CMS Physics Analysis Summary

Contact: cms-pag-conveners-higgs@cern.ch 2016/06/10

Search for the standard model Higgs boson produced through vector boson fusion and decaying to $b\bar{b}$ with proton-proton collisions at $\sqrt{s} = 13$ TeV

The CMS Collaboration

Abstract

A search for a standard model Higgs boson (H) that is produced through vector boson fusion and decays to a bottom-quark pair is performed using proton-proton collisions at $\sqrt{s} = 13$ TeV. The data sample corresponds to an integrated luminosity of 2.3 fb^{-1} recorded in 2015 with the CMS experiment at the CERN LHC. The observed 95% confidence level upper limit on the production of a $H \to b\bar{b}$ signal at a mass of 125 GeV is 3.0 times the standard model expectation, compared to 5.0 expected in absence of a signal. The combination of this result with the previous CMS search using proton-proton collisions at $\sqrt{s} = 8$ TeV yields an observed 95% confidence level upper limit on the signal production of 3.4 times the standard model expectation, where 2.2 is expected in the absence of signal, and a corresponding fitted signal strength $\mu = \sigma/\sigma_{SM} = 1.3^{+1.2}_{-1.1}$.

1 Introduction

In the minimal standard model (SM) [1] [2] [3] the Brout-Englert-Higgs mechanism [4] [5] [6] explains the electroweak symmetry breaking and allows electroweak gauge bosons to acquire mass. The mechanism predicts the existence of a Higgs scalar field, and its observation in 2012 with the LHC Run 1 proton-proton collision data by both the CMS [7] and ATLAS [8] collaborations achieved one of the main goals of the LHC physics program.

The observed Higgs boson mass is now precisely determined to be $m_{\mathrm{H}} = 125.09 \pm 0.24\,\mathrm{GeV}$ [9], i.e. with a 0.2% precision level. On the other hand, the observed properties and couplings are consistent with those of a minimal SM Higgs boson but only with a precision at the level of 10% or worse [10]. In particular, while a SM Higgs boson with a mass of 125.1 GeV is predicted to have a dominant decay branching ratio to b-quarks (58.2% [11, 12]), the LHC Run 1 data did not yet clearly establish the coupling of the Higgs boson to b-quarks, nor in general to down-type quarks.

At the LHC, a SM Higgs boson can be produced through a variety of mechanisms, and the most abundant channel after gluon-fusion (GF) is Vector Boson Fusion (VBF) production. While the inclusive observation of the SM Higgs boson decaying to $\mathrm{b\bar{b}}$ pairs is not pursuable in proton collisions, the observation of the $\mathrm{b\bar{b}}$ decay channel in the VBF production context can be pursued thanks to the peculiar kinematics generated by the VBF process. Signal measurements in this channel allow to access both the cross section of the VBF Higgs boson production, and its coupling to b-quarks.

The first search for a SM Higgs boson produced through vector boson fusion and decaying to a bottom-quark pair was performed by CMS on the Run 1 $\sqrt{s} = 8$ TeV proton-proton collision data [13]. The search yielded a fitted signal strength of $\mu = \sigma/\sigma_{\mathrm{SM}} = 2.8^{+1.6}_{-1.4}$ at a Higgs boson mass of 125 GeV, with a corresponding observed significance of 2.2 standard deviations, where 0.8 standard deviations were expected with a SM signal.

The combination of the Run 1 VBF $\mathrm{H} \to \mathrm{b\bar{b}}$ results with other CMS $\mathrm{H} \to \mathrm{b\bar{b}}$ searches where the Higgs boson is produced in association with a vector boson (VH) [14], or with a top quark pair ($\mathrm{t\bar{t}H}$) [15, 16], yielded a $\mathrm{H} \to \mathrm{b\bar{b}}$ signal strength $\mu = 1.03^{+0.44}_{-0.42}$ with a signal significance of 2.6 standard deviations for $m_{\mathrm{H}} = 125$ GeV [13].

The prominent feature of the VBF process $\mathrm{qqH} \to \mathrm{qqb\bar{b}}$ is the presence of four energetic jets in the final state. Two jets are expected to originate from a light-quark pair (u or d), which are typically two valence quarks from each of the colliding protons scattered away from the beam line in the VBF process. These "VBF-tagging" jets are expected to be roughly in the forward and backward directions relative to the beam direction. Two additional jets are expected from the Higgs boson decay to a $\mathrm{b\bar{b}}$ pair in more central regions of the detector. Another important property of the signal events is that, being produced through an electroweak process, no quantum chromodynamics (QCD) color is exchanged at leading order in the production. As a result, in the most probable color evolution of these events, the VBF-tagging jets connect to the proton remnants in the forward and backward beam line directions, while the two b-quark jets connect to each other as decay products of the color neutral Higgs boson. Consequently very little additional QCD radiation and hadronic activity is expected in the space outside the color-connected regions, in particular in the whole rapidity interval (rapidity gap) between the two VBF-tagging jets, with the exception of the Higgs boson decay products.

The dominant background to this search is from QCD production of multijet events. Other backgrounds arise from: (i) hadronic decays of Z or W bosons produced in association with additional jets, (ii) hadronic decays of top quark pairs, and (iii) hadronic decays of singly-

produced top quarks. The contribution of the Higgs boson in GF processes with two or more associated jets is included in the expected signal yield.

The search is performed on selected four-jet events that are characterized by the response of a multivariate discriminant trained to separate signal events from background without making use of kinematic information on the two b-jet candidates. Subsequently, the invariant mass distribution of two b jets is analyzed in each category in the search for a signal "bump" above the smooth contribution from the SM background.

This paper is organized as follows: Section 2 highlights the features of the CMS detector needed to perform this analysis. Section 3 details the production of simulated samples used to study the signal and main backgrounds, and Section 4 presents the employed triggers. Event reconstruction and selection are described in Sections 5 and 6, respectively. The unique features of the analysis are discussed in Section 7, which include the improvement of the resolution in jet transverse momentum (p_T) by regression techniques, discrimination between quark- and gluon-originated jets, and soft QCD activity. The search for a SM Higgs boson is discussed in Section 8 and the associated systematic uncertainties are presented in Section 9. The final results are discussed in Section 10. The combination with Run 1 results is reported in Section 11 and a summary is given in Section 12.

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. A silicon pixel and strip tracker, a lead tungstate crystal electromagnetic calorimeter, and a brass and scintillator hadron calorimeter are located within the axial field. Muons are measured in gas-ionization detectors embedded in the steel flux-return yoke of the solenoid. Forward calorimetry (pseudorapidity $3 < |\eta| < 5$) complements the coverage provided by the barrel ($|\eta| < 1.3$) and endcap ($1.3 < |\eta| < 3$) detectors. The first level (L1) of the CMS trigger system, composed of specialized processors, uses information from the calorimeters and muon detectors to select the most interesting events in a time interval of less than 4 μs. The high-level trigger (HLT) processor farm decreases the event rate from about 100 kHz to about 1 kHz, before data storage. A more detailed description of the CMS apparatus and the main kinematic variables used in the analysis can be found in Ref. [17].

3 Simulated samples

Samples of simulated Monte Carlo (MC) signal and background events are used to guide the analysis optimization and to estimate signal yields. Several event generators are used to produce the MC events. The NNPDF3.0 parton distribution functions (PDF) [18] are used for all samples.

The samples of VBF and GF signal processes are generated using the next-to-leading order (NLO) perturbative QCD program POWHEG 2.0 [19–21], and PYTHIA 8 [22, 23] for the hadronization process and modelling of the underlying event (UE), using the CUETP8M1 tune [24]. The signal samples are generated using only $H \to b\bar{b}$ decays, for m_H=125 GeV, and are normalized using cross sections and decay rates computed by the LHC Higgs cross section working group [12].

Background samples of QCD multijet, Z +jets and W+jets events are simulated using leading-order (LO) MADGRAPH 5 [25] interfaced with PYTHIA 8. The $t\bar{t}$ and single top quark background samples are produced using POWHEG [26, 27], interfaced with PYTHIA 8. The produc-

tion cross sections for W+jets and Z +jets are rescaled to next-to-next-to-leading-order (NNLO) cross sections calculated using the FEWZ 3.1 program [28–30]. The $t\bar{t}$ and single top quark samples are also rescaled to their cross sections based on NNLO calculations [31, 32].

To accurately simulate the LHC luminosity conditions during data taking, additional simulated pp interactions overlapping in the same or neighboring bunch crossings of the main interaction, denoted as pileup, are added to the simulated events with a multiplicity distribution that matches the one in data.

4 Triggers

The data used for this analysis were collected using two trigger strategies that result in two different data samples for analysis. The two trigger event selection (paths) were specifically designed and deployed for the VBF qqH $\rightarrow$ qqbb signal search, both for the L1 trigger and the HLT, and operated during the full 2015 data taking. Both paths use a common L1 selection (seed) of events with three jets, and a common initial HLT selection of four jets, while differentiating on the presence of b-jet candidates and on the final kinematic requirements of the four jets. The total collected integrated luminosity used in this analysis was 2.32 fb^{-1} for both paths.

4.1 Common part

The L1 seed required the presence of at least three jets with p_{T} above decreasing thresholds $p_{\mathrm{T}}^{(1)}$=84 GeV, $p_{\mathrm{T}}^{(2)}$=68 GeV, $p_{\mathrm{T}}^{(3)}$=48 GeV. Among the three jets, one and only one of the two leading jets (with $p_{\mathrm{T}} > p_{\mathrm{T}}^{(1)}$, $p_{\mathrm{T}}^{(2)}$) can be in the forward region with pseudorapidity $2.6 < |\eta| \leq 5.2$, while the other two jets are required to be central ($|\eta| \leq 2.6$). The estimated efficiency of the L1 seed on signal events is 31%.

The common HLT event selection required the presence of at least four particle-flow jets (PFJets, see Section 5) with p_{T} above decreasing thresholds p_{T} >92, 76, 64, 15 GeV, so that essentially the three p_{T}-leading jets are required to be energetic, while the fourth jet presence is just required above a minimal threshold.

The HLT b-tagging algorithms used in the HLT paths is the "Combined Secondary Vertex" (CSV) [33, 34], evaluated using HLT regional tracking around the jets, and at least one of the event jets must fulfill minimum b-tagging requirements. At this point HLT events can be classified in two different paths if they contain one b-tagged jet (SingleB path), or more than one b-tagged jet (DoubleB path), and proceed through different selections as detailed below.

4.2 SingleB sample

The SingleB path selects events that pass the above common HLT selection and some final cuts on the PFJets kinematics. For these final cuts only the four p_{T}-leading PFJets are considered: the most b-tagged jet is labeled as a "b"-jet and, among the remaining three, the two jets with the largest pseudorapidity opening are labeled as the two "q"-jets, so that the second "b"-jet is then identified as the remaining jet. Final cuts are applied on the q-jets pseudorapidity separation $\Delta\eta_{qq} > 4.1$ and invariant mass $m_{qq} > 460$ GeV, as well as on the b-jets azimuthal opening $\Delta\phi_{bb} < 1.6$.

4.3 DoubleB sample

The DoubleB paths select events that pass the common HLT selection and have a second b-tag, among the six p_{T}-leading PFJets. Final cuts on the PFJets kinematics are also applied.

For the final cuts the two most b-tagged jets are labeled as "b"-jets and among the collection of remaining PFJets, the two p_T-leading ones are labeled as the two "q"-jets. Final cuts are applied on the q-jets pseudorapidity separation $\Delta\eta_{qq} > 1.2$ and invariant mass $m_{qq} > 200$ GeV.

4.4 Trigger efficiency

To evaluate trigger efficiencies, two prescaled control paths are used. A path requiring two PFJets with average $p_T > 60$ GeV is used to determine the trigger efficiencies of the kinematic requirements on the jets. A second path requiring the above four-jet common selection, but without any b-tagging condition, is used to determine the online b-tagging efficiency. Simulated trigger efficiency are then corrected to match the efficiency in data, with a scale factors that are parametrized a function of the jets p_T and one or two b-tag output values, respectively for the SingleB and DoubleB samples. The estimated efficiency of the HLT paths on signal events is 2.3% for SingleB and 3.9% for DoubleB, while their combined efficiency is estimated to be 6.2%.

5 Event reconstruction

The offline analysis uses reconstructed charged-particle tracks and candidates from the particle-flow (PF) algorithm [35–37]. In the PF event reconstruction all stable particles in the event, i.e. electrons, muons, photons, and charged and neutral hadrons, are reconstructed as PF candidates using information from all CMS subdetectors to obtain an optimal determination of their direction, energy, and type. The PF candidates are then used to reconstruct the jets and missing transverse energy.

Jets are reconstructed by clustering PF candidates with the anti-k_T algorithm [38, 39] with a distance parameter of 0.4. Reconstructed jets require a small additional energy correction, mostly due to thresholds on reconstructed tracks and clusters in the PF algorithm and various reconstruction inefficiencies [40]. Jet identification criteria are also applied to reject misreconstructed jets resulting from detector noise, as well as jets heavily contaminated with pileup energy (clustering of energy deposits not associated with a parton from the primary pp interaction) [41]. The efficiency of the jet identification criteria is greater than 99%, with the rejection of 90% of background pileup jets with $p_T \simeq 50$ GeV.

The identification of jets that originate from the hadronization of b quarks is done with the CSV b tagger [33, 34], also implemented for the HLT paths, as described in Section 4. The CSV algorithm combines the information from track impact parameters and secondary vertices identified within a given jet, and provides a continuous discriminator output.

Events are required to have at least four reconstructed jets. All the jets found in an event are ordered according to their p_T, and the most probable b jet and VBF-tagging q-jet candidates are searched among the seven leading ones. For the DoubleB sample, as in the HLT logic, the two jets with the highest CSV b-tagging outputs are selected as the two b-jet candidates, among the remaining the two highest-p_T jets are selected as q-jet candidates. For the SingleB sample, the distinction between the two jet types is done by means of a multivariate discriminant that takes into account the jet b-tag, η and p_T values, together with their respective rankings among the other jets in the event. A boosted decision tree (BDT), implemented with the TMVA package [42], is trained on simulated signal events using these discriminating variables and its output is used as a b-jet likelihood score; among all jets in the event, the two with the highest score are identified as the b jets, while the other two jets with the largest pseudorapidity separation are identified as the VBF-tagging jets. With the use of the multivariate b-jet assignment

the signal efficiency is increased by ≈5% compared to the interpretation based on CSV output only.

6 Event selection

The offline event selection is based upon the b-jet and VBF-tagging jet assignment described in Section 5, and is adjusted to the two different samples (SingleB and DoubleB) presented in Section 4. These selections are summarized in Table 1.

Following the trigger selections, events in either sample are required to have at least four PF jets with $p_{\mathrm{T}}^{1,2,3,4} > 92, 76, 64, 30$ GeV and $|\eta| < 4.7$. In each sample the two b-jets and q-jets are chosen with the criteria defined above in Section 5, and further requirements on these jet pairs also follow the HLT criteria in order to ensure an adequate trigger efficiency at this level.

For events in the SingleB sample the VBF topology is ensured by requiring $m_{\mathrm{qq}} > 460$ GeV and $|\Delta\eta_{\mathrm{qq}}| > 4.1$, and the $\Delta\phi_{\mathrm{bb}} < 1.6$ radians. For events in the DoubleB sample at least two jets must have a CSV [33, 34] output greater than 0.5, while the requirements on the jet pairs are $m_{\mathrm{qq}} > 200$ GeV, $|\Delta\eta_{\mathrm{qq}}| > 1.2$ and $\Delta\phi_{\mathrm{bb}} <$2.4. Events in the DoubleB sample are also required not to belong to the SingleB sample, to avoid double counting.

Figure 1 shows the normalized distributions of m_{qq} in the two SingleB and DoubleB samples, for the sum of all simulated backgrounds, and the VBF and GF Higgs boson production.

After all the selection requirements, 1.6% of the simulated VBF signal events end up in the SingleB sample, and 0.6% end up in DoubleB. The fraction of events in SingleB that also satisfy the requirements of DoubleB (except for the SingleB veto) amounts to 69%. The DoubleB selection recovers signal events presenting less pronounced VBF q-jets, but better tagged b-jets.

Table 1: Summary of selection requirements for the two analyses.

	SingleB	DoubleB
Trigger	one b-tagged jet	two b-tagged jets
jets p_{T}	$p_{\mathrm{T}}^{1,2,3,4} > 92, 76, 64, 30$ GeV	
jets $\|\eta\|$	<4.7	
b tag	no cut	two jets with CSV>0.5
$\Delta\phi_{\mathrm{bb}}$	<1.6 radians	<2.4 radians
VBF topology	$m_{\mathrm{qq}} > 460$ GeV	$m_{\mathrm{qq}} > 200$ GeV
	$\|\Delta\eta_{\mathrm{qq}}\| > 4.1$	$\|\Delta\eta_{\mathrm{qq}}\| > 1.2$
Veto	None	Events that belong to SingleB

7 Signal properties

The analysis described in this paper relies on characteristic properties of the studied final state, which provide a significant improvement of the overall sensitivity. First, the resolution of the invariant mass of the two b jets is improved by applying multivariate regression techniques. Then jet composition properties are used to separate jets originating from light quarks or gluons. Third, soft QCD activity outside the jets is quantified and used as a discriminant between QCD processes with strong color flow and the VBF signal without color flow.

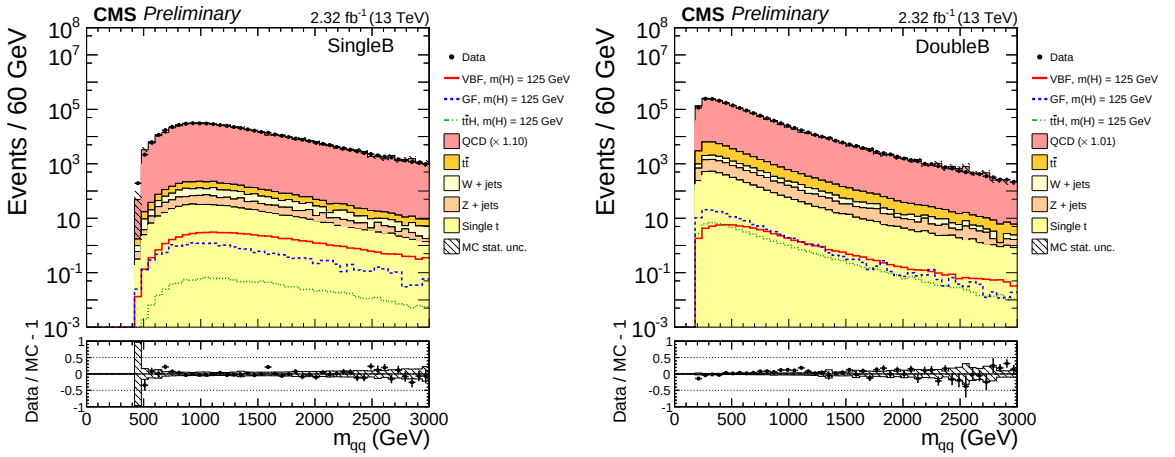

Figure 1: Distributions of the invariant mass of the two VBF-jet candidates (m_{qq}) for events in the SingleB (left) and DoubleB (right) samples. Data are shown by the points, while the simulated backgrounds are stacked. The LO QCD cross section is scaled such that the total number of background events matches the number of events in the data for each category, with a resulting multiplicative factor of 1.10 and 1.01 for the SingleB and DoubleB samples respectively. The VBF Higgs boson signal is displayed by a solid line, and the GF Higgs boson signal is shown by a dashed line. The panels at the bottom show the fractional difference between data and background simulation, with the shaded band representing the statistical uncertainties in the MC samples.

7.1 Jet transverse-momentum regression

The $b\bar{b}$ mass resolution is improved by using a regression technique similar to those used in the Run 1 $H \to b\bar{b}$ searches [13, 14]. A refined calibration is carried out for individual b jets, beyond the default jet energy corrections, that takes into account the jet composition properties and targets semileptonic b decays that lead to a substantial mismeasurement of the jet p_T due to the presence of an escaping neutrino. For this purpose a regression BDT is trained on simulated ditop events with inputs including information about the jet properties and structure. The target of the regression is the p_T of the associated particle-level jet, clustered from all stable particles (with lifetime $c\tau > 1\,\text{cm}$). The inputs include: (i) the jet p_T, η, and transverse mass; (ii) the p_T of the leading track in the jet; (iii) the jet energy fractions carried by hadrons and photons [35–37]; (iv) the p_T, mass and number of charged tracks associated to the secondary vertex, when present; (v) the decay length and uncertainty, of the secondary vertex, when present; (vi) the p_T component and $\Delta R = \sqrt{\Delta\phi^2 + \Delta\eta^2}$ distance relative to the jet axis of the soft-lepton candidate (with $p_T > 3$ GeV), when present; and (vii) the number of reconstructed primary vertices in the event.

The additional energy correction of b jets leads to an improvement of the jet p_T resolution, which in turn improves the dijet invariant mass resolution by approximately 5% in the phase space of the offline event selections. As a final correction to recover harder final state radiation (FSR) gluon jets outside the b-jet cone, the four-momenta of other eventual jets with $p_T > 15$ GeV adjacent to the b-jet (within $\Delta R < 0.8$) are added to the regressed b-jet four-momentum further improving the resolution by about 2%, and the final reconstructed peak position.

Figure 2 shows the reconstructed dijet invariant mass of the b-jet candidates (m_{bb}) before and after the regression for simulated events passing the SingleB or DoubleB selections. The measured distribution of the regressed m_{bb} in both samples is shown in Fig. 3.

The validation of the regression technique in data is done with samples of $Z \to \ell\ell$ events with one or two b-tagged jets. When the jets are corrected by the regression procedure, the p_T balance distribution, between the Z boson, reconstructed from the leptons, and the b-tagged jet or dijet system is improved to be better centered at zero and narrower than when the regression correction is not applied. In both cases the distributions for data and the simulated samples are in good agreement after the regression correction is applied.

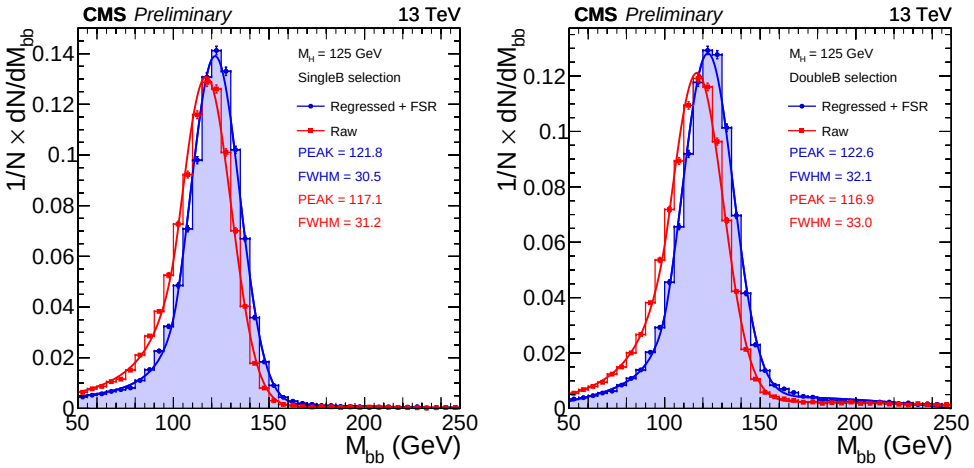

Figure 2: Simulated invariant mass distribution of the two b-jet candidates before and after the jet p_T regression, for VBF signal events. The generated Higgs boson signal mass is 125 GeV and the event selection corresponds to the SingleB (left) and DoubleB (right) samples. By FWHM we denote the width of the distribution at the middle of its maximum height.

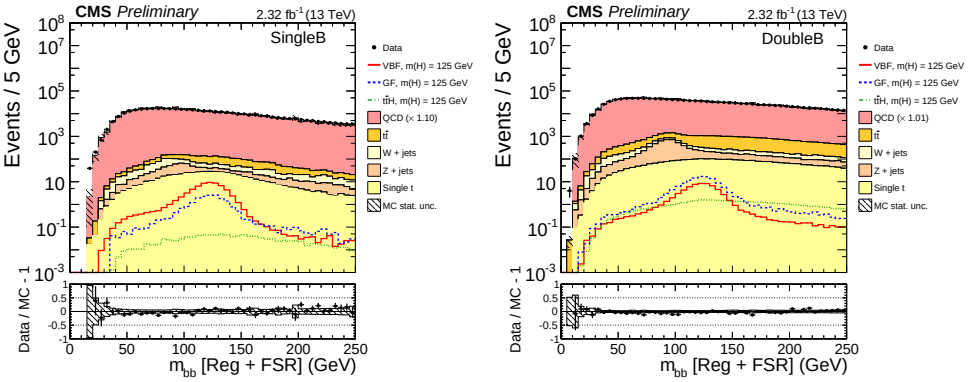

Figure 3: Distribution in invariant mass of the two b-jet candidates, after the jet p_T regression, for events in the SingleB (left) and DoubleB (right) samples. Data are shown by the points, while the simulated backgrounds are stacked. The LO QCD cross section is scaled such that the total number of background events matches the number of events in the data for each category, with a resulting multiplicative factor of 1.10 and 1.01 for the SingleB and DoubleB samples respectively. The panel at the bottom shows the fractional difference between the data and the background simulation, with the shaded band representing the statistical uncertainties in the MC samples.

7.2 Discrimination between quark- and gluon-originated jets

To further identify whether the two jets selected as q-jets candidates are likely to originate from the hadronization of a light (u,d,s-type) quark, as expected for signal VBF jets, or from gluons, as is more probable for jets produced in QCD processes, the jets internal properties

can be investigated. For this analysis, instead of a full quark-gluon discriminant [43, 44], we use a single jet-observable σ_2, the minor Root-Mean-Square (RMS) of the distribution of jet constituents in the η-ϕ plane [45].

Figure 4 shows the normalized distribution of σ_2 for the first VBF q-jet candidate, for background and signal events. As expected, VBF signal events, dominated by quark jets, have narrower jets, while the background and GF events are enriched in wider gluon jets. The σ_2 distributions of the two q-jets candidates are further used as input to the signal vs. background discriminants (Section 8.1).

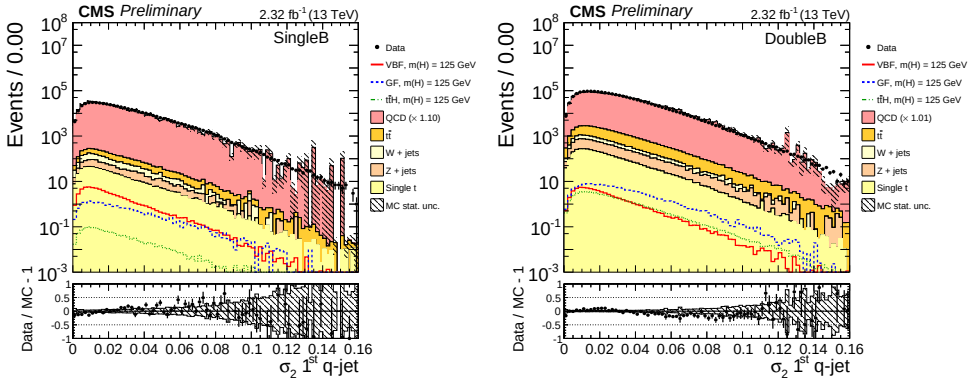

Figure 4: Distribution of σ_2, the minor RMS for the p_T-leading q-jet in the SingleB (left) and DoubleB (right) sample events. Data are shown by the points, while the simulated backgrounds are stacked. The LO QCD cross section is scaled such that the total number of background events matches the number of events in the data for each category, with a resulting multiplicative factor of 1.10 and 1.01 for the SingleB and DoubleB samples respectively. The panel at the bottom shows the fractional difference between the data and the background simulation, with the shaded band representing the statistical uncertainties in the MC samples.

7.3 Soft QCD activity

To measure the additional hadronic activity between the VBF-tagging jets, excluding the more centrally produced Higgs boson decay products, only reconstructed charged tracks are used. This is done to measure the hadronic activity associated with the primary vertex (PV), defined as the reconstructed vertex with the largest sum of squared transverse momenta of tracks used to reconstruct it.

A collection of "additional tracks" is assembled using reconstructed tracks that (i) satisfy the high purity quality requirements defined in Ref. [46] and $p_T > 300\,\text{MeV}$; (ii) are not associated with any of the four leading PF jets in the event; (iii) have a minimum longitudinal impact parameter, $|d_z(\text{PV})|$, with respect to the main PV, rather than to other pileup interaction vertices; (iv) satisfy $|d_z(\text{PV})| < 2\,\text{mm}$; and (v) are not in the region between the two best b-tagged jets. This is defined as an ellipse in the η-ϕ plane, centered on the midpoint between the two jets, with major axis of length $\Delta R(\text{bb}) + 1$, where $\Delta R(\text{bb}) = \sqrt{(\Delta\eta_{\text{bb}})^2 + (\Delta\phi_{\text{bb}})^2}$, oriented along the direction connecting the two b jets, and with minor axis of length 1.

The additional tracks are then clustered into "soft TrackJets" using the anti-k_T clustering algorithm with a distance parameter of 0.5. The use of TrackJets represents a clean and validated method [47] to reconstruct the hadronization of partons with very low energies down to a few GeV [48]; an extensive study of the soft TrackJet activity can be found in Refs. [43, 44].

For the purpose of separating the signal from the QCD background, we make use of clustered

soft TrackJets, and consider the soft TrackJet multiplicity N^{soft} with transverse momentum $p_{\text{T}} > 5$ GeV, N_5^{soft}; The discriminating variable N_5^{soft} is shown in Fig. 5.

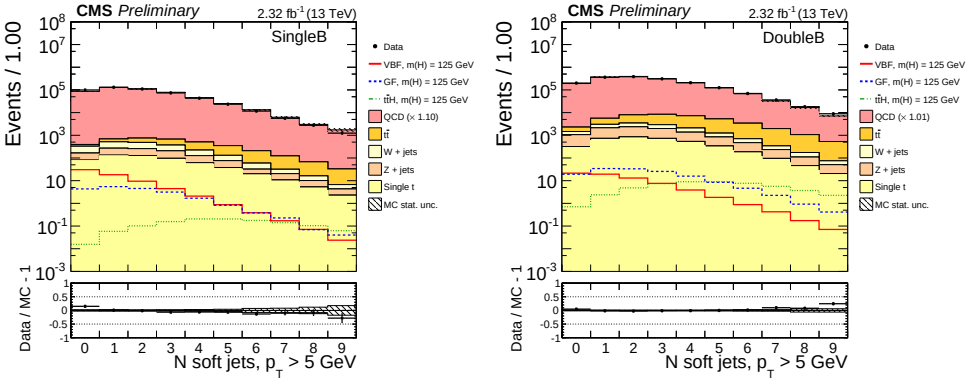

Figure 5: Distribution of the soft TrackJet multiplicity N_5^{soft} with transverse momentum $p_{\text{T}} > 5$ GeV in the SingleB (left) and DoubleB (right) sample events. Data are shown by the points, while the simulated backgrounds are stacked. The LO QCD cross section is scaled such that the total number of background events matches the number of events in the data for each category, with a resulting multiplicative factor of 1.10 and 1.01 for the SingleB and DoubleB samples respectively. The panel at the bottom shows the fractional difference between the data and the background simulation, with the shaded band representing the statistical uncertainties in the MC samples.

8 Search for a Higgs boson

The search for a Higgs boson follows closely the methodology applied in the Run 1 data analysis [13]. A multivariate discriminant is employed (Section 8.1) to divide the events into seven categories that are subsequently fit simultaneously with m_{bb} templates (Section 8.2).

8.1 Higgs boson signal vs. background discrimination

In order to separate the overwhelmingly large QCD background from the Higgs boson signal, all possible discriminating features have to be used in an optimal way. This is best achieved by using a multivariate discriminant, which in this case is a BDT implemented with the TMVA package. The variables used as an input to the BDT are chosen such that they are very weakly correlated with the dynamics of the $\text{b}\bar{\text{b}}$ system, in particular with m_{bb}, and are grouped into five distinct groups: (i) the dynamics of the VBF-jet system, expressed by $\Delta\eta_{\text{qq}}$, $\Delta\phi_{\text{qq}}$, and m_{qq}; (ii) the b-jet content of the event, expressed by the CSV output for the two b-jet candidates; (iii) the minor RMS of the two q-jet candidates; (iv) the gap activity, quantified by the number N_5^{soft} of soft TrackJets with $p_{\text{T}} > 5$ GeV, and the p_{T} of the fifth p_{T}-leading jet; (v) the angular dynamics of the production, expressed by the difference between the average pseudorapidity of the two q-jets and the average pseudorapidity of the two b-jets. (vi) the kinematics of the production, expressed by the total transverse and longitudinal momenta of the qqbb four-jet system.

In practice, two BDTs are trained with the same input variables using the selections corresponding to the two sets of events. This distinction is necessary because the properties of the selected events are significantly different between the two sets. Figure 6 shows the output of the BDT for the two sets of events.

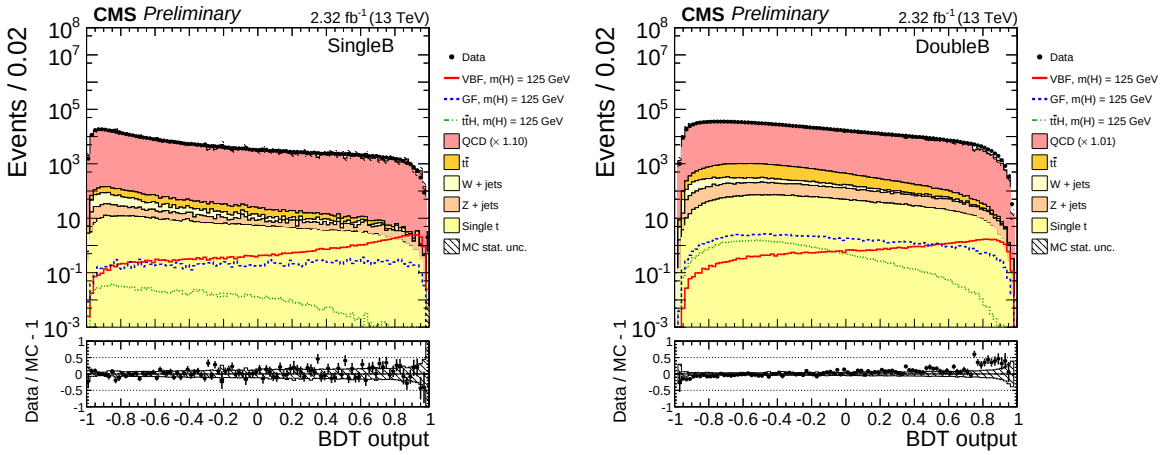

Figure 6: Distribution of the BDT output for the events in the SingleB (left) and DoubleB (right) sets. Data are shown by the points, while the simulated backgrounds are stacked. The LO QCD cross sections are scaled such that the total number of background events equals the number of events in data. The panels at the bottom show the fractional difference between the data and the background simulation, with the shaded band representing the statistical uncertainties of the MC samples.

Table 2: Definition of the event categories and corresponding yields in the m_{bb} interval $[80, 200]$ GeV, for the data and the MC expectation, with statistical uncertainties only. The BDT output boundary values refer to the distributions shown in Fig. 6.

BDT boundary values	SingleB				DoubleB		
	Cat. 1	Cat. 2	Cat. 3	Cat. 4	Cat. 5	Cat. 6	Cat. 7
	0.28 – 0.72	0.72 – 0.87	0.87 – 0.93	0.93 – 1.0	0.36 – 0.76	0.76 – 0.89	0.89 – 1.0
Data	25298	5834	1281	302	69963	9831	1462
Z +jets	49± 4	12.5± 2.0	4.1± 1.1	1.7± 0.7	448± 11	50± 4	8.4± 1.7
W+jets	25.8± 3.5	1.6± 0.9	0.1± 0.1	<0.1	74± 6	4.6± 1.3	0.9± 0.6
tt̄	53± 1	5.1± 0.2	0.7± 0.1	0.2± 0.04	534± 2	22.6± 0.4	1.1± 0.1
Single t	52± 1	9.7± 0.5	1.8± 0.2	0.4± 0.1	221± 3	23.2± 0.8	1.8± 0.2
VBF $m_{\mathrm{H}}(125)$	19.5± 0.2	13.7± 0.1	7.2± 0.1	4.2± 0.1	21.7± 0.2	10.5± 0.1	3.8± 0.1
GF $m_{\mathrm{H}}(125)$	5.5± 0.2	1.8± 0.1	0.6± 0.07	0.2± 0.04	18.7± 0.4	3.1± 0.1	0.6± 0.07

8.2 Fit of the dijet invariant mass spectrum

Taking into account the expected sensitivity of the analysis and the available number of MC events (necessary to build the various m_{bb} templates), seven categories are defined, according to the BDT output: four for SingleB and three for DoubleB. The boundaries of the categories and the respective event yields are summarized in Table 2. In an m_{bb} interval of twice the width of the Gaussian core of the signal distribution ($m_{\mathrm{H}} = 125$ GeV), the signal-over-background ratio reaches 3% in the most sensitive category (Category 4). It should be noted that both the VBF and GF contributions are added to the Higgs boson signal, with the fraction of the latter ranging from ∼40% in Category 5 to ∼5% in Category 4.

The analysis relies on the assumption that the QCD m_{bb} spectrum shape is the same in all BDT categories of the same set of events. In reality, a small correction is needed to account for residual differences between the m_{bb} spectrum in Category 1 vs. Categories 2,3,4, and in Category 5 vs. Categories 6,7. The correction factor (transfer function) is a linear function of m_{bb} in SingleB and a quadratic one in DoubleB (because a stronger dependence is observed in DoubleB between m_{bb} and the multivariate discriminant). With the introduction of the transfer

functions, the fit model for the Higgs boson signal is given by Eq. (1):

$$\begin{aligned} f_i(m_{\mathrm{bb}}) = \mu_{\mathrm{H}}\, N_{i,\mathrm{H}}\, H_i(m_{\mathrm{bb}}; k_{\mathrm{JES}}, k_{\mathrm{JER}}) + N_{i,\mathrm{Z}}\, Z_i(m_{\mathrm{bb}}; k_{\mathrm{JES}}, k_{\mathrm{JER}}) \\ + N_{i,\mathrm{t}}\, T_i(m_{\mathrm{bb}}; k_{\mathrm{JES}}, k_{\mathrm{JER}}) + N_{i,\mathrm{QCD}}\, K_i(m_{\mathrm{bb}})\, B(m_{\mathrm{bb}}; \vec{p}_{\mathrm{set}}), \end{aligned} \tag{1}$$

where the subscript i denotes the category and μ_{H}, $N_{i,\mathrm{QCD}}$ are free parameters for the signal strength and the QCD event yield. $N_{i,\mathrm{H}}$, $N_{i,\mathrm{Z}}$, and $N_{i,\mathrm{t}}$ are the expected yields for the Higgs boson signal, the Z +jets, and the top quark background respectively. The shape of the top quark background $T_i(m_{\mathrm{bb}}; k_{\mathrm{JES}}, k_{\mathrm{JER}})$ is taken from the simulation (sum of the $\mathrm{t\bar{t}}$ and single top quark contributions) and is described by a broad Gaussian. The Z/W+jets background $Z_i(m_{\mathrm{bb}}; k_{\mathrm{JES}}, k_{\mathrm{JER}})$ and the Higgs boson signal $H_i(m_{\mathrm{bb}}; k_{\mathrm{JES}}, k_{\mathrm{JER}})$ shapes are taken from the simulation and are parameterized as a Crystal ball function (Gaussian core with power-law tail) on top of a polynomial. The position and the width of the Gaussian core of the MC templates (signal and background) are allowed to vary within their uncertianties by the factors k_{JES} and k_{JER}, respectively, which quantify any mismatch of the jet energy scale and resolution between data and simulation. Finally, the QCD shape is described by a polynomial $B(m_{\mathrm{bb}}; \vec{p}_{\mathrm{set}})$, common within the categories of each set, and a multiplicative transfer function $K_i(m_{\mathrm{bb}})$ per category, accounting for the shape differences between the categories. The parameters of the polynomial, $\vec{p}_{\mathrm{set}}$, and those of the transfer functions, are determined by the fit, which is performed simultaneously in all categories in each set. For SingleB, the polynomial is of fifth order, while for DoubleB it is of fourth order.

9 Systematic uncertainties

Table 3 summarizes the sources of uncertainty related to both the background and to the signal processes. The leading uncertainty comes from the QCD background description: both the parameters of its shape and the overall normalization in each category are allowed to float freely, being determined by the simultaneous fit to the data. The resulting covariance matrix is used to compute the uncertainty. For the smaller background contributions from the Z/W+jets and top quark production, the m_{bb} shapes are taken from the simulation, while their corresponding yields are allowed to float in the fit with a 30% log-normal constraint centered on the SM expectations.

The experimental uncertainties on the JES and JER affect the signal acceptance and the shape of the multivariate discriminant output, and are included as nuisance parameters. The effect of the JES and JER uncertainties on the m_{bb} shape is taken into account in the fit function. By varying the JES and JER by their measured uncertainties [40], the impact of the signal yield per analysis category is estimated. These variations affect the acceptance by up to 10%, while the peak position of the m_{bb} shape is shifted by 2%, and the width by 2%.

Additional uncertainties are assigned to the flavor tagging of the jets. The CSV discriminant outputs are weighted according to the observed differences between data and simulation and the effect on the signal acceptance is estimated to range from 3% to 10%. The impact of the CSV shift is significant, both because it is used for the event selection, and because the multivariate discriminant depends more strongly on the b tagging of the jets.

The trigger uncertainty is estimated by propagating the uncertainty in the data vs. MC simulation scale factors for the efficiency. As a result, the uncertainty in the signal yield ranges from 8% to 15% for the VBF process, and from 6% to 11% for the GF.

Theoretical uncertainties affect the signal simulation. First, the uncertainty due to PDFs and strong coupling constant α_S variation is computed to be 2.1% (VBF) and 9% (GF) [49]. A resid-

ual uncertainty from these sources is estimated for the particular kinematical phase space of the search comparing the results to those obtained when using the sub-PDFs of the NNPDF set the PDF and α_S uncertainty ranges from 1% to 2%, while the renormalization and factorization scale variations in the signal simulation induce an uncertainty of around 1% (VBF) and 15% (GF) in the analysis categories, on top of a global cross section uncertainty of 0.4% (VBF) and 8% (GF). Finally, the variation of the UE and parton-shower (PS) model (using HERWIG [50, 51] instead of the default PYTHIA 8) affects the signal acceptance by 2% to 7% (VBF) and by 10% to 45% (GF). Lastly, an uncertainty of 2.7% is assigned to the total integrated luminosity measurement [52].

Table 3: Sources of systematic uncertainty and their impact on the shape and normalization of the background and signal processes.

Background uncertainties		
QCD shape parameters	determined by the fit	
QCD bkg. normalization	determined by the fit	
Top quark bkg. normalization	30%	
Z/W+jets bkg. normalization	30%	
Uncertainties affecting the signal	**VBF signal**	**GF signal**
JES (signal shape)	2%	
JER (signal shape)	2%	
Integrated luminosity	2.7%	
Branching fraction ($H \to b\bar{b}$)	1.3%	
JES (acceptance)	1–4%	2–11%
JER (acceptance)	1–2%	1–3%
b-jet tagging	3–9%	2–10%
Trigger	8–15%	6–11%
Theory uncertainties	**VBF signal**	**GF signal**
UE & PS	2–7%	10–45%
Scale variation (global)	0.4%	8%
Scale variation (categories)	1%	15%
PDF (global)	2%	3%
PDF (categories)	1–2%	1–2%

10 Results

The m_{bb} distributions in data, for all categories, are fitted simultaneously with the parametric functions described in Section 8.2 under two different hypotheses: background only and background plus a Higgs boson signal. The fit is a binned likelihood fit incorporating the systematic uncertainties discussed in Section 9 as nuisance parameters. Due to the smallness of the GF contribution in the most signal-sensitive categories we do not attempt to fit independently the VBF and the GF signal strengths. The fits of the data in the SingleB and DoubleB samples are shown in Figs. 7 and 8, respectively. The limits on the signal strength are computed with the asymptotic CL_s method [53–55]. For the 125 GeV Higgs boson signal the observed 95% confidence level upper limit is 3.0 times the standard model expectation, compared to 5.0 expected in absence of a signal.

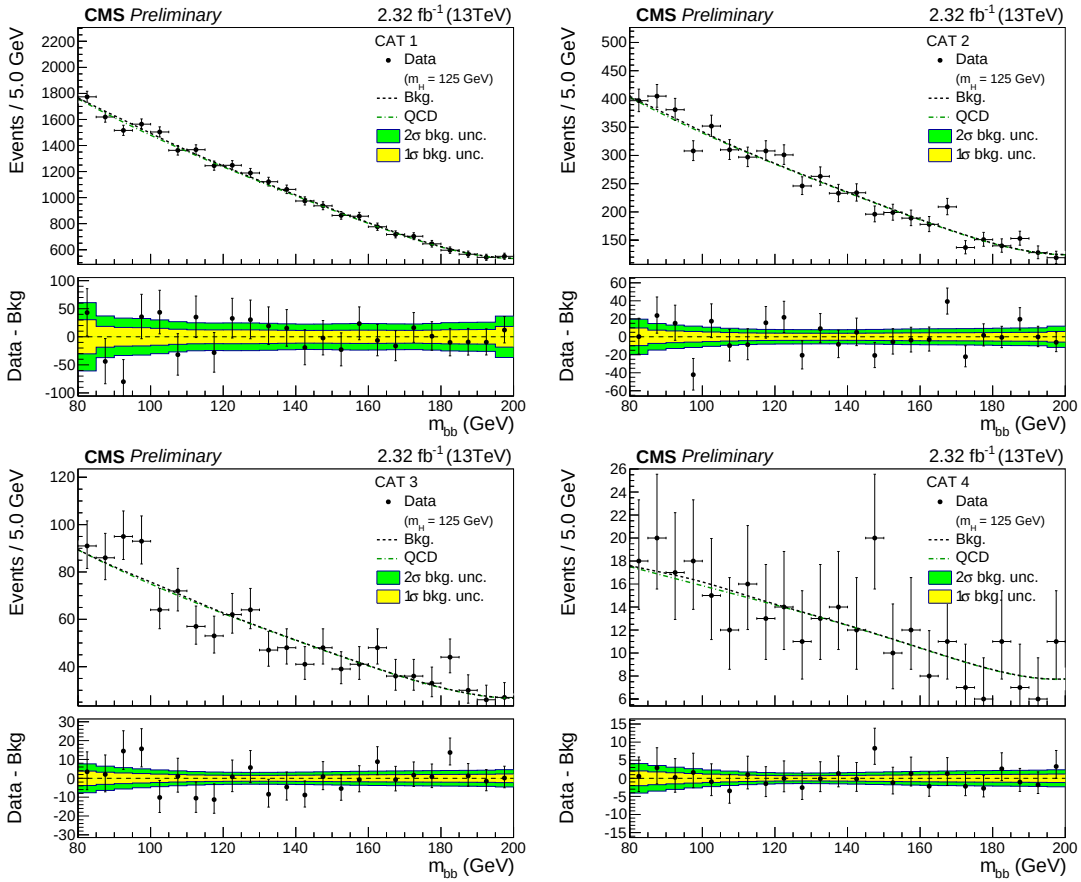

Figure 7: Fit of the invariant mass of the two b-jet candidates for the Higgs boson signal ($m_{\mathrm{H}} = 125$ GeV) in the four event categories of SingleB in the background-only hypothesis. Data are shown with markers. The dashed line is the background component, and the dashed-dotted line is the QCD component alone. The bottom panel shows the background-subtracted distribution, and with the 1σ and 2σ background uncertainty bands.

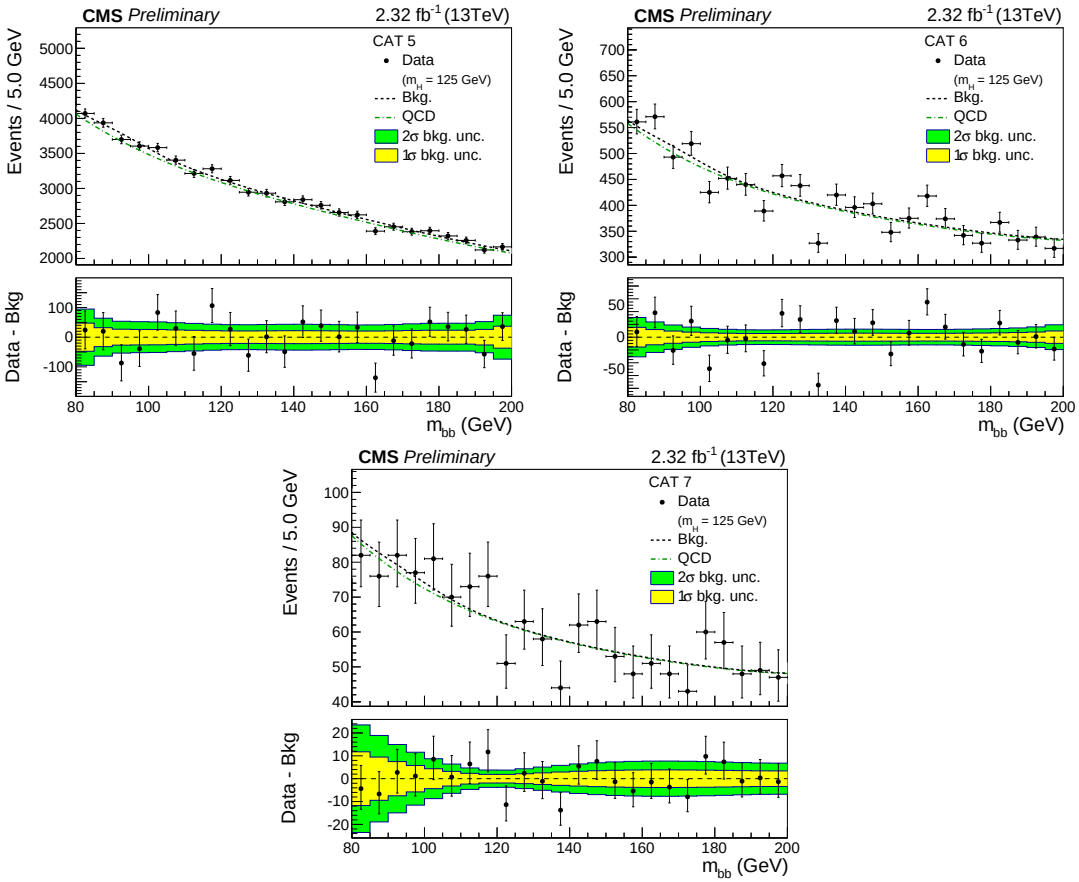

Figure 8: Fit of the invariant mass of the two b-jet candidates for the Higgs boson signal ($m_{\mathrm{H}} = 125$ GeV) in the three event categories of DoubleB in the background-only hypothesis. Data are shown with markers. The dashed line is the background component, and the dashed-dotted line is the QCD component alone. The bottom panel shows the background-subtracted distribution, and with the 1σ and 2σ background uncertainty bands.

11 Combination with Run 1 results

The results above have been combined with the previous CMS search for a SM Higgs boson produced through VBF and decaying to a bottom-quark pair on the Run 1 proton-proton collision data at $\sqrt{s} = 8$ TeV [13]. The combination methodology is based on the likelihood ratio test statistics employed in Section 10, and takes into account correlations among sources of systematic uncertainty. Theoretical uncertainties are assumed to be fully correlated among the two measurements.

For $m_{\mathrm{H}} = 125$ GeV the Run 2 data yields alone a fitted signal strength of $\mu = \sigma/\sigma_{\mathrm{SM}} = -3.7^{+2.4}_{-2.5}$, that is compatible with the SM Higgs boson prediction $\mu = 1$ at the 3% level. The combination of Run 1 and Run 2 results yields an observed (expected) upper limit of 3.4 (2.3) times the SM prediction, and a signal strength $\mu = 1.3^{+1.2}_{-1.1}$ with a significance of 1.2 standard deviations. Figure 9 shows the likelihood scans of the expected and observed best-fit signal strength separately with Run 1 8 TeV data, with Run 2 13 TeV data, and for their combination.

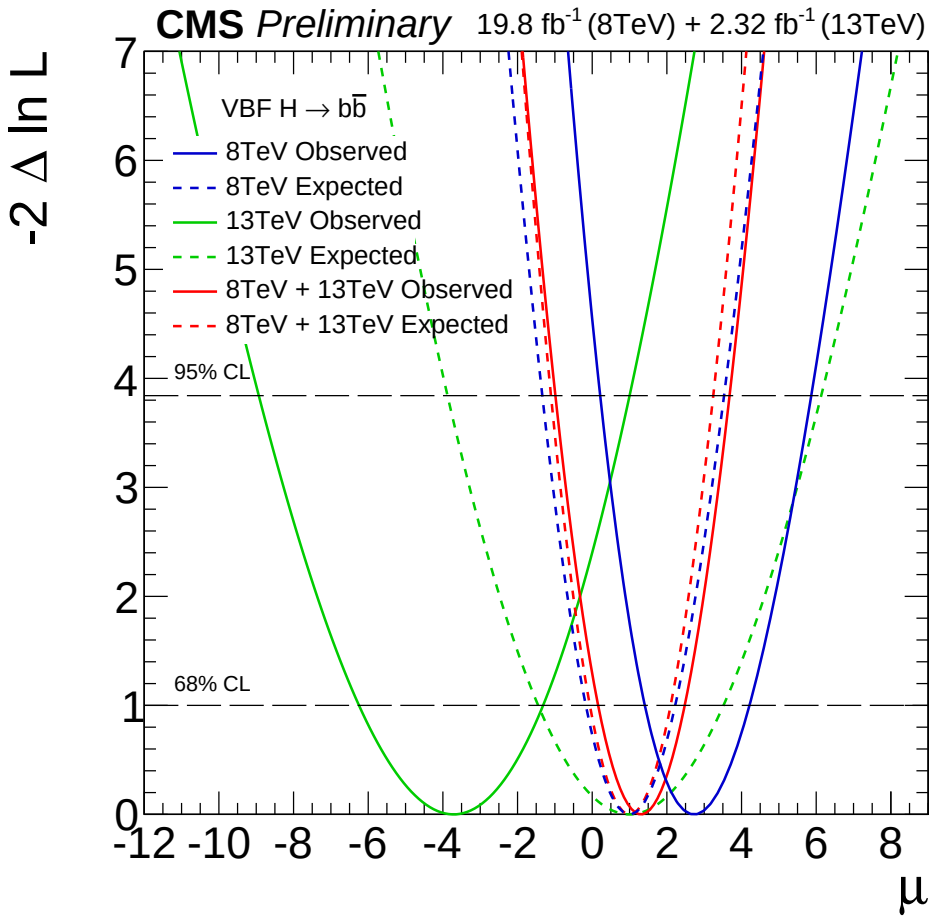

Figure 9: Observed and SM-expected likelihood profile of the signal strength $\mu = \sigma/\sigma_{\mathrm{SM}}$ with $m_{\mathrm{H}} = 125$ GeV, using Run 1 8 TeV data, Run 2 13 TeV data, and for the combination of 8 TeV and 13 TeV data.

12 Summary

A search has been carried out for the SM Higgs boson produced in vector boson fusion and decaying to $\mathrm{b\bar{b}}$ with a data sample of pp collisions at $\sqrt{s} = 13$ TeV collected with the CMS detector at the LHC corresponding to an integrated luminosity of $2.3\,\mathrm{fb}^{-1}$. Upper limits, at the 95% confidence level, on the production cross section times the $\mathrm{H} \to \mathrm{b\bar{b}}$ branching fraction, relative to expectations for a SM Higgs boson, are extracted for a Higgs boson with mass 125 GeV. The expected upper limits in the absence of a signal is 5.0 times the SM prediction, while the observed upper limit is 3.0, and the fitted signal strength is $\mu = \sigma/\sigma_{\mathrm{SM}} = -3.7^{+2.4}_{-2.5}$.

The combination of the results obtained in this search with the similar CMS search with Run 1 data yields an observed (expected) upper limit of 3.4 (2.3) times the SM prediction. The combined fitted $\mathrm{H} \to \mathrm{b\bar{b}}$ signal strength is $\mu = 1.3^{+1.2}_{-1.1}$, with a signal significance of 1.2 standard deviations for $m_{\mathrm{H}} = 125$ GeV.

References

[1] A. Salam and J. C. Ward, "On a Gauge Theory of Elementary Interactions", *Nuovo Cim.* **19** (1961) 165–170, doi:10.1007/BF02812723.

[2] S. Glashow, "Partial Symmetries of Weak Interactions", *Nucl.Phys.* **22** (1961) 579–588, doi:10.1016/0029-5582(61)90469-2.

[3] S. Weinberg, "A Model of Leptons", *Phys.Rev.Lett.* **19** (1967) 1264–1266, doi:10.1103/PhysRevLett.19.1264.

[4] F. Englert and R. Brout, "Broken Symmetry and the Mass of Gauge Vector Mesons", *Phys. Rev. Lett.* **13** (1964) 321–323, doi:10.1103/PhysRevLett.13.321.

[5] P. W. Higgs, "Broken Symmetries and the Masses of Gauge Bosons", *Phys. Rev. Lett.* **13** (1964) 508–509, doi:10.1103/PhysRevLett.13.508.

[6] G. S. Guralnik, C. R. Hagen, and T. W. B. Kibble, "Global Conservation Laws and Massless Particless", *Phys. Rev. Lett.* **13** (1964) 585–587, doi:10.1103/PhysRevLett.13.585.

[7] CMS Collaboration, "Observation of a new boson at a mass of 125 GeV with the CMS experiment at the LHC", *Phys. Lett.* **B716** (2012) 30–61, doi:10.1016/j.physletb.2012.08.021, arXiv:1207.7235.

[8] 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.* **B716** (2012) 1–29, doi:10.1016/j.physletb.2012.08.020, arXiv:1207.7214.

[9] ATLAS and CMS Collaboration, "Combined Measurement of the Higgs Boson Mass in pp Collisions at $\sqrt{s} = 7$ and 8 TeV with the ATLAS and CMS Experiments", *Phys. Rev. Lett.* **114** (2015) 191803, doi:10.1103/PhysRevLett.114.191803, arXiv:1503.07589.

[10] The ATLAS and CMS Collaborations, "Measurements of the Higgs boson production and decay rates and constraints on its couplings from a combined ATLAS and CMS analysis of the LHC pp collision data at $\sqrt{s}$ = 7 and 8 TeV",.

[11] LHC Higgs Cross Section Working Group et al., "Handbook of LHC Higgs Cross Sections: 2. Differential Distributions", *CERN-2012-002* (CERN, Geneva, 2012) arXiv:1201.3084.

[12] LHC Higgs Cross Section Working Group et al., "Handbook of LHC Higgs Cross Sections: 4. Deciphering the nature of the Higgs sector". https://twiki.cern.ch/twiki/bin/view/LHCPhysics/LHCHXSWG#SM_Higgs, CERN, Geneva, 2016, in preparation.

[13] CMS Collaboration, "Search for the standard model Higgs boson produced through vector boson fusion and decaying to $b\bar{b}$", *Phys. Rev.* **D92** (2015), no. 3, 032008, doi:10.1103/PhysRevD.92.032008, arXiv:1506.01010.

[14] CMS Collaboration, "Search for the standard model Higgs boson produced in association with W or Z bosons and decaying to bottom quarks", *Phys. Rev. D* **89** (2014) 012003, doi:10.1103/PhysRevD.89.012003, arXiv:1310.3687.

[15] CMS Collaboration, "Search for the standard model Higgs boson produced in association with a top-quark pair in pp collisions at the LHC", *JHEP* **05** (2013) 145, doi:10.1007/JHEP05(2013)145, arXiv:1303.0763.

[16] CMS Collaboration, "Search for the associated production of the Higgs boson with a top-quark pair", *JHEP* **09** (2014) 087, doi:10.1007/JHEP09(2014)087,10.1007/JHEP10(2014)106, arXiv:1408.1682. [Erratum: JHEP10,106(2014)].

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

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

[19] 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.

[20] 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.

[21] 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.

[22] T. Sjöstrand, S. Mrenna, and P. 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.

[23] T. Sjöstrand et al., "An Introduction to PYTHIA 8.2", *Comput. Phys. Commun.* **191** (2015) 159–177, doi:10.1016/j.cpc.2015.01.024, arXiv:1410.3012.

[24] CMS Collaboration, "Event generator tunes obtained from underlying event and multiparton scattering measurements", arXiv:1512.00815.

[25] 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.

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

[27] E. Re, "Single-top Wt-channel production matched with parton showers using the POWHEG method", *Eur. Phys. J.* **C71** (2011) 1547, doi:10.1140/epjc/s10052-011-1547-z, arXiv:1009.2450.

[28] R. Gavin, Y. Li, F. Petriello, and S. Quackenbush, "FEWZ 2.0: A code for hadronic Z production at next-to-next-to-leading order", *Comput. Phys. Commun.* **182** (2011) 2388, doi:10.1016/j.cpc.2011.06.008, arXiv:1011.3540.

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

[30] R. Gavin, Y. Li, F. Petriello, and S. Quackenbush, "W physics at the LHC with FEWZ 2.1", *Comput. Phys. Commun.* **184** (2013) 208, doi:10.1016/j.cpc.2012.09.005, arXiv:1201.5896.

[31] M. Czakon, P. Fiedler, and A. Mitov, "Total Top-Quark Pair-Production Cross Section at Hadron Colliders Through O(α_S^4)", *Phys. Rev. Lett.* **110** (2013) 252004, doi:10.1103/PhysRevLett.110.252004, arXiv:1303.6254.

[32] N. Kidonakis, "Differential and total cross sections for top pair and single top production", (2012). arXiv:1205.3453.

[33] CMS Collaboration, "Identification of b-quark jets with the CMS experiment", *JINST* **8** (2013) P04013, doi:10.1088/1748-0221/8/04/P04013, arXiv:1211.4462.

[34] CMS Collaboration Collaboration, "Identification of b quark jets at the CMS Experiment in the LHC Run 2", Technical Report CMS-PAS-BTV-15-001, CERN, Geneva, 2016.

[35] CMS Collaboration, "Particle–flow event reconstruction in CMS and performance for jets, taus, and $E_{\mathrm{T}}^{\mathrm{miss}}$", CMS Physics Analysis Summary CMS-PAS-PFT-09-001, CERN, 2009.

[36] CMS Collaboration, "Commissioning of the Particle-flow Event Reconstruction with the first LHC collisions recorded in the CMS detector", CMS Physics Analysis Summary CMS-PAS-PFT-10-001, CERN, 2010.

[37] CMS Collaboration, "Commissioning of the particle-flow reconstruction in minimum-bias and jet events from pp collisions at 7 TeV", CMS Physics Analysis Summary CMS-PAS-PFT-10-002, CERN, 2010.

[38] M. Cacciari and G. P. Salam, "Dispelling the N^3 myth for the k_t jet-finder", *Phys. Lett. B* **641** (2006) 57, doi:10.1016/j.physletb.2006.08.037, arXiv:hep-ph/0512210.

[39] 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.

[40] CMS Collaboration, "Determination of Jet Energy Calibration and Transverse Momentum Resolution in CMS", *JINST* **6** (2011) P11002, doi:10.1088/1748-0221/6/11/P11002, arXiv:1107.4277.

[41] CMS Collaboration, "Pileup jet identification", CMS Physics Analysis Summary CMS-PAS-JME-13-005, CERN, 2013.

[42] H. Voss, A. Höcker, J. Stelzer, and F. Tegenfeldt, "TMVA, the Toolkit for Multivariate Data Analysis with ROOT", in *XIth International Workshop on Advanced Computing and Analysis Techniques in Physics Research (ACAT)*, p. 40. 2007. arXiv:physics/0703039.

[43] CMS Collaboration, "Measurement of the hadronic activity in events with a Z and two jets and extraction of the cross section for the electroweak production of a Z with two jets in pp collisions at $\sqrt{s}$ = 7 TeV", *JHEP* **10** (2013) 062, doi:10.1007/JHEP10(2013)062, arXiv:1305.7389.

[44] CMS Collaboration, "Measurement of electroweak production of two jets in association with a Z boson in proton-proton collisions at $\sqrt{s} = 8$ TeV", *Eur. Phys. J.* **C75** (2015), no. 2, 66, doi:10.1140/epjc/s10052-014-3232-5, arXiv:1410.3153.

[45] CMS Collaboration, "Performance of quark/gluon discrimination in 8 TeV pp data", CMS Physics Analysis Summary CMS-PAS-JME-13-002, CERN, 2013.

[46] CMS Collaboration, "Description and performance of track and primary-vertex reconstruction with the CMS tracker", *JINST* **9** (2014), no. 10, P10009, doi:10.1088/1748-0221/9/10/P10009, arXiv:1405.6569.

[47] CMS Collaboration, "Commissioning of trackjets in pp collisions at $\sqrt{s}$=7 TeV", CMS Physics Analysis Summary CMS-PAS-JME-10-006, CERN, 2010.

[48] CMS Collaboration, "Performance of jet reconstruction with charged tracks only", CMS Physics Analysis Summary CMS-PAS-JME-08-001, CERN, 2009.

[49] LHC Higgs Cross Section Working Group et al., "Handbook of LHC Higgs Cross Sections: 3. Higgs Properties", *CERN-2013-004* (CERN, Geneva, 2013) arXiv:1307.1347.

[50] P. Richardson and A. Wilcock, "Monte Carlo Simulation of Hard Radiation in Decays in Beyond the Standard Model Physics in Herwig++", *Eur. Phys. J.* **C74** (2014) 2713, doi:10.1140/epjc/s10052-014-2713-x, arXiv:1303.4563.

[51] J. Bellm et al., "Herwig++ 2.7 Release Note", arXiv:1310.6877.

[52] CMS Collaboration, "CMS Luminosity Measurement for the 2015 Data Taking Period", CMS Physics Analysis Summary CMS-PAS-LUM-15-001, CERN, 2016.

[53] T. Junk, "Confidence level computation for combining searches with small statistics", *Nucl. Instrum. Meth.* **A434** (1999) 435–443, doi:10.1016/S0168-9002(99)00498-2, arXiv:hep-ex/9902006.

[54] A. L. Read, "Presentation of search results: The CL(s) technique", *J. Phys.* **G28** (2002) 2693–2704, doi:10.1088/0954-3899/28/10/313. [,11(2002)].

[55] G. Cowan, K. Cranmer, E. Gross, and O. Vitells, "Asymptotic formulae for likelihood-based tests of new physics", *Eur. Phys. J.* **C71** (2011) 1554, doi:10.1140/epjc/s10052-011-1554-0, 10.1140/epjc/s10052-013-2501-z, arXiv:1007.1727. [Erratum: Eur. Phys. J.C73,2501(2013)].