\eta| < 2.5$ passing the loose WP of the DeepCSV algorithm, or at least one jet passing the medium WP, are discarded. Events are also required to have at least one additional lepton (electron or muon), which is expected from the

Table 2: The product of acceptance and selection efficiency for the different signal production processes, the total expected number of signal events with $m_{\mathrm{H}} = 125.38\,\mathrm{GeV}$, the HWHM of the signal peak, the estimated number of background events and the observation in data within $\pm$ HWHM, and the $\mathrm{S}/(\mathrm{S}+\mathrm{B})$ and the $\mathrm{S}/\sqrt{\mathrm{B}}$ ratios within $\pm$ HWHM, for each of the optimized ggH event categories.

Category	Sig.	ggH (%)	VBF (%)	VH + $\mathrm{t\bar{t}H}$ (%)	HWHM (GeV)	Bkg. in HWHM	$\mathrm{S}/(\mathrm{S}+\mathrm{B})$ (%) in HWHM	$\mathrm{S}/\sqrt{\mathrm{B}}$ in HWHM	Data in HWHM
ggH-cat1	267.6	93.7	2.9	3.4	2.12	86359	0.20	0.60	86632
ggH-cat2	311.5	93.5	3.4	3.1	1.75	46347	0.46	0.98	46393
ggH-cat3	131.4	93.2	4.0	2.8	1.60	12655	0.70	0.80	12738
ggH-cat4	125.6	91.5	5.5	3.0	1.47	8259	1.03	0.96	8377
ggH-cat5	53.8	83.5	14.3	2.2	1.50	1678	2.16	0.91	1711

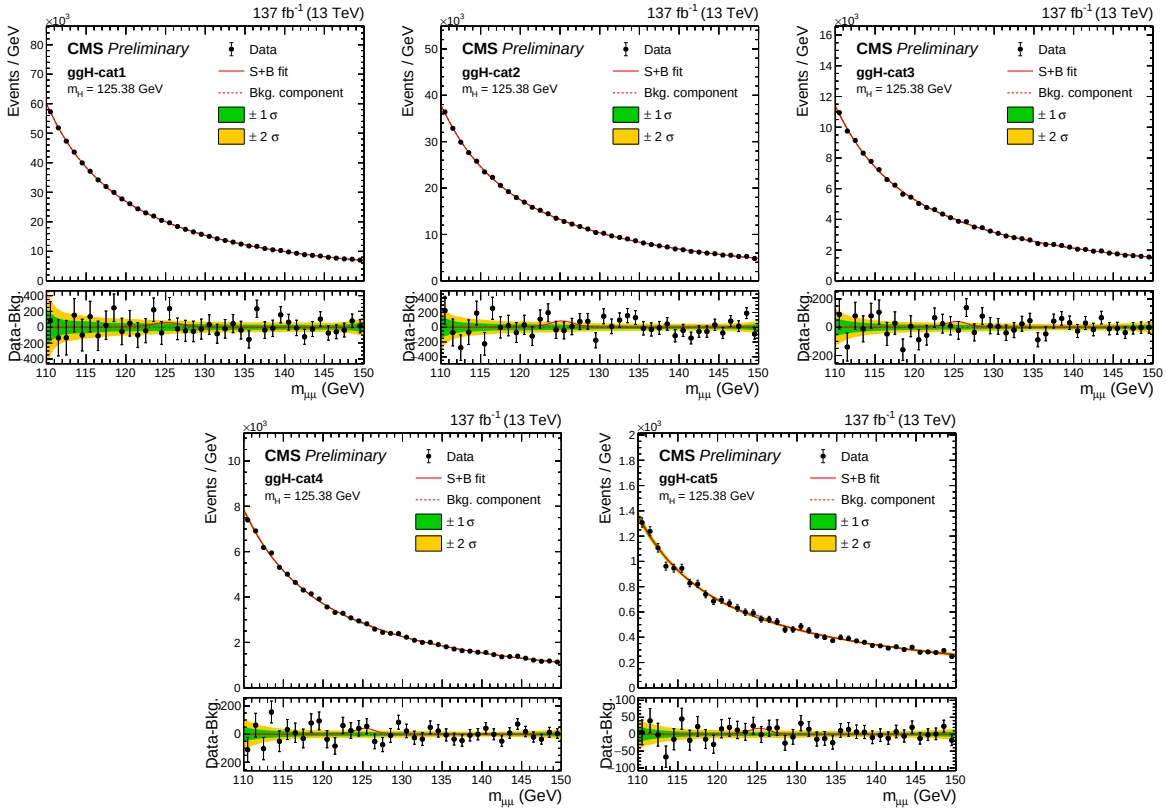

Figure 5: Comparison between the observed data and the total background extracted from a signal-plus-background fit performed across the ggH categories. First row, from left to right: ggH-cat1, ggH-cat2, and ggH-cat3. Second row, from left to right: ggH-cat4 and ggH-cat5. The one (green) and two (yellow) standard deviation bands include the uncertainties in the background component of the fit. The lower panel shows the residuals after background subtraction and the red line indicates the signal with $m_{\mathrm{H}} = 125.38\,\mathrm{GeV}$ extracted from the fit.

leptonic decay of the W or Z boson. The additional muons (electrons) must have $p_{\mathrm{T}} > 20\,\mathrm{GeV}$, $|\eta| < 2.4$ (2.5), and pass certain isolation and identification requirements with an average efficiency of 95 (90)%. Furthermore, electrons and muons are required to pass the medium WP of a multivariate discriminant developed to identify and suppress nonprompt leptons [89], with a selection efficiency of about 90 (85)% per prompt muon (electron).

Events containing exactly one additional lepton belong to the WH category, which targets signal events where the Higgs boson is produced in association with a leptonically decaying W bo-

son. If the additional lepton is a muon, the two pairs of oppositely charged muons are required to have $m_{\mu\mu} > 12\,\text{GeV}$ to suppress background events from quarkonium decays. Moreover, neither of the two oppositely charged muon pairs can have an invariant mass consistent with m_Z within 10 GeV. Finally, at least one of these two muon pairs must have $m_{\mu\mu}$ in the range 110–150 GeV. If both $m_{\mu\mu}$ pairs satisfy this criterion, the highest-p_T pair is considered as the Higgs boson candidate. If the additional lepton is an electron, the only requirement imposed is that $110 < m_{\mu\mu} < 150\,\text{GeV}$.

The ZH category targets signal events where the Higgs boson is produced in association with a Z boson that decays to a pair of electrons or muons. Events in the ZH category are therefore required to contain four leptons, with a combined lepton number and electric charge of zero. As in the WH category, the invariant mass of each pair of same-flavour opposite-charge leptons is required to be greater than 12 GeV. An event is rejected if it does not contain exactly one pair of same-flavour and oppositely charged leptons with invariant mass compatible with the Z boson within 10 (20) GeV for muon (electron) pairs. In addition, each event must contain one oppositely charged muon pair satisfying $110 < m_{\mu\mu} < 150\,\text{GeV}$. For events with four muons, the muon pair with $m_{\mu\mu}$ closer to m_Z is chosen as the Z boson candidate, while the other muon pair is selected as the Higgs boson candidate. A summary of the selection criteria applied in the WH and ZH production categories is reported in Table 3.

Table 3: Summary of the kinematic selection used to define the WH and ZH production categories.

Selection	WH leptonic		ZH leptonic	
	$\mu\mu\mu$	$\mu\mu$e	4μ	2μ2e
Number of loose (medium) b-tagged jets	≤ 1 (0)	≤ 1 (0)	≤ 1 (0)	≤ 1 (0)
N(μ) passing id.+iso.	3	2	4	2
N(e) passing id.+iso.	0	1	0	2
Lepton charge	$\sum q(\ell) = \pm 1$		$\sum q(\ell) = 0$	
Low mass resonance veto	$m_{\ell\ell} > 12\,\text{GeV}$			
N($\mu^+\mu^-$) pairs with $110 < m_{\mu\mu} < 150\,\text{GeV}$	≥ 1	1	≥ 1	1
N($\mu^+\mu^-$) pairs with $\lvert m_{\mu\mu} - m_Z \rvert < 10\,\text{GeV}\rvert$	0	0	1	0
N(e^+e^-) pairs with $\lvert m_{ee} - m_Z \rvert < 20\,\text{GeV}\rvert$	0	0	0	1

Two BDT discriminants are trained to discriminate between signal and background events in the WH and ZH categories. The input variables are selected such that the BDT outputs are not significantly correlated with the $m_{\mu\mu}$ of the Higgs boson candidate. This is required by the chosen analysis strategy, which is analogous to that adopted in the ggH category. The impact of the $m_{\mu\mu}$ resolution, which evolves as a function of muon p_T and η, is taken into account during the BDT training by applying weights to the simulated signal events that are inversely proportional to the per-event uncertainty on the measured $m_{\mu\mu}$, following the same strategy described in Section 7. The BDT discriminant used in the WH category takes as inputs several variables that exploit the kinematic features of the three leptons in the event, as well as the p_T^{miss}. These variables include the full kinematics, apart from the invariant mass, of the dimuon system corresponding to the Higgs boson candidate. In addition, angular distances in ϕ and η between the additional lepton (ℓ_W) and the Higgs boson candidate, ℓ_W and both muons from the Higgs boson candidate, and ℓ_W and $\vec{H}_T^{miss}$ are considered. The $\vec{H}_T^{miss}$ is defined as the negative vector sum of the p_T of all jets in the event with $p_T > 30\,\text{GeV}$ and $|\eta| < 4.7$. Finally, the transverse mass of the combined ℓ_W and $\vec{H}_T^{miss}$ system and the flavour and the p_T of ℓ_W are added as inputs to the BDT. The particular kinematic properties of the ℓ_W and H_T^{miss} distributions for signal events enable a large suppression of the residual DY background. The

BDT discriminant trained in the ZH category considers several input observables constructed from the lepton pair corresponding to the Z boson candidate and the muon pair considered as the Higgs boson candidate. The flavour of the lepton pair associated to the Z boson decay is also included as an input variable.

Figure 6 shows the output of the BDT classifiers in the WH (left) and ZH (right) categories. Based on these outputs, events in the WH category are further divided into three subcategories termed "WH-cat1", "WH-cat2", and "WH-cat3". Similarly, events in the ZH category are divided into two subcategories, labelled "ZH-cat1" and "ZH-cat2". The boundaries of these categories, defined in terms of the BDT discriminant and indicated in Fig. 6 by black dashed vertical lines, are chosen via an optimization strategy analogous to that described in Section 7 for the ggH category. In this category, the BWZ function is used to estimate the total background instead of the mBW.

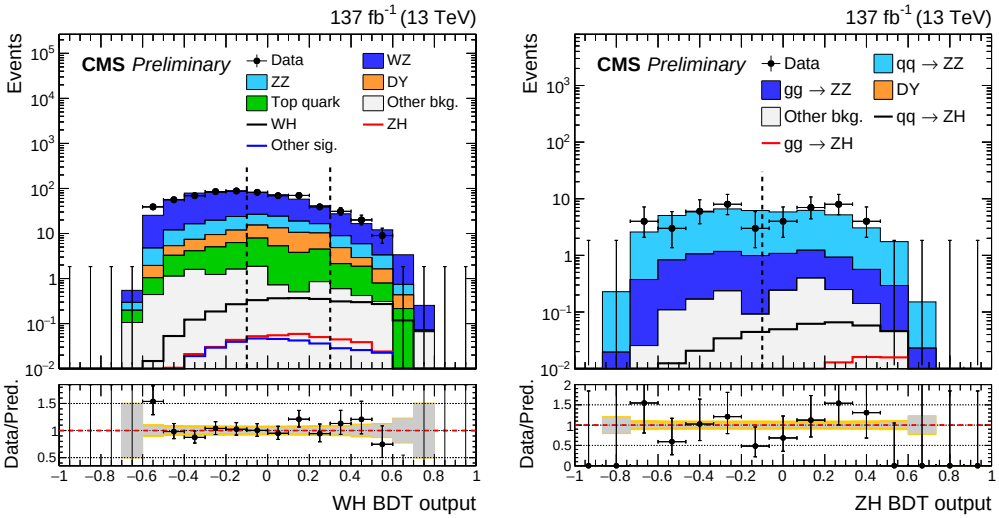

Figure 6: The observed BDT output distribution in the WH (left) and ZH (right) categories compared to the prediction from the simulation of various SM background processes. Signal distributions expected from different production modes of the 125 GeV Higgs boson are overlaid. The description of the ratio panel is the same as in Fig. 4. The dashed vertical lines indicate the boundaries of the optimized event categories.

The systematic uncertainties considered in this analysis account for possible mismodeling in the signal shape or rate. The shape of the reconstructed Higgs boson resonance, modelled using the DCB function defined in Eq.(3), is affected by the uncertainty in the muon momentum scale and resolution. Uncertainties in the calibration of these values are propagated to the shape of the $m_{\mu\mu}$ distribution of the Higgs boson, yielding variations of up to 0.2% in the peak position and up to 10% in the width. Experimental systematic uncertainties from the measurement of the electron and muon selection efficiencies (1–3% per category), jet energy scale and resolution (0.5–2% per category), the efficiency of vetoing b quark jets (1–3% per category), and the pileup model (0.5–2% per category) affect the predicted signal rate. Furthermore, theoretical uncertainties in the prediction of the Higgs boson production cross section, decay rate, and acceptance are also considered. Rate uncertainties are taken into account in the signal extraction as nuisance parameters acting on the relative signal yield with log-normal constraints.

Figure 7 show the $m_{\mu\mu}$ distributions in the WH (first row) and ZH (second row) event categories. The signal is extracted via a binned maximum-likelihood fit in each event category, where the signal is modelled with a DCB function and the background is modelled with the

BWZGamma function in WH-cat1, as defined in Eq.(6) and the BWZ function in the remaining categories, as defined in Eq.(5). In order to estimate the potential bias arising from the choice of the parametric function used to model the background, alternative functions able to fit the data with a χ^2 p-value larger than 5% are considered. These include Bernstein polynomials, series of exponentials, and series of power laws. In each event category, background-only fits to the data are performed with each function listed above. From each of these fits, pseudodata sets are generated taking into account the uncertainties in the fit parameters and their correlations, and injecting a certain amount of signal events. A signal-plus-background fit is then performed on these pseudodata sets using either the BWZGamma (in WH-cat1) or the BWZ (in the remaining categories) function to model the background. The corresponding bias is observed to be small and is therefore neglected in the signal extraction. The chosen functions maximize the expected sensitivity to the 125 GeV Higgs boson. Finally, Table 4 reports the signal composition in the WH and ZH categories, along with the HWHM of the expected signal shape. In addition, the estimated number of background events, the $S/(S+B)$ and $S/\sqrt{B}$ ratios, and the observation in data within the HWHM of the signal peak are also listed.

Table 4: The product of acceptance and selection efficiency for the different signal production processes, the total expected number of signal events with $m_H = 125.38$ GeV, the HWHM of the signal peak, the estimated number of background events and the observed number of events within $\pm$ HWHM, and the $S/(S+B)$ and the $S/\sqrt{B}$ ratios computed within the HWHM of the signal peak for each of the optimized event categories defined along the WH and ZH BDT outputs.

Category	Sig.	WH (%)	qqZH (%)	ggZH (%)	$t\bar{t}H + tH$ (%)	HWHM (GeV)	Bkg. in HWHM	$S/(S+B)$ (%) in HWHM	$S/\sqrt{B}$ in HWHM	Data in HWHM
WH-cat1	0.82	76.2	9.6	1.6	12.6	2.00	32.0	1.54	0.09	34
WH-cat2	1.72	80.1	9.1	1.5	9.3	1.80	23.1	4.50	0.23	27
WH-cat3	1.14	85.7	6.7	1.8	4.8	1.90	5.48	12.6	0.35	4
ZH-cat1	0.11	—	82.8	17.2	—	2.07	2.05	3.29	0.05	4
ZH-cat2	0.31	—	79.6	20.4	—	1.80	2.19	8.98	0.14	4

9 Event categories for ttH production

The $t\bar{t}H$ process has the smallest cross section among the main Higgs boson production modes at the LHC. However, the presence of a pair of top quarks in addition to the Higgs boson helps to reduce the background to a level that is comparable to the expected signal rate. Top quarks decay predominantly into a b quark and a W boson, which then decays either to a lepton and a neutrino ($\mathcal{B}(W \to \ell\nu) \approx 0.33$), or into two quarks ($\mathcal{B}(W \to qq') \approx 0.66$). Therefore, events in the $t\bar{t}H$ category are required to contain at least two jets passing the loose WP of the DeepCSV algorithm, or at least one jet passing the medium WP. This requirement suppresses background processes not enriched in jets originating from the hadronization of b quarks, such as DY and diboson production. This selection also ensures mutual exclusivity between the $t\bar{t}H$ analysis and the other production categories considered in this search.

The $t\bar{t}H$ signal events may contain additional leptons, depending on the decay of the top quarks. The muon isolation definition is modified compared to the baseline event selection detailed in Section 5. In order to increase the signal selection efficiency in events with large hadronic activity, the isolation requirement is relaxed to be less than 40% of the muon p_T. In addition, the isolation cone size decreases dynamically with the muon p_T ($R = 0.2$ for $p_T < 50$ GeV, $R = 10/p_T$ for $50 < p_T < 200$ GeV, and $R = 0.05$ for $p_T > 200$ GeV), following the same approach used in Ref. [90]. Electron candidates are required to have $p_T > 20$ GeV, $|\eta| < 2.5$,

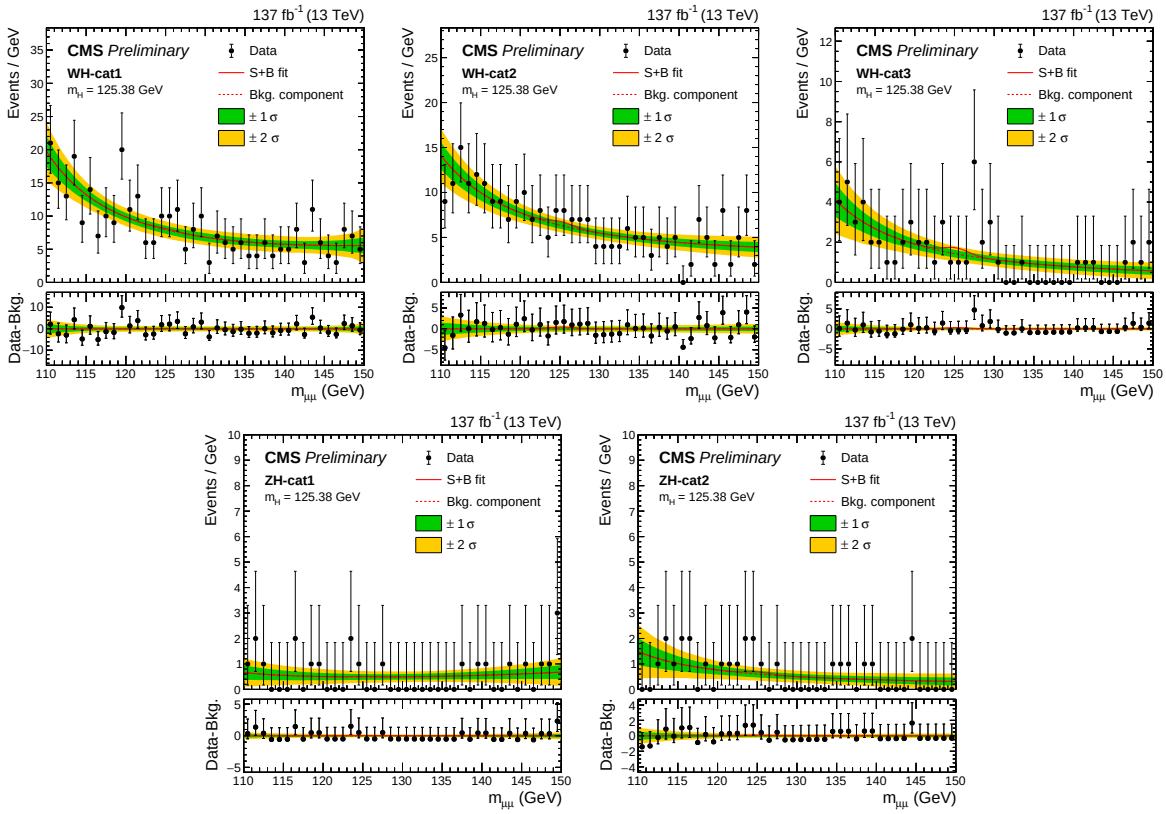

Figure 7: Comparison between the observed data and the total background extracted from a signal-plus-background fit performed across WH (first row) and ZH (second row) event categories. First row, from left to right: WH-cat1, WH-cat2, and WH-cat3. Second row, from left to right: ZH-cat1 and ZH-cat2. The one (green) and two (yellow) standard deviation bands include the uncertainties in the background component of the fit. The lower panel shows the residuals after the background subtraction, where the red line indicates the signal with $m_{\mathrm{H}} = 125.38\,\mathrm{GeV}$ extracted from the fit.

and to pass identification requirements imposed on the properties of the ECAL energy cluster associated with the electron track and the consistency between the electron momentum measured by the inner tracker and its ECAL energy deposit. Each electron is also required to be isolated following the same strategy as for muons, and the magnitude of the transverse and longitudinal impact parameters must be smaller than 0.05 and 0.1 cm, respectively. With these requirements, electrons are selected with an efficiency of about 90%. In order to suppress backgrounds containing nonprompt leptons produced in the decay of heavy quarks, electrons and muons are rejected when the jet with $p_{\mathrm{T}} > 15\,\mathrm{GeV}$ that is nearest to the lepton in ΔR separation is b-tagged according to the DeepCSV medium WP. Furthermore, electrons and muons are required to pass the medium WP of a multivariate lepton identification discriminant specifically designed to reject nonprompt leptons [89], resulting in a selection efficiency of about 90 (85)% per prompt muon (electron).

Events with exactly two oppositely charged muons with $110 < m_{\mu\mu} < 150\,\mathrm{GeV}$ and at least three jets in the final state with invariant mass (m_{jjj}) between 100 and 300 GeV belong to the $\mathrm{t\bar{t}H}$ hadronic category. Each jet must have $p_{\mathrm{T}} > 25\,\mathrm{GeV}$ and $|\eta| < 4.7$. Events with one or two additional leptons in the final state are grouped in the $\mathrm{t\bar{t}H}$ leptonic category, in which at least one of the two top quarks decays leptonically. An event in the $\mathrm{t\bar{t}H}$ leptonic category containing three (four) leptons is further required to have the net sum of the lepton electric charges equal

to one (zero). In the case of events with more than one pair of oppositely charged muons with $110 < m_{\mu\mu} < 150\,\text{GeV}$, the pair with the largest dimuon p_T is chosen as the Higgs boson candidate. The invariant mass of each pair of same-flavour opposite-charge leptons is required to be greater than 12 GeV to suppress backgrounds arising from quarkonium decays. An event is vetoed if it contains a pair of oppositely charged electrons or muons with an invariant mass in the range 81–101 GeV, consistent with the decay of an on-shell Z boson. A summary of the selection criteria used to define the $\text{t}\bar{\text{t}}\text{H}$ hadronic and leptonic categories is reported in Table 5.

Table 5: Summary of the kinematic selections used to define the $\text{t}\bar{\text{t}}\text{H}$ hadronic and leptonic production categories.

Selection	$\text{t}\bar{\text{t}}\text{H}$ hadronic	$\text{t}\bar{\text{t}}\text{H}$ leptonic
Number of b quark jets	> 0 medium or > 1 loose b-tagged jets	
Number of leptons	2	3 or 4
Lepton charge	$\sum q(\ell) = 0$	$\text{N}(\ell) = 3\ (4) \rightarrow \sum q(\ell) = \pm 1\ (0)$
Jet multiplicity ($p_\text{T} > 25\,\text{GeV}$, $\lvert\eta\rvert < 4.7$)	≥ 3	≥ 2
Leading jet p_T	$> 50\,\text{GeV}$	$> 35\,\text{GeV}$
Jet triplet mass	$100 < m_{jjj} < 300\,\text{GeV}$	—
Z mass veto	—	$\lvert m_{\ell\ell} - m_\text{Z}\rvert > 10\,\text{GeV}$
Low mass resonance veto	—	$m_{\ell\ell} > 12\,\text{GeV}$

The dominant background in the $\text{t}\bar{\text{t}}\text{H}$ hadronic category is from fully leptonic $\text{t}\bar{\text{t}}$ decays, while the main backgrounds in the $\text{t}\bar{\text{t}}\text{H}$ leptonic category are due to $\text{t}\bar{\text{t}}\text{Z}$ and $\text{t}\bar{\text{t}}$ processes. In order to obtain an optimal discrimination between the $\text{t}\bar{\text{t}}\text{H}$ signal and the expected backgrounds, BDT-based multivariate discriminants are trained in both the $\text{t}\bar{\text{t}}\text{H}$ hadronic and leptonic categories. The input variables are chosen to account for both the kinematics of the dimuon system and the properties of the top quark decay products, while ensuring that the BDT outputs remain uncorrelated with $m_{\mu\mu}$. A common set of observables is used as input to the two BDT discriminants. These include variables that characterize the production and decay of the Higgs boson candidate, namely $p_\text{T}(\mu\mu)$, $y^{\mu\mu}$, ϕ_CS, and $\cos\theta_\text{CS}$. In addition, the η of the two muons and the p_T relative to $m_{\mu\mu}$ are also considered. In order to account for the large hadronic activity in $\text{t}\bar{\text{t}}\text{H}$ signal events, the p_T and η of the three leading jets, the maximum DeepCSV value of jets not in overlap with leptons, the number of jets, and the scalar (vectorial) p_T sum H_T (H_T^miss) of all identified leptons and jets with $\lvert\eta\rvert < 2.5$ are included. The p_T^miss is also considered along with the $\Delta\zeta$ variable, which is defined as the projection of the $\vec{p}_\text{T}^\text{miss}$ on the bisector of the dimuon system in the transverse plane. The BDT discriminants learn the $m_{\mu\mu}$ resolution because signal events are weighted during the training with the inverse of the per-event uncertainty on the measured $m_{\mu\mu}$, following the same approach used in the ggH and VH categories.

In the $\text{t}\bar{\text{t}}\text{H}$ leptonic category, several additional variables are used in the BDT discriminant that target the kinematic properties of a leptonic top quark decay. These include the azimuthal separation between the Higgs boson candidate and the highest-p_T additional lepton, the invariant mass formed by the leading additional lepton and the jet with the highest DeepCSV score, and the transverse mass formed by the additional lepton and $\vec{E}_\text{T}^\text{miss}$ in the event. In the $\text{t}\bar{\text{t}}\text{H}$ hadronic category, the resolved hadronic top tagger (RHTT), which combines a kinematic fit and a BDT-based multivariate discriminant, is used to identify top quark decays to three resolved jets. The jet triplet with the highest RHTT score is selected as a hadronic top quark candidate. The corresponding RHTT score is used as input to the BDT discriminant. Furthermore, the p_T of the top quark candidate and the p_T balance of the top quark and the muon pair are also considered.

Figure 8 shows the output of the BDT discriminant in the $\text{t}\bar{\text{t}}\text{H}$ hadronic (left) and leptonic (right) categories. The high BDT score region of the $\text{t}\bar{\text{t}}\text{H}$ hadronic category is enriched in events with

large jet multiplicity, where the $t\bar{t}$ and DY background predictions rely on a significant number of jets from the PS and are known to not entirely reproduce the data. The signal prediction, however, relies largely on jets from the ME. Since the background prediction is extracted from data, the observed differences between data and background simulation do not affect the fit result. Based on the BDT output, events in the $t\bar{t}H$ leptonic category are further divided into two subcategories, termed "$t\bar{t}H$lep-cat1" and "$t\bar{t}H$lep-cat2". Similarly, events in the $t\bar{t}H$ hadronic category are divided into three subcategories labelled "$t\bar{t}H$had-cat1", "$t\bar{t}H$had-cat2", and "$t\bar{t}H$had-cat3". The BDT score boundaries of these event categories, indicated in Fig. 8 by black dashed vertical lines, are optimized following the same strategy described in Section 7 for the ggH category. Bernstein polynomials are chosen for the analytical function used to model the background in the "$t\bar{t}H$had-cat1" and "$t\bar{t}H$had-cat2", while a sum of two exponentials and a single exponential functions are used in the "$t\bar{t}H$had-cat3" and $t\bar{t}H$ leptonic categories, respectively.

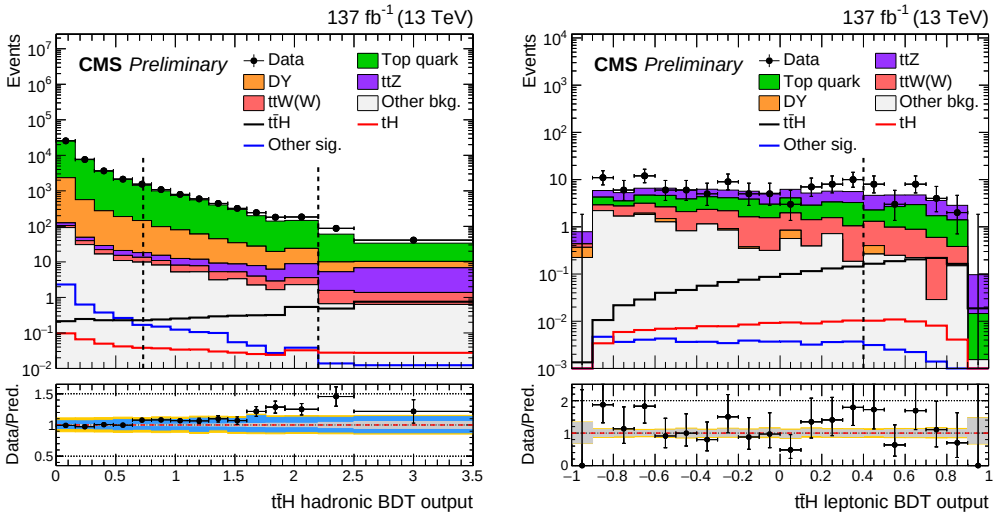

Figure 8: The observed BDT output distribution in the $t\bar{t}H$ hadronic (left) and leptonic (right) categories compared to the prediction from the simulation of various SM background processes. Signal distributions expected from different production modes of the 125 GeV Higgs boson are overlaid. The dashed vertical lines indicate the boundaries of the optimized event categories. The description of the ratio panels is the same as in Fig. 4.

The systematic uncertainties considered account for possible mismodelling of the signal shape or rate. Uncertainties in the calibration of muon momentum scale and resolution are propagated to the shape of the signal $m_{\mu\mu}$ distribution, yielding variations of up to 0.1% in the peak position and up to 10% in width. Experimental systematic uncertainties from the measurement of the electron and muon selection efficiencies (0.5–1.5% per category), muon momentum scale and resolution (0.1–0.8% per category), jet energy scale and resolution (2–6% per category), efficiency of identifying b quark jets (1–3% per category), and modelling of the pileup conditions (0.2–1% per category) affect the predicted signal rate. Furthermore, theoretical uncertainties in the prediction of the Higgs boson production cross sections, decay rate, and acceptance are also included, as already described for the ggH, VBF, and VH analyses. Rate uncertainties included in the signal extraction as nuisance parameters acting on the relative signal yield with log-normal priors.

Figure 9 shows the $m_{\mu\mu}$ distributions in the $t\bar{t}H$ hadronic (first row) and leptonic (second row) event categories. The signal is extracted by performing a binned maximum-likelihood fit to these $m_{\mu\mu}$ distributions, where the signal is modelled using the DCB function and the back-

ground is modelled using a second-order Bernstein polynomial or a sum of two exponentials (single exponential) in the $t\bar{t}H$ hadronic (leptonic) categories. The potential bias due to the choice of the parametric function used to model the background is estimated using the same procedure employed in the VH analysis, detailed in Section 8. The set of analytical functional forms considered in the bias studies includes series of exponentials, Bernstein polynomials, and series of power laws. The chosen parametrization maximizes the expected sensitivity without introducing a significant bias in the measured signal yield. Table 6 reports the signal composition of each $t\bar{t}H$ category, along with the HWHM of the expected signal shape. In addition, the estimated number of background events, the observation in data, and the $S/(S+B)$ and $S/\sqrt{B}$ ratios within the HWHM of the signal shape are shown.

Table 6: The product of acceptance and selection efficiency for the different signal production processes, the total expected number of signal events with $m_H = 125.38$ GeV, the HWHM of the signal peak, the estimated number of background events and the observed number of events within $\pm$ HWHM, and the $S/(S+B)$ and $S/\sqrt{B}$ ratios computed within the HWHM of the signal peak, for each of the optimized event categories defined along the $t\bar{t}H$ hadronic and leptonic BDT outputs.

Category	Sig.	$t\bar{t}H$ (%)	ggH (%)	VH (%)	tH (%)	VBF+$b\bar{b}H$ (%)	HWHM (GeV)	Bkg. in HWHM	$S/(S+B)$ (%) in HWHM	$S/\sqrt{B}$ in HWHM	Data in HWHM
$t\bar{t}$Hhad-cat1	6.87	32.3	40.3	17.2	6.2	4.0	1.85	4298	1.07	0.07	4251
$t\bar{t}$Hhad-cat2	1.62	84.3	3.8	5.6	6.2	—	1.81	82.0	1.32	0.12	89
$t\bar{t}$Hhad-cat3	1.33	94.0	0.3	1.3	4.2	0.2	1.80	12.3	6.87	0.26	12
$t\bar{t}$Hlep-cat1	1.06	85.8	—	4.7	9.5	—	1.92	9.00	7.09	0.22	13
$t\bar{t}$Hlep-cat2	0.99	94.7	—	1.0	4.3	—	1.75	2.08	24.5	0.47	4

10 Results

A simultaneous fit is performed across all the event categories, with a single overall signal strength modifier (μ) with a flat prior. The signal strength modifier is defined as the ratio between the observed Higgs boson rate in the $H \to \mu\mu$ decay channel and the SM expectation, $\mu = (\sigma\mathcal{B}(H \to \mu\mu))_{obs}/(\sigma\mathcal{B}(H \to \mu\mu))_{SM}$. The relative contributions from the different Higgs boson production modes are fixed to the SM prediction within uncertainties. Confidence intervals on the signal strength are estimated using a profile likelihood ratio test statistic [91], in which systematic uncertainties are modelled as nuisance parameters following a frequentist approach [92]. The profile likelihood ratio is defined as

$$q = -2 \ln \frac{\mathcal{L}(\text{data}|\mu, \hat{\theta}_a)}{\mathcal{L}(\text{data}|\hat{\mu}, \hat{\theta})},$$

where $\hat{\mu}$ represents the value of the signal strength that maximizes the likelihood $\mathcal{L}$ for the data and $\hat{\theta}$ ($\hat{\theta}_a$) denotes the best-fit estimate for the nuisance parameters given a freely floating (fixed) value of μ. Theoretical uncertainties affecting the signal prediction are correlated among the event categories. Similarly, experimental uncertainties in the measurement of the integrated luminosity in each year, jet energy scale and resolution, modelling of the pileup conditions, and selection efficiencies of muons and electrons are also correlated across categories. Uncertainties in the b quark jet identification are uncorrelated. Because of the different analysis strategy employed in the VBF category, the acceptance uncertainties from the muon energy scale and resolution are correlated only among the ggH, WH, ZH, and $t\bar{t}H$ categories. Furthermore, their effect on the position and width of the signal peak are assumed to be uncorrelated across event categories.

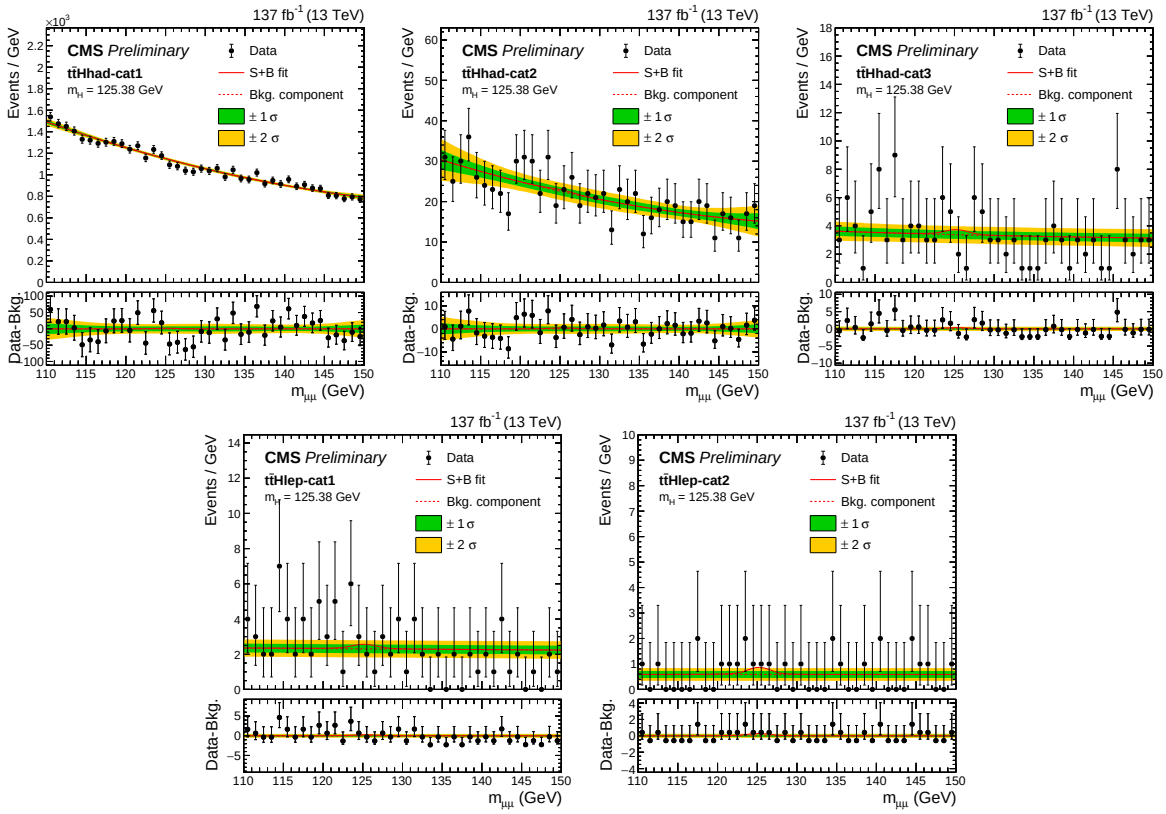

Figure 9: Comparison between the observed data and the total background extracted from a signal-plus-background fit performed across $t\bar{t}H$ hadronic (first row) and leptonic (second row) event categories. First row, from left to right: $t\bar{t}$Hhad-cat1, $t\bar{t}$Hhad-cat2, and $t\bar{t}$Hhad-cat3. Second row, from left to right: $t\bar{t}$Hlep-cat1 and $t\bar{t}$Hlep-cat2. The one (green) and two (yellow) standard deviation bands include the uncertainties in the background component of the fit. The lower panel shows the residuals after the background subtraction, where the red line indicates the signal with $m_{\mathrm{H}} = 125.38\,\mathrm{GeV}$ extracted from the fit.

An unbiased mass distribution representative of the fit result in the VBF category is obtained by weighting both simulated and data events from the VBF-SR and VBF-SB regions by the per-event $\mathrm{S}/(\mathrm{S}+\mathrm{B})$ ratio, computed as a function of the mass-decorrelated DNN output, defined in Section 6, for events within $m_{\mu\mu} = 125.38\,\mathrm{GeV} \pm \mathrm{HWHM}$. The best-fit estimates for the nuisance parameters and signal strength are propagated to the $m_{\mu\mu}$ distribution. Figure 10 (left) shows the observed and predicted weighted $m_{\mu\mu}$ distributions for events in the VBF-SB and VBF-SR regions, combining 2016, 2017, and 2018 data. The lower panel shows the residuals between the data and the post-fit background prediction, along with the post-fit uncertainty obtained from the background-only fit. The best-fit signal contribution with $m_{\mathrm{H}} = 125.38\,\mathrm{GeV}$ is indicated by the blue line. An excess is observed in the weighted data distribution that is consistent with the expected resonant mass distribution for the signal with m_{H} near 125 GeV and compatible with the excess observed at high DNN score in Fig. 3. The signal and background distributions are then interpolated with a spline function in order to obtain a continuous spectrum that can be summed with the analytical fit results in the ggH, WH, ZH, and $t\bar{t}H$ categories. Figure 10 (right) shows the $m_{\mu\mu}$ distribution for the weighted combination of all event categories. The ggH, VH, and $t\bar{t}H$ categories are weighted proportionally to the corresponding $\mathrm{S}/(\mathrm{S}+\mathrm{B})$ ratio, where S and B are the number of expected signal and background events with mass within $\pm\,\mathrm{HWHM}$ of the expected signal peak with $m_{\mathrm{H}} = 125.38\,\mathrm{GeV}$. The upper panel is dominated

by the ggH categories with many data events but relatively small $S/(S+B)$. The lower panel shows the residuals after background subtraction, with the best-fit SM signal contribution with $m_H = 125.38$ GeV indicated by the red line. An excess of events over the background-only expectation is observed near $m_{\mu\mu} = 125$ GeV.

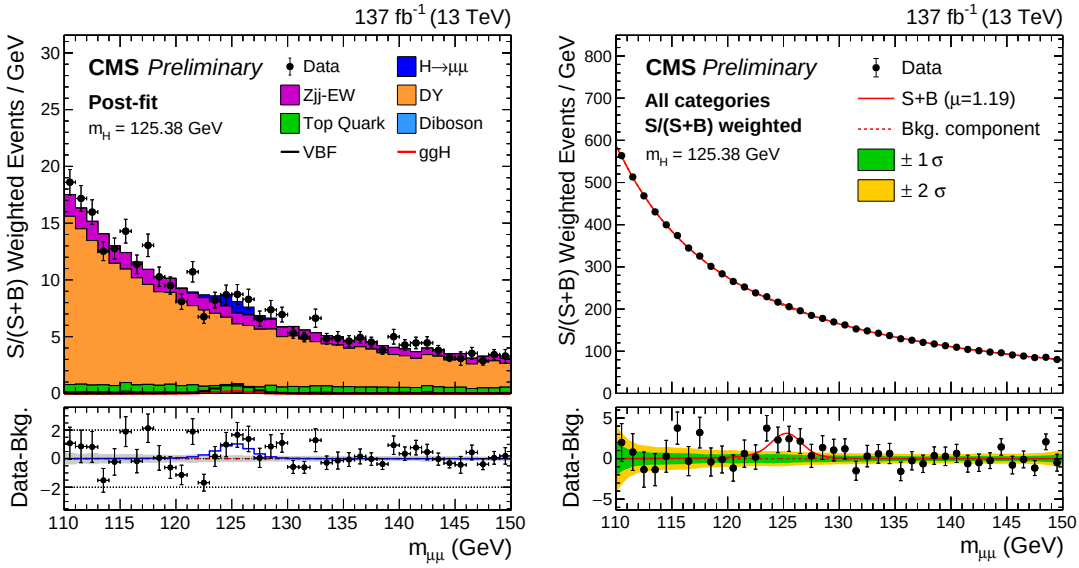

Figure 10: Left: the $m_{\mu\mu}$ distribution for the weighted combination of VBF-SB and VBF-SR events. Each event is weighted proportionally to the $S/(S+B)$ ratio, calculated as a function of the mass-decorrelated DNN output. The lower panel shows the residuals after subtracting the background prediction from the signal-plus-background fit. The best-fit $H \to \mu\mu$ signal contribution is indicated by the blue line, and the grey band indicates the total background uncertainty from the background-only fit. Right: the $m_{\mu\mu}$ distribution for the weighted combination of all event categories. The upper panel is dominated by the ggH categories with many data events but relatively small $S/(S+B)$. The lower panel shows the residuals after background subtraction, with the best-fit SM $H \to \mu\mu$ signal contribution with $m_H = 125.38$ GeV indicated by the red line.

The local p-value quantifies the probability for the background to produce a fluctuation larger than the apparent signal observed in the search region. Figure 11 (left) shows the observed local p-value for the combined fit and for each individual production category as a function of m_H in a 5 GeV window around the expected Higgs boson mass. Figure 11 (right) shows the expected p-values computed for the combined fit and for each production category on an Asimov data set [91] generated from the background expectation obtained from the signal-plus-background fit injecting a signal at $m_H = 125.38$ GeV. The solid markers indicate the mass points for which the observed p-values are computed. In the ggH, VH, and $t\bar{t}H$ categories, in order to evaluate p-values for masses different than 125 GeV, signal models are derived using additional alternative $H \to \mu\mu$ signal samples generated with m_H fixed to 120 and 130 GeV. Signal shape parameters and the expected rate for each production mode in each event category are then interpolated within $120 < m_H < 130$ GeV, providing a signal model for any mass value in the $m_H = 125 \pm 5$ GeV range. A different strategy is employed in the VBF category since $m_{\mu\mu}$ is a DNN input variable. As described in Section 6, the DNN output can be decorrelated from the dimuon mass information by fixing the value of $m_{\mu\mu}$ to 125 GeV. Therefore, a potential signal with mass m' different from 125 GeV can be extracted by fitting the data with an alternative set of signal and background templates, obtained by shifting the mass value used as input to the DNN evaluation by $\Delta m = 125\,\text{GeV} - m'$ and adjusting the expected signal yields

by the corresponding differences in the production cross section and decay rate. Variations in the acceptance per DNN bin as a function of Δm are found to be negligible in the mass range of interest. This procedure is also applied to the data, yielding for each tested mass hypothesis a different observed DNN distribution to fit. Throughout the explored mass range, $120 < m_{\mathrm{H}} < 130$ GeV, the VBF category has the highest expected sensitivity to $\mathrm{H} \to \mu\mu$ decays, followed by the ggH, $\mathrm{t\bar{t}H}$, and VH categories, respectively. The observed (expected for $\mu = 1$) significance at $m_{\mathrm{H}} = 125.38$ GeV of the incompatibility with the background-only hypothesis is 3.0 (2.5) σ. Fluctuations in the observed p-value of the VBF category and for the combined fit are due to the nature of the signal extraction fit used in the VBF analysis. When evaluating the DNN for each tested mass point, event migrations in data between neighbouring bins in the high score DNN region produce discrete variations in the observed p-value. The 95% CL upper limit (UL) on the signal strength, computed with the asymptotic $\mathrm{CL_s}$ criterion [91, 93, 94], is also derived from the combined fit performed across all event categories. The observed (expected for $\mu = 0$) upper limit on μ at 95% CL for $m_{\mathrm{H}} = 125.38$ GeV is 1.9 (0.8).

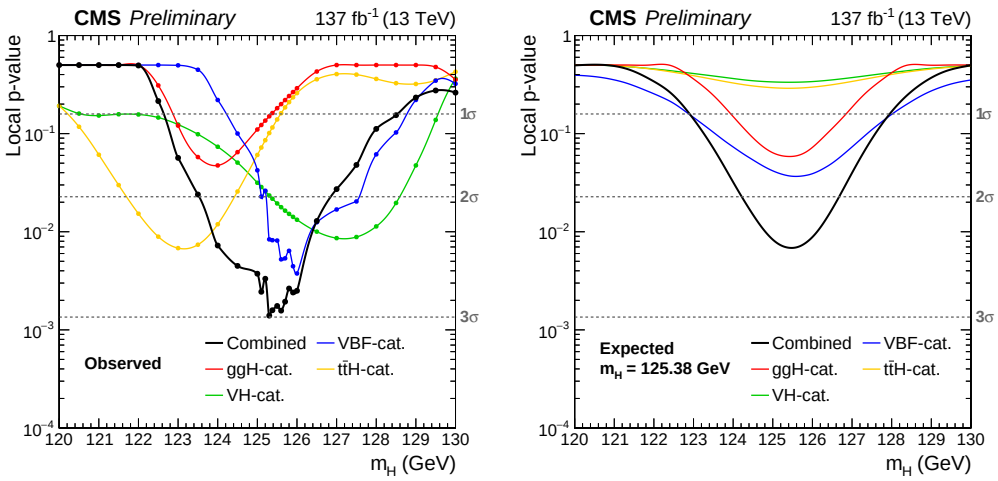

Figure 11: Left: observed local p-values as a function of m_{H}, extracted from the combined fit as well as from each individual production category, are shown. Right: the expected p-values are calculated using the background expectation obtained from the signal-plus-background fit and injecting a signal with $m_{\mathrm{H}} = 125.38$ GeV and $\mu = 1$.

The best-fit signal strength for the Higgs boson with mass of 125.38 GeV, and the corresponding 68% CL interval, is $\hat{\mu} = 1.19\ ^{+0.41}_{-0.39}\,(\mathrm{stat})^{+0.17}_{-0.16}\,(\mathrm{syst})$. Assuming SM production cross sections for the various modes, the $\mathrm{H} \to \mu\mu$ branching fraction is constrained at 95% CL to be within $0.8 \times 10^{-4} < \mathcal{B}(\mathrm{H} \to \mu\mu) < 4.5 \times 10^{-4}$. The statistical component of the post-fit uncertainty is separated by performing a likelihood scan as a function of μ in which systematic uncertainties are removed. The systematic uncertainty component is then taken as the difference in quadrature between the total and the statistical uncertainties. The individual contributions to the uncertainty in the measured signal strength from experimental uncertainties, the limited size of the simulated samples, and theory uncertainties are also evaluated following a similar procedure. The individual uncertainty components are summarized in Table 7. The uncertainty in the measured signal rate is dominated by the limited number of data events.

Figure 12 (left) reports a summary of the best-fit values for the signal strength and the corresponding 68% CL intervals obtained from a profile likelihood scan in each production category. A likelihood scan is performed in which the four main Higgs boson production mechanisms are associated to either fermion (ggH and $\mathrm{t\bar{t}H}$) or vector boson (VBF and VH) couplings. Two signal strength modifiers, denoted as $\mu_{\mathrm{ggH,t\bar{t}H}}$ and $\mu_{\mathrm{VBF,VH}}$, are varied independently as un-

constrained parameters in the fit. Figure 12 (right) shows the 1σ and 2σ contours, computed as variations around the likelihood maximum for $m_{\mathrm{H}} = 125.38$ GeV, for the signal strength modifiers $\mu_{\mathrm{ggH,t\bar{t}H}}$ and $\mu_{\mathrm{VBF,VH}}$. The best-fit values for these parameters are $\hat{\mu}_{\mathrm{ggH,t\bar{t}H}} = 0.66^{+0.67}_{-0.66}$ and $\hat{\mu}_{\mathrm{VBF,VH}} = 1.84^{+0.89}_{-0.77}$, consistent with the SM expectation.

Table 7: Major sources of uncertainty in the measurement of the signal strength μ and their impact. The total post-fit uncertainty on μ is separated into four components: statistical, size of the simulated samples, experimental, and theoretical.

Uncertainty source	$\Delta\mu$	
Total uncertainty	$+0.44$	-0.42
Statistical uncertainty	$+0.41$	-0.39
Total systematic uncertainty	$+0.17$	-0.16
Size of simulated samples	$+0.07$	-0.06
Total experimental uncertainty	$+0.12$	-0.10
Total theoretical uncertainty	$+0.10$	-0.11

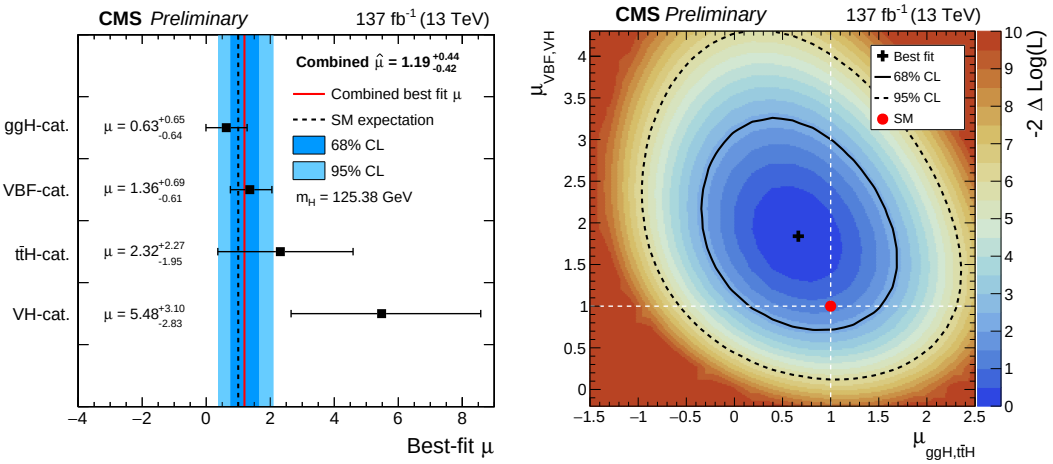

Figure 12: Left: signal strength modifiers measured for $m_{\mathrm{H}} = 125.38$ GeV in each production category (black points) are compared to the result of the combined fit (solid red line) and the SM expectation (dashed grey line). Right: scan of the profiled likelihood ratio as a function of $\mu_{\mathrm{ggH,t\bar{t}H}}$ and $\mu_{\mathrm{VBF,VH}}$ with the corresponding 1σ and 2σ uncertainty contours. The black cross indicates the best-fit values $(\hat{\mu}_{\mathrm{ggH,t\bar{t}H}}, \hat{\mu}_{\mathrm{VBF,VH}}) = (0.66, 1.84)$, while the red circle represents the SM expectation.

The result is combined with that obtained from data recorded at centre-of-mass energies of 7 and 8 TeV. The 7+8 TeV search is identical to the one described in Ref. [80], except that the values used for the Higgs boson production cross sections and the branching fraction are updated to those reported in Ref. [21]. Systematic uncertainties in the inclusive signal production cross sections and $\mathcal{B}(\mathrm{H} \to \mu\mu)$ are correlated across the 7, 8, and 13 TeV analyses. Experimental uncertainties affecting the measured properties of the various physics objects (muons, electrons, jets, and b quark jets), the measurement of the integrated luminosity, and the modelling of the pileup conditions are assumed to be uncorrelated between the 7+8 and 13 TeV analyses. The combination improves upon the 13 TeV result by about 1%. Table 8 reports the observed and expected significances over the background-only expectation at $m_{\mathrm{H}} = 125.38$ GeV and the 95% CL upper limits on μ in each production category as well as for the 13 TeV and the 7+8+13 TeV

combined fits. Figure 13 shows the observed (solid black) and the expected (dashed black) local p-values derived from the 7+8+13 TeV combined fit as a function of m_{H} in a 5 GeV window around the expected Higgs boson mass. The expected p-value is computed on an Asimov data set [91] generated from the background expectation obtained from the signal-plus-background fit injecting a signal at $m_{\mathrm{H}} = 125.38$ GeV. As in Fig. 11, the solid markers indicate the mass points for which the observed p-values are computed.

Table 8: Observed and expected significances for the incompatibility with the background-only hypothesis for $m_{\mathrm{H}} = 125.38$ GeV and the corresponding 95% CL upper limits on μ (in absence of $\mathrm{H} \to \mu\mu$ decays) for each production category as well as for the 13 TeV and the 7+8+13 TeV combined fits.

Production category	Observed (expected) Signif.	Observed (expected) UL on μ
VBF	2.40 (1.77)	2.57 (1.22)
ggH	0.99 (1.56)	1.77 (1.28)
$\mathrm{t\bar{t}H}$	1.20 (0.54)	6.48 (4.20)
VH	2.02 (0.42)	10.8 (5.13)
Combined $\sqrt{s} = 13$ TeV	2.95 (2.46)	1.94 (0.82)
Combined $\sqrt{s} = 7,\ 8,\ 13$ TeV	2.98 (2.48)	1.93 (0.81)

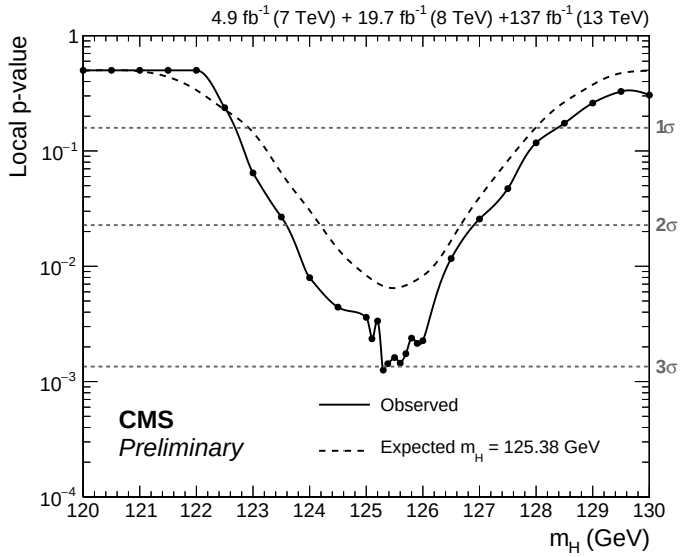

Figure 13: Observed (solid black) and expected (dashed black) local p-values as a function of m_{H}, extracted from the combined fit performed on data recorded at $\sqrt{s} = 7,\ 8,$ and 13 TeV, are shown. The expected p-values are calculated using the background expectation obtained from the signal-plus-background fit and injecting a signal with $m_{\mathrm{H}} = 125.38$ GeV and $\mu = 1$.

The results presented in this note are the most precise measurements in the $\mathrm{H} \to \mu\mu$ decay channel reported to date, and can be used to improve constraints on the coupling between the Higgs boson and fermions of the second generation. The signal strength measured in the $\mathrm{H} \to \mu\mu$ analysis cannot be translated directly into a measurement of the Higgs boson coupling to muons because it is also sensitive to the interactions between the Higgs boson and several SM particles involved in the considered production processes, primarily the top quark and vector boson couplings. These Higgs boson couplings to other particles are constrained by combining the result of this analysis with those presented in Ref. [95], based on pp collision data recorded by the CMS experiment at $\sqrt{s} = 13$ TeV corresponding to an integrated luminosity of up to 137 fb^{-1}.

Under the assumption that there are no BSM particles contributing to the Higgs boson total width, Higgs boson production and decay rates in each category are expressed in terms of coupling modifiers within the so-called κ-framework. Six free coupling parameters are introduced in the likelihood (κ_W, κ_Z, κ_t, κ_τ, κ_b, and κ_μ) and are extracted from a simultaneous fit across all categories. In the combined fit, the coupling modifiers are constrained to be positive defined and the event categories of the $H \to \mu\mu$ analysis described in this document supersede those considered in Ref. [95]. Figure 14 (left) shows the observed profile likelihood ratio as a function of κ_μ for $m_H = 125.38$ GeV. The best-fit value for κ_μ, as well as those for the other couplings, are compatible with the SM prediction. The corresponding 68% and 95% CL intervals for the κ_μ parameter are $0.91 < \kappa_\mu < 1.34$ and $0.65 < \kappa_\mu < 1.53$, respectively. Note that the observed (expected) significances reported in Table 8 and Fig. 11 are computed assuming SM production cross sections and decay rates, constrained within the corresponding theoretical uncertainties. In contrast, in the result presented in Fig. 11 (left) the freely floating coupling modifiers are allowed to simultaneously modify both Higgs boson production cross sections and decay rates within the constraint of keeping the total Higgs boson width fixed to the SM value.

In the SM, the Yukawa coupling between the Higgs boson and the fermions (λ_F) is proportional to the fermion mass (m_F), while the coupling to weak bosons (g_V) is proportional to the square of the vector boson masses (m_V). The results from the κ-fit can therefore be translated in terms of reduced coupling strength modifiers, defined as $y_V = \sqrt{\kappa_V}\frac{m_V}{v}$ for weak bosons and $y_F = \kappa_F \frac{m_F}{v}$ for fermions, where v is the vacuum expectation value of the Higgs field of 246.22 GeV. Figure 14 (right) shows the best-fit estimates for the six reduced coupling strength modifiers as a function of particle mass, where lepton, vector boson, and quark masses are taken from Ref. [84]. The compatibility between the measured coupling strength modifiers and their SM expectation is derived from the $-2 \times \Delta\mathrm{Log}(L)$ separation between the best-fit and an alternative one, performed by fixing the six coupling modifiers to the SM prediction ($\kappa_W = \kappa_Z = \kappa_t = \kappa_\tau = \kappa_b = \kappa_\mu = 1$), yielding a p-value of 44%.

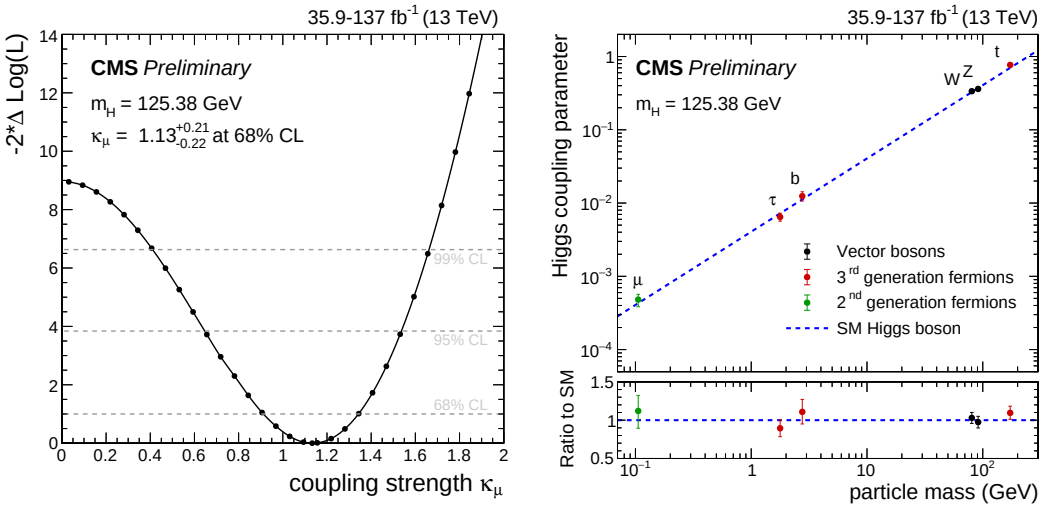

Figure 14: Left: observed profile likelihood ratio as a function of κ_μ for $m_H = 125.38$ GeV, obtained from a combined fit with Ref. [10] in the κ-framework model. The best-fit value for κ_μ is 1.13 and the corresponding observed 68% CL interval is $0.91 < \kappa_\mu < 1.34$. Right: the best-fit estimates for the reduced coupling modifiers extracted for fermions and weak bosons from the resolved κ-framework model compared to their corresponding prediction from the SM. The error bars represent 68% CL intervals for the measured parameters. The lower panel shows the ratios of the measured coupling modifiers values to their SM predictions.

11 Summary

A measurement of the Higgs boson decay to a pair of muons is presented. This result combines searches in four exclusive categories targeting the production of the Higgs boson via gluon fusion, via vector boson fusion, in association with a weak vector boson, and in association with a pair of top quarks. The measurement is performed using $\sqrt{s} = 13\,\text{TeV}$ proton-proton (pp) collision data, corresponding to an integrated luminosity of $137\,\text{fb}^{-1}$, recorded by the CMS experiment at the CERN LHC. An excess of events is observed in data with a significance of 3.0 standard deviations, where the expectation for the standard model (SM) Higgs boson with $m_{\text{H}} = 125.38\,\text{GeV}$ is 2.5. The measured signal strength, relative to the SM expectation, is $1.19^{+0.41}_{-0.39}(\text{stat})^{+0.17}_{-0.16}(\text{sys})$. The combination of this result with that from data recorded at centre-of-mass energies of 7 and 8 TeV improves both expected and observed sensitivity by 1%. This result constitutes the first evidence for the Higgs boson decay to fermions of the second generation.

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.

[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] CMS Collaboration, "Observation of the Higgs boson decay to a pair of τ leptons with the CMS detector", *Phys. Lett. B* **779** (2018) 283, doi:10.1016/j.physletb.2018.02.004, arXiv:1708.00373.

[5] CMS Collaboration, "Observation of $\text{t}\bar{\text{t}}\text{H}$ production", *Phys. Rev. Lett.* **120** (2018) 231801, doi:10.1103/PhysRevLett.120.231801, arXiv:1804.02610.

[6] CMS Collaboration, "Observation of Higgs boson decay to bottom quarks", *Phys. Rev. Lett.* **121** (2018) 121801, doi:10.1103/PhysRevLett.121.121801, arXiv:1808.08242.

[7] CMS Collaboration, "Measurements of Higgs boson properties in the diphoton decay channel in proton-proton collisions at $\sqrt{s} = 13\,\text{TeV}$", *JHEP* **11** (2018) 185, doi:10.1007/JHEP11(2018)185, arXiv:1804.02716.

[8] CMS Collaboration, "Measurements of properties of the Higgs boson decaying to a W boson pair in pp collisions at $\sqrt{s} = 13\,\text{TeV}$", *Phys. Lett. B* **791** (2019) 96, doi:10.1016/j.physletb.2018.12.073, arXiv:1806.05246.

[9] CMS Collaboration, "Measurements of properties of the Higgs boson decaying into the four-lepton final state in pp collisions at $\sqrt{s} = 13\,\text{TeV}$", *JHEP* **11** (2017) 047, doi:10.1007/JHEP11(2017)047, arXiv:1706.09936.

[10] CMS Collaboration, "Combined measurements of Higgs boson couplings in proton-proton collisions at $\sqrt{s} = 13\,\text{TeV}$", *Eur. Phys. J. C* **79** (2019) 421, doi:10.1140/epjc/s10052-019-6909-y, arXiv:1809.10733.

[11] ATLAS Collaboration, "Cross-section measurements of the Higgs boson decaying into a pair of τ-leptons in proton-proton collisions at $\sqrt{s} = 13\,\text{TeV}$ with the ATLAS detector", *Phys. Rev D.* **99** (2019) 072001, doi:10.1103/PhysRevD.99.072001, arXiv:1811.08856.

[12] ATLAS Collaboration, "Observation of $\text{H} \to \text{b}\overline{\text{b}}$ decays and VH production with the ATLAS detector", *Phys. Lett. B* **786** (2018) 59, doi:10.1016/j.physletb.2018.09.013, arXiv:1808.08238.

[13] ATLAS Collaboration, "Observation of Higgs boson production in association with a top quark pair at the LHC with the ATLAS detector", *Phys. Lett. B* **784** (2018) 173, doi:10.1016/j.physletb.2018.07.035, arXiv:1806.00425.

[14] ATLAS Collaboration, "Measurements of gluon-gluon fusion and vector-boson fusion Higgs boson production cross-sections in the $\text{H} \to \text{WW}^* \to e\nu\mu\nu$ decay channel in pp collisions at $\sqrt{s} = 13\,\text{TeV}$ with the ATLAS detector", *Phys. Lett. B* **789** (2019) 508, doi:10.1016/j.physletb.2018.11.064, arXiv:1808.09054.

[15] ATLAS Collaboration, "Measurement of the Higgs boson coupling properties in the $\text{H} \to \text{ZZ}^* \to 4\ell$ decay channel at $\sqrt{s} = 13\,\text{TeV}$ with the ATLAS detector", *JHEP* **03** (2018) 095, doi:10.1007/JHEP03(2018)095, arXiv:1712.02304.

[16] ATLAS Collaboration, "Measurements of Higgs boson properties in the diphoton decay channel with $36\,\text{fb}^{-1}$ of pp collision data at $\sqrt{s} = 13\,\text{TeV}$ with the ATLAS detector", *Phys. Rev. D* **98** (2018) 052005, doi:10.1103/PhysRevD.98.052005, arXiv:1802.04146.

[17] ATLAS Collaboration, "Combined measurements of Higgs boson production and decay using up to $80\,\text{fb}^{-1}$ of proton-proton collision data at $\sqrt{s} = 13\,\text{TeV}$ collected with the ATLAS experiment", *Phys. Rev. D* **101** (2020) 012002, doi:10.1103/PhysRevD.101.012002, arXiv:1909.02845.

[18] F. Englert and R. Brout, "Broken symmetry and the mass of gauge vector mesons", *Phys. Rev. Lett.* **13** (1964) 321, doi:10.1103/PhysRevLett.13.321.

[19] P. W. Higgs, "Broken symmetries and the masses of gauge bosons", *Phys. Rev. Lett.* **13** (1964) 508, doi:10.1103/PhysRevLett.13.508.

[20] P. W. Higgs, "Spontaneous symmetry breakdown without massless bosons", *Phys. Rev.* **145** (1966) 1156, doi:10.1103/PhysRev.145.1156.

[21] LHC Higgs Cross Section Working Group, "Handbook of LHC Higgs Cross Sections: 4. Deciphering the Nature of the Higgs Sector", *CERN* (2016) doi:10.23731/CYRM-2017-002, arXiv:1610.07922.

[22] CMS Collaboration, "Search for the Higgs boson decaying to two muons in proton-proton collisions at $\sqrt{s} = 13\,\text{TeV}$", *Phys. Rev. Lett.* **122** (2019) 021801, doi:10.1103/PhysRevLett.122.021801, arXiv:1807.06325.

[23] ATLAS Collaboration, "A search for the dimuon decay of the Standard Model Higgs boson with the ATLAS detector", *Submitted to Phys. Lett. B* (7, 2020) arXiv:2007.07830.

[24] CMS Collaboration, "A measurement of the Higgs boson mass in the diphoton decay channel", *Phys. Lett. B* **805** (2020) 135425, doi:10.1016/j.physletb.2020.135425, arXiv:2002.06398.

[25] CMS Collaboration, "The CMS trigger system", *JINST* **12** (2017) P01020, doi:10.1088/1748-0221/12/01/P01020, arXiv:1609.02366.

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

[27] 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.

[28] 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.

[29] 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.

[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] CMS Collaboration, "Performance of missing transverse momentum reconstruction in proton-proton collisions at $\sqrt{s} = 13$ TeV using the CMS detector", *JINST* **14** (2019) P07004, doi:10.1088/1748-0221/14/07/P07004, arXiv:1903.06078.

[32] 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.

[33] 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.

[34] CMS Collaboration, "Performance of electron reconstruction and selection with the CMS detector in proton-proton collisions at $\sqrt{s} = 8$ TeV", *JINST* **10** (2015) P06005, doi:10.1088/1748-0221/10/06/P06005, arXiv:1502.02701.

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

[36] 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.

[37] 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.

[38] 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.

[39] 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.

[40] E. Bagnaschi, G. Degrassi, P. Slavich, and A. Vicini, "Higgs production via gluon fusion in the POWHEG approach in the SM and in the MSSM", *JHEP* **02** (2012) 088, doi:10.1007/JHEP02(2012)088, arXiv:1111.2854.

[41] P. Nason and C. Oleari, "NLO Higgs boson production via vector-boson fusion matched with shower in POWHEG", *JHEP* **02** (2010) 037, doi:10.1007/JHEP02(2010)037, arXiv:0911.5299.

[42] G. Luisoni, P. Nason, C. Oleari, and F. Tramontano, "$HW^{\pm}/HZ + 0$ and 1 jet at NLO with the POWHEG BOX interfaced to GoSam and their merging within MiNLO", *JHEP* **10** (2013) 083, doi:10.1007/JHEP10(2013)083, arXiv:1306.2542.

[43] C. Anastasiou et al., "High precision determination of the gluon fusion Higgs boson cross-section at the LHC", *JHEP* **05** (2016) 058, doi:10.1007/JHEP05(2016)058, arXiv:1602.00695.

[44] M. Cacciari et al., "Fully Differential Vector-Boson-Fusion Higgs Production at Next-to-Next-to-Leading Order", *Phys. Rev. Lett.* **115** (2015) 082002, doi:10.1103/PhysRevLett.115.082002, arXiv:1506.02660. [Erratum: Phys.Rev.Lett. 120, 139901 (2018)].

[45] O. Brein, A. Djouadi, and R. Harlander, "NNLO QCD corrections to the Higgs-strahlung processes at hadron colliders", *Phys. Lett. B* **579** (2004) 149, doi:10.1016/j.physletb.2003.10.112, arXiv:hep-ph/0307206.

[46] S. Dawson et al., "Associated Higgs production with top quarks at the large hadron collider: NLO QCD corrections", *Phys. Rev. D* **68** (2003) 034022, doi:10.1103/PhysRevD.68.034022, arXiv:hep-ph/0305087.

[47] S. Frixione et al., "Weak corrections to Higgs hadroproduction in association with a top-quark pair", *JHEP* **09** (2014) 065, doi:10.1007/JHEP09(2014)065, arXiv:1407.0823.

[48] A. Djouadi, J. Kalinowski, and M. Spira, "HDECAY: A Program for Higgs boson decays in the standard model and its supersymmetric extension", *Comput. Phys. Commun.* **108** (1998) 56, doi:10.1016/S0010-4655(97)00123-9, arXiv:hep-ph/9704448.

[49] M. Spira, "QCD effects in Higgs physics", *Fortsch. Phys.* **46** (1998) 203, doi:10.1002/(SICI)1521-3978(199804)46:3<203::AID-PROP203>3.0.CO;2-4, arXiv:hep-ph/9705337.

[50] K. Hamilton, P. Nason, E. Re, and G. Zanderighi, "NNLOPS simulation of Higgs boson production", *JHEP* **10** (2013) 222, doi:10.1007/JHEP10(2013)222, arXiv:1309.0017.

[51] K. Hamilton, P. Nason, and G. Zanderighi, "Finite quark-mass effects in the NNLOPS POWHEG+MiNLO Higgs generator", *JHEP* **05** (2015) 140, doi:10.1007/JHEP05(2015)140, arXiv:1501.04637.

[52] Y. Li and F. Petriello, "Combining QCD and electroweak corrections to dilepton production in FEWZ", *Phys. Rev D.* **86** (2012) 094034, doi:10.1103/PhysRevD.86.094034, arXiv:1208.5967.

[53] M. Grazzini, S. Kallweit, D. Rathlev, and M. Wiesemann, "$W^{\pm}Z$ production at the LHC: fiducial cross sections and distributions in NNLO QCD", *JHEP* **05** (2017) 139, doi:10.1007/JHEP05(2017)139, arXiv:1703.09065.

[54] M. Grazzini, S. Kallweit, and D. Rathlev, "ZZ production at the LHC: fiducial cross sections and distributions in NNLO QCD", *Phys. Lett. B* **750** (2015) 407, doi:10.1016/j.physletb.2015.09.055, arXiv:1507.06257.

[55] T. Gehrmann et al., "W^+W^- production at hadron colliders in Next to Next to Leading Order QCD", *Phys. Rev. Lett.* **113** (2014) 212001, doi:10.1103/PhysRevLett.113.212001, arXiv:1408.5243.

[56] J. M. Campbell, R. K. Ellis, and C. Williams, "Vector boson pair production at the LHC", *JHEP* **07** (2011) 018, doi:10.1007/JHEP07(2011)018, arXiv:1105.0020.

[57] 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.

[58] 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.

[59] NNPDF Collaboration, "Parton distributions for the LHC Run II", *JHEP* **04** (2015) 040, doi:10.1007/JHEP04(2015)040, arXiv:1410.8849.

[60] 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.

[61] T. Sjöstrand, S. Mrenna, and P. Z. Skands, "A Brief Introduction to PYTHIA 8.1", *Comput. Phys. Commun.* **178** (2008) 852, doi:10.1016/j.cpc.2008.01.036, arXiv:0710.3820.

[62] CMS Collaboration, "Event generator tunes obtained from underlying event and multiparton scattering measurements", *Eur. Phys. J. C* **76** (2016) 155, doi:10.1140/epjc/s10052-016-3988-x, arXiv:1512.00815.

[63] 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.

[64] J. Alwall et al., "Comparative study of various algorithms for the merging of parton showers and matrix elements in hadronic collisions", *Eur. Phys. J. C* **53** (2008) 473, doi:10.1140/epjc/s10052-007-0490-5, arXiv:0706.2569.

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

[66] J. Bellm et al., "Herwig 7.0/Herwig++ 3.0 release note", *Eur. Phys. J. C* **76** (2016) 196, doi:10.1140/epjc/s10052-016-4018-8, arXiv:1512.01178.

[67] CMS Collaboration, "Electroweak production of two jets in association with a Z boson in proton-proton collisions at $\sqrt{s} = 13\,\text{TeV}$", *Eur. Phys. J. C* **78** (2018) 589, doi:10.1140/epjc/s10052-018-6049-9, arXiv:1712.09814.

[68] CMS Collaboration, "Extraction and validation of a set of HERWIG 7 tunes from CMS underlying-event measurements", Technical Report CMS-PAS-GEN-19-001, CERN, Geneva, 2020.

[69] B. Cabouat and T. Sjöstrand, "Some Dipole Shower Studies", *Eur. Phys. J. C* **78** (2018) 226, doi:10.1140/epjc/s10052-018-5645-z, arXiv:1710.00391.

[70] B. Jäger et al., "Parton-shower effects in Higgs production via Vector-Boson Fusion", (2020). arXiv:2003.12435.

[71] A. Bodek et al., "Extracting muon momentum scale corrections for hadron collider experiments", *Eur. Phys. J. C* **72** (2012) 2194, doi:10.1140/epjc/s10052-012-2194-8, arXiv:1208.3710.

[72] F. Chollet et al., "Keras", 2015. Software available from keras.io. https://keras.io.

[73] M. Abadi et al., "TensorFlow: Large-scale machine learning on heterogeneous systems", 2015. Software available from tensorflow.org. https://www.tensorflow.org/.

[74] J. C. Collins and D. E. Soper, "Angular distribution of dileptons in high-energy hadron collisions", *Phys. Rev. D* **16** (1977) 2219, doi:10.1103/PhysRevD.16.2219.

[75] F. Schissler and D. Zeppenfeld, "Parton shower effects on W and Z production via vector boson fusion at NLO QCD", *JHEP* **04** (2013) 057, doi:10.1007/JHEP04(2013)057, arXiv:1302.2884.

[76] CMS Collaboration, "Performance of quark/gluon discrimination in 8 TeV pp data", Technical Report CMS-PAS-JME-13-002, CERN, Geneva, 2013.

[77] CMS Collaboration, "Jet algorithms performance in 13 TeV data", Technical Report CMS-PAS-JME-16-003, CERN, Geneva, 2017.

[78] N. Srivastava et al., "Dropout: A simple way to prevent neural networks from overfitting", (2014).

[79] G. Cowan, "Discovery sensitivity for a counting experiment with background uncertainty", 2012. https://www.pp.rhul.ac.uk/~cowan/stat/notes/medsigNote.pdf.

[80] CMS Collaboration, "Search for a standard model-like Higgs boson in the $\mu^+\mu^-$ and e^+e^- decay channels at the LHC", *Phys. Lett. B* **744** (2015) 184, doi:10.1016/j.physletb.2015.03.048, arXiv:1410.6679.

[81] R. Barlow and C. Beeston, "Fitting using finite monte carlo samples", *Computer Physics Communications* **77** (1993) 219, doi:10.1016/0010-4655(93)90005-W.

[82] A. Hoecker et al., "TMVA - Toolkit for multivariate data analysis", (2007). arXiv:physics/0703039.

[83] CMS Collaboration, "Measurements of differential Z boson production cross sections in proton-proton collisions at $\sqrt{s} = 13\,\text{TeV}$", *JHEP* **12** (2019) 061, doi:10.1007/JHEP12(2019)061, arXiv:1909.04133.

[84] Particle Data Group Collaboration, "Review of Particle Physics", *Phys. Rev. D* **98** (2018) 030001, doi:10.1103/PhysRevD.98.030001.

[85] D. Bourilkov, "Photon-induced Background for dilepton searches and measurements in pp collisions at 13 TeV", (2016). arXiv:1606.00523.

[86] D. Bourilkov, "Exploring the LHC landscape with dileptons", (2016). arXiv:1609.08994.

[87] P. D. Dauncey, M. Kenzie, N. Wardle, and G. J. Davies, "Handling uncertainties in background shapes", *JINST* **10** (2015) P04015, doi:10.1088/1748-0221/10/04/P04015, arXiv:1408.6865.

[88] R. A. Fisher, "On the interpretation of χ^2 from contingency tables, and the calculation of p", volume 85, p. 87. [Wiley, Royal Statistical Society], 1922.

[89] CMS Collaboration, "Observation of single top quark production in association with a Z boson in proton-proton collisions at $\sqrt{s} = 13\,\text{TeV}$", *Phys. Rev. Lett.* **122** (2019) 132003, doi:10.1103/PhysRevLett.122.132003, arXiv:1812.05900.

[90] CMS Collaboration, "Evidence for associated production of a Higgs boson with a top quark pair in final states with electrons, muons, and hadronically decaying τ leptons at $\sqrt{s} = 13\,\text{TeV}$", *JHEP* **08** (2018) 066, doi:10.1007/JHEP08(2018)066, arXiv:1803.05485.

[91] 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: Eur. Phys. J.C73,2501(2013)].

[92] The ATLAS and CMS Collaborations, and the LHC Higgs Combination Group, "Procedure for the LHC Higgs boson search combination in Summer 2011", Technical Report CMS-NOTE-2011-005, ATL-PHYS-PUB-2011-11, 2011.

[93] 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.

[94] 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.

[95] CMS Collaboration, "Combined Higgs boson production and decay measurements with up to $137\,\text{fb}^{-1}$ of proton-proton collision data at $\sqrt{s} = 13\,\text{TeV}$", Technical Report CMS-PAS-HIG-19-005, CERN, Geneva, 2020.