CMS Physics Analysis Summary

Search for a standard model like Higgs boson in the 
$H \rightarrow ZZ \rightarrow \ell^+\ell^- q\bar{q}$ decay channel at $\sqrt{s}=8$ TeV

The CMS Collaboration

Abstract

A search for a standard-model-like Higgs boson decaying into two Z bosons with subsequent decay into two leptons and two quarks, $H \rightarrow ZZ \rightarrow \ell^+\ell^- q\bar{q}$ is performed. The analysis uses 19.7 fb$^{-1}$ of data collected by the CMS experiment from proton-proton collisions produced in LHC at $\sqrt{s} = 8$ TeV. The search exploits the kinematic information and the flavor tagging of the leading particles of the event in order to isolate hypothetical Higgs boson-like signals in the 230 GeV to 1000 GeV mass range. At high mass the two quarks from the Z-decay merge into a single jet and are analyzed with jet-substructure methods. The sensitivity of the analysis is further improved compared to previous similar searches by a dedicated study of a Vector Boson Fusion signatures. We interpret the data in terms of a standard-model-like Higgs boson as well as an electroweak singlet visible through the interference with the recently discovered 125 GeV boson. No evidence of a signal is found and upper limits are set on the production cross section.
1 Introduction

The Higgs mechanism is an essential element of the standard model (SM) of particles and their interactions explaining the origin of mass and playing a key role in electroweak symmetry breaking. A suitable candidate has recently been found with a mass around 126 GeV [1, 2]. However, several models of physics beyond the SM predict a more complex Higgs sector than the SM, which may contain additional particles coupling to vector bosons [3]. Of particular interest here is the mixing of the SM Higgs boson with an electroweak singlet [4], which appears as an additional particle with Higgs-like couplings. As shown in [5], a direct search holds a promise to better constrain such models compared to indirect results derived from the measured properties of the 126 GeV boson. Thus we present here further searches for Higgs-like particles using the SM Higgs as a benchmark model.

In this note we report on a search for a Higgs-like boson in $H \rightarrow ZZ$ when one of $Z$ decays as $Z \rightarrow \ell^+\ell^-$ (where $\ell$ is either an electron or a muon) and the other as $Z \rightarrow q\bar{q}$, using LHC proton-proton collision data at $\sqrt{s} = 8$ TeV. Previous searches for a resonant decay to $H \rightarrow ZZ \rightarrow \ell^+\ell^-q\bar{q}$ include recent results from CMS [6] and ATLAS [7], and a similar search for a graviton decays by CDF [8], CMS [9] and ATLAS [10].

The analysis is performed by studying the mass of the ZZ system, $M_{\ell\ell jj}$, which is expected to exhibit a narrow peak for hypothetical signals at low mass, significantly increasing in width at high mass. The dominant background is $Z+jets$ production from Drell-Yan processes, with smaller contributions from top-quark and diboson decays. Unlike for the signal, the $M_{\ell\ell jj}$ distribution of the background events is not resonant, providing a useful handle for isolating signal events. The analysis exploits the kinematic information and the flavor tagging of the leading particles of the event to enhance a hypothetical signal over the contamination. Data from a signal-free control region are used to constrain the normalization and tune the dominant Drell-Yan plus jets background reducing the dependence on the simulation. The contamination from $t\bar{t}$ events is directly extracted from data.

The Higgs boson may be produced through gluon-gluon fusion (ggH) or vector boson fusion (VBF) processes, where the VBF production plays a small role at low Higgs masses, but reaches almost 50% of the total production cross section at a Higgs mass of 1 TeV. In the VBF process two additional jets are produced in the forward directions, which occurs rarely for the dominant backgrounds. We make use of this topology by separating events with suitable forward jets into a VBF category with improved signal to background ratio.

For high Higgs boson masses, the quarks from the hadronic decay of a $Z$ boson are sufficiently boosted for the two jets to be reconstructed as a single jet. To include these events in the search and avoid efficiency losses, we also consider events that contain two leptons and a single jet with a substructure compatible with the hadronic decay of a $Z$ boson.

This study is an improvement over [11], widening the search range to resonance masses up to 1 TeV and adding the tagging of Vector Boson Fusion (VBF) events.

2 The CMS detector and simulation

A detailed description of the CMS detector can be found in Ref. [12]. The central feature of the CMS detector is a 3.8 T superconducting solenoid of 6 m internal diameter. Within the field volume are the silicon tracker, the crystal electromagnetic calorimeter (ECAL), and the brass-scintillator hadron calorimeter (HCAL). The muon system is installed outside the solenoid and embedded in the steel return yoke. The CMS tracker [13] consists of 1440 silicon pixel and
15,148 silicon strip detector modules. The ECAL consists of nearly 76,000 lead tungstate crystals, which provide coverage in pseudorapidity $|\eta| < 1.479$ in the central barrel region and $1.479 < |\eta| < 3.0$ in the two forward endcap regions. The ECAL has an energy resolution of better than 0.5% for unconverted photons with transverse energies above 100 GeV. The HCAL [14] consists of a set of sampling calorimeters which utilize alternating layers of brass as absorber and plastic scintillator as active material. The muon system includes barrel drift tubes covering the pseudorapidity range $|\eta| < 1.2$, endcap cathode strip chambers ($0.9 < |\eta| < 2.5$), and resistive plate chambers ($|\eta| < 1.6$) [15]. The pseudorapidity $\eta$ is defined as $\ln \cot(\theta/2)$ with the polar angle $\theta$ measured in the laboratory frame. The first level of the CMS trigger system, composed of custom hardware processors, is designed to select the most interesting events in less than 3 $\mu$s, using information from the calorimeters and muon detectors. The High Level Trigger processor farm further reduces the event rate to a few hundred Hz before data storage.

Background simulation samples are generated using MADGRAPH V5_1.3.30 [16] for Z+jets, POWHEG [17–19] for $t\bar{t}$, and PYTHIA 6.4.22 [20] for an alternative $t\bar{t}$ sample and for diboson samples. Signal events are generated using POWHEG [17–19].

The signal simulation does not include the propagator effects discussed in Refs. [21–23], which describe the Complex-Pole Scheme (CPS). This effect is important for high Higgs boson masses ($> 400$ GeV) and included in the samples here via a reweighting of the simulated samples. Moreover, the effects on the line shape due to the interference between Higgs-boson signal and the $gg \rightarrow ZZ$ background have been included [24] in the gluon-gluon fusion production channel.

Parton distribution functions (PDF) are modeled using the parameterizations CT10 at NLO [25] and CTEQ6 [26] at LO. For both signal and background Monte Carlo (MC) samples, events are simulated using a GEANT4-based model [27] of the CMS detector and processed using the same reconstruction algorithms as for data.

3 Event reconstruction and selection

We search for a fully reconstructed decay expected for a SM Higgs-like boson $H \rightarrow ZZ \rightarrow \ell^+\ell^-q\bar{q}$, where the charged leptons $\ell^\pm$ are either muons or electrons. Quarks are identified as two jets or a single jet with large momentum transverse to the pp beam direction ($p_T$). Events are recorded with triggers that require two isolated leptons. Leptons are required to have $p_T$ above 17 (8) GeV for the leading (subleading) lepton respectively. Electrons are additionally required to pass loose isolation and track-cluster matching requirements. The trigger efficiency for signal events that pass the full event selection is estimated to be about 88% in the $\mu^+\mu^-$ final state and about 94% in the $e^+e^-$ final state.

The data are classified into several categories, depending on lepton flavor, jet number and VBF tagging criteria to optimize the sensitivity of the search over a wide mass range. In the following we discuss common selection criteria before describing the categories in detail.

3.1 Lepton reconstruction and selection

The details of electron and muon identification criteria are described elsewhere [28]. Muons are measured with the all-silicon tracker and the muon system. Electrons are detected as trajectories in the tracker pointing to energy clusters in the ECAL. Both muons and electrons are required to have $p_T$ greater than 20 GeV and 40 GeV, for the lower and higher $p_T$ lepton, respectively. Muons (electrons) in the pseudorapidity range $|\eta| < 2.4$ ($|\eta| < 2.5$) are consid-
3.2 Jet reconstruction and selection

Electrons in the transition region between the barrel and endcap, $1.44 < |\eta| < 1.57$, are excluded. Both the $p_T$ and $\eta$ requirements are consistent with those in the online trigger algorithms. Leptons are required to be isolated from hadronic activity in the event. Isolation is defined by using the charged and neutral particles in a cone of $\Delta R \equiv \sqrt{(\Delta \eta)^2 + (\Delta \phi)^2} < 0.4$ ($0.3$) around the muon (electron) direction, where $\Delta \eta$ and $\Delta \phi$ are the differences in pseudorapidity and in azimuthal angle in radians, respectively. The individual particles considered in the isolation sum are reconstructed with the particle flow (PF) [29, 30] algorithm, which, through the combination of information from all sub-detectors, aims to reconstruct all particles produced in a collision event.

The isolation sum $S_{\ell}^{\text{Iso}}$ is defined as the scalar sum of the transverse momenta of charged hadrons originating from the primary vertex, plus the scalar sums of the transverse energies for neutral hadrons and photons, respectively. This sum is corrected to take into account the contribution from neutrals originating from additional pp collisions in the same bunch crossing (pile-up). The number of true interactions per bunch crossing in the simulated samples is weighed to match the distributions in data events, so that pile-up effects in the simulation are a good representation of those in real collisions. Muons (electrons) are considered isolated if $S_{\ell}^{\text{Iso}}$ is less than 12% (15%) of the measured $p_T$. Reconstruction efficiencies for leptons and their uncertainties are evaluated from data with a tag-and-probe [28] technique, where one lepton from an inclusive sample of Z decays serves as a tag and the efficiency for the reconstruction of the other lepton is measured.

Each pair of oppositely charged leptons is considered to be a Z candidate. The dilepton invariant mass $m_{\ell\ell}$ is restricted to $76 < m_{\ell\ell} (\text{GeV}) < 106$ to reduce the combinatorial background from lepton pairs such as $t\bar{t}$ that do not have a real Z while having a large acceptance to signal events.

3.2 Jet reconstruction and selection

Reconstructed particles are clustered to form PF jets with the anti-$k_T$ algorithm [31] with radius parameter $R = 0.5$ (AK5 jets) and with the Cambridge-Aachen algorithm [32] with $R = 0.8$ (CA8 jets), as implemented in the FASTJET package [33]. Any reconstructed jet overlapping with isolated leptons within $\Delta R = 0.5$ (0.8) for AK5 (CA8) jets is removed in order to avoid double counting of the lepton as a jet.

Jet-energy corrections are applied to account for the non-linear response of the calorimeters to the particle energies and other instrumental effects. These corrections are based on in-situ measurements using dijet and $\gamma^* / Z + \text{jet}$ data samples [34]. Uncertainties due to these corrections are evaluated by varying the jet energy within calibration uncertainties.

Pile-up and the underlying event affect the jet reconstruction by contributing additional energy to the reconstructed jets. The median energy density due to pile-up is evaluated in each event and the corresponding energy is subtracted from each jet [35, 36]. A requirement based on the energy balance between charged and neutral hadrons in a jet is applied to remove misidentified jets. Furthermore, in order to remove jets originating from pile-up interactions, we select only jets having $\beta \geq 0.2$, where $\beta$ is defined as the sum of transverse momenta of all charged particles in the jet coming from the primary vertex, normalized to the total sum of transverse momenta of all charged particles in the jet. The corresponding uncertainty is studied by varying by 5% the minimum-bias cross section used to estimate the number of pile-up interactions. The central value for this cross section is 69.30 mb.

For high resonance masses and corresponding high $Z p_T$, the two quarks from the Z decay are
expected to be reconstructed as a single CA8 jet. Thus we consider single CA8 jets as additional hadronic Z candidates. To improve background rejection and jet mass resolution, we apply a jet-pruning algorithm described in detail in [37]. Additionally, the “n-subjettiness” variable $\tau_{21}$ [37] is used to reduce the backgrounds. This defines a measure, $\tau_N$, of the compatibility of a jet having $N$ subjets. The value of $\tau_N$ tends to zero as a jet becomes more consistent with $N$ subjets. For discriminating merged Z jets with two subjets and QCD jets with a single subjet, the ratio of $\tau_{21}$ has been found to be especially effective [37]. The distributions for $\tau_{21}$ and $M_{\text{pruned}}$ are shown in Fig. 1.

Jets considered for the reconstruction of a Higgs candidate are required to be inside the tracker acceptance ($|\eta| < 2.4$), while AK5 jets up to $|\eta| < 4.7$ are considered to tag the VBF event topology. After all corrections, a selection requirement $p_T > 30$ GeV is applied to all accepted jets. Each pair of jets is considered as a Z candidate. The dijet invariant mass $m_{jj}$ is restricted to $71 < m_{jj}$(GeV) $< 111$ to reduce the Z+jets background. Data with dijet masses in the ranges $[60, 71]$ GeV and $[111, 130]$ GeV, which we refer as sidebands, are used to validate and normalize the simulation in the signal region. The $m_{jj}$ distribution is displayed in Fig. 2. Additionally, in events with $p_T^{\ell\ell} > 200$ GeV we consider individual CA8 jets with $p_T^{\ell} > 100$ GeV as hadronic Z candidates. We require $\tau_{21} < 0.5$ to avoid contamination from jets originating from the hadronization of gluons and single quarks. CA8 jets with a pruned jet mass [37] $71 \leq M_{\text{pruned}} \leq 111$ GeV and $p_T^{\ell} > 100$ GeV are considered as signal candidates. Candidates with pruned jet mass within $60 \leq M_{\text{pruned}} \leq 130$ GeV, but excluding the signal region are used as a sideband for background studies. Events are considered for further analysis if they contain at least one leptonic and one hadronic Z candidate.

The parton flavor of the jets provides a powerful tool for background discrimination. Jets from the hadronic Z decay in the signal have higher b-quark content than jets from the Z+jets background. The flavor of quarks in Z decays is almost equally distributed among the five
3.3 Event categories and specialized selections

We use the hadronic and leptonic Z candidates to reconstruct a Higgs boson candidate. In order to obtain the best possible sensitivity, events are classified according to the lepton flavor, hadronic Z type (dijet or merged jet), b-tag criteria and the possible presence of VBF tagging jets. Further selections are applied to suppress backgrounds, some of which are specific to the event category in question. The categories are constructed to be exclusive. When more than one Higgs boson candidate is present in an event, the best candidate is chosen according to the prescription below. The spectra of the reconstructed Higgs boson mass are used to set limits on or possibly discover a new resonance using a combined likelihood function.
If the event contains two additional AK5 jets with pseudorapidity difference $\Delta \eta > 3.5$ and an invariant mass above 500 GeV, the event is considered as a VBF candidate. Events that do not contain suitable tagged jets are further categorized by the type of hadronic Z decay: events with $p_T^{\ell\ell} > 200$ GeV, that contain a CA8 jet passing the $t_{21}$ and $M_{\text{pruned}}$ requirements discussed above are considered as merged candidates. The remaining events constitute the dijet category. The merged and dijet categories are further split into three exclusive b-tag categories each. Events in the 2 b-tag category are required to have one jet identified with medium ($\sim 65\%$ efficiency) and the other jet with loose ($\sim 80\%$ efficiency) requirements. Events not selected in the 2 b-tag region are categorized as 1 b-tag if they have one jet satisfying the loose-tag requirements. The 0 b-tag category contains the remaining events. In the VBF category, b-tagging information is only used to determine the best candidate. The candidate with the highest number of b-tags is preferred, should more than one Higgs boson candidate be reconstructed in the event.

Candidates that satisfy these selection criteria are classified into 14 exclusive categories for further analysis, which are grouped into three broad categories:

- **VBF category**: 2 VBF channels split by lepton flavor. Candidates with dijets and merged jets are combined into a single category, regardless of the number of b-tags in the event. The b-tagging information is still used to choose the best candidate in case more than one Higgs boson candidate is found.

- **Merged jet category**: 6 high mass channels, using single CA8 jets and vetoing VBF candidates, split into two lepton flavors and three b-tag categories.

- **Dijet category**: 6 low mass channels, using dijets constructed from AK5 jets and vetoing VBF and merged candidates, split into two lepton flavors and three b-tag categories.

Since a Higgs-like boson carries no spin, the angular distribution of its decay products is independent of the production mechanism. Five angles ($\theta^*, \theta_1, \theta_2, \Phi_1, \Phi$) defined in the Higgs boson rest frame (see Ref. [44, 45]) are used to fully describe the kinematics of the $gg \rightarrow H \rightarrow ZZ \rightarrow \ell^+\ell^- q\bar{q}$ decay. The variables are only weakly correlated with the invariant masses of the H and the two Z bosons, and with the longitudinal and transverse momenta of the boson candidate and provide a good discriminating power between signal and background. We construct an angular likelihood discriminant (LD) based on the probability ratio of the signal and background hypotheses, described in detail in Ref. [6]. Distributions of the angular likelihood discriminant for signal and background are shown in Fig. 3. The disagreement in the background-dominated region with LD < 0.4 in the dijet analysis is due to the mis-modeling of the $\cos \theta^*$ distribution for Z+jets events at $|\cos \theta^*| > 0.85$ in the MC samples. In the case of the merged jet final state, the angles are computed from the subjets provided by the CA8 jet clustering algorithm. All selected events are required to have a value of the likelihood discriminant larger than 0.5 (0.2) for the dijet and merged (VBF) categories, which reduces the Z+jets background by about a factor of two, while retaining a signal efficiency of 75% to 90%, depending on $M_H$ hypothesis and analysis category.

Unlike $t\bar{t}$ background which has significant missing transverse energy, $E_T$, signal events are typically well-balanced in transverse momentum. Thus the $t\bar{t}$ background is suppressed by a $E_T$-based selection. The so-called $E_T$ significance variable, $\lambda$, is constructed using the resolution functions of the individual $E_T$ inputs and can be interpreted as significance of the measured $E_T$ [46]. We apply a loose requirement, $\lambda < 10$, in the dijet category, where $t\bar{t}$ background is most prominent.

For VBF events we additionally construct a multivariate discriminant based on the following
3.3 Event categories and specialized selections

Figure 3: Distribution of the likelihood discriminant in the dijet (left), merged jet (middle) and VBF (right) categories. In addition to data (black points) and background simulation (filled histograms) as well as the top background estimate derived from $e\mu$ data (in the dijet and merged categories), we show SM Higgs boson simulations, scaled for better visibility. The gray band shows the background statistical and systematic uncertainties. Background normalizations are derived from the fit performed to extract a hypothetical signal.

Figure 4: Distribution of the multivariate VBF discriminant in MC and data, for all categories combined, after all selections excluding the VBF discriminant itself. The gray band shows the background statistical uncertainties. Backgrounds are normalized to data.

variables:

- $\Delta \eta$ between the VBF jets, VBF jets invariant mass, $\Delta \phi$ between the VBF jets;
- Leading VBF jet energy, $p_T$ and pseudorapidity;
- Second VBF jet energy, $p_T$ and pseudorapidity;

and require the discriminant (shown in Fig. 4) to have a value above 0.4. This cut has an efficiency between 70% and 80% for VBF signals and around 20% for signals produced in gluon fusion processes. The efficiency on backgrounds is 10%. These efficiencies are relative to the number of events that pass the full selection excluding the VBF discriminant.

The main selection requirements are summarized in Tables 1–4. When an event contains mul-
4 Background determination

The three main background components, in order of importance, are: Z+jets, \( t\bar{t} \) and diboson. Other contributions, such as W+jets or QCD multijet events with jets misidentified as leptons, are negligible due to the tight dilepton selection. The small diboson contribution is taken directly from simulation, but its effect is negligible due to the small cross section of these processes. Other backgrounds are treated slightly differently in the three categories.

Figure 5: Distribution of the Jet Probability b-tagging discriminator after final selection for dijets (left), and the hardest subjet of the merged jet (right). The red dashed histogram indicates the expected distribution for a Higgs boson hypothesis. The gray band shows the background statistical and systematic uncertainties. Background normalizations are derived from the fit performed to extract a hypothetical signal.

Table 1: Summary of common selection requirements.

<table>
<thead>
<tr>
<th>Requirement</th>
<th>preselection</th>
</tr>
</thead>
<tbody>
<tr>
<td>( p_T(\ell^\pm) )</td>
<td>lowest ( p_T &gt; 20 \text{ GeV} ), highest ( p_T &gt; 40 \text{ GeV} )</td>
</tr>
<tr>
<td>( p_T(\text{jets}) )</td>
<td>( &gt; 30 \text{ GeV} )</td>
</tr>
<tr>
<td>(</td>
<td>\eta</td>
</tr>
<tr>
<td>(</td>
<td>\eta</td>
</tr>
<tr>
<td>b-tag</td>
<td>0 b-tag</td>
</tr>
<tr>
<td></td>
<td>1 b-tag</td>
</tr>
<tr>
<td></td>
<td>2 b-tag</td>
</tr>
<tr>
<td>( m_{jj} )</td>
<td>( \in [71, 111] \text{ GeV} )</td>
</tr>
<tr>
<td>( m_{\ell\ell} )</td>
<td>( \in [76, 106] \text{ GeV} )</td>
</tr>
</tbody>
</table>

Multiple candidates passing the selection requirements, we retain the one with jets in the highest b-tag category. Distributions of the b-tag discriminant are shown in Fig. 5. Further ambiguity between multiple candidates is resolved by selecting the candidate with \( m_{jj} \) and \( m_{\ell\ell} \) values closest to the Z boson mass \( m_Z \).
Table 2: Selection criteria specific to the merged jet events.

<p>| | |</p>
<table>
<thead>
<tr>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>$p_T(\ell^+\ell^-)$</td>
<td>$&gt; 200$ GeV</td>
</tr>
<tr>
<td>$M_{\text{pruned}}(j)$</td>
<td>$&gt; 50$ GeV</td>
</tr>
<tr>
<td>$\Delta R(\ell j)$</td>
<td>$&gt; 0.8$</td>
</tr>
<tr>
<td>$\tau_{21}$</td>
<td>$&lt; 0.5$</td>
</tr>
</tbody>
</table>

Table 3: Selection for the VBF category. The negation of these requirements are used for the dijet and merged categories to obtain exclusive categories.

<p>| | |</p>
<table>
<thead>
<tr>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>$p_T(\text{additional jets})$</td>
<td>$&gt; 30$ GeV</td>
</tr>
<tr>
<td>$</td>
<td>\eta</td>
</tr>
<tr>
<td>$M(\text{additional jet pair})$</td>
<td>$&gt; 500$ GeV</td>
</tr>
<tr>
<td>$\Delta \eta(\text{additional jet pair})$</td>
<td>$&gt; 3.5$</td>
</tr>
</tbody>
</table>

4.1 $t\bar{t}$ background

In the VBF category, the $t\bar{t}$ plays a minor role and is estimated from simulation. The $t\bar{t}$ background is an important source of contamination in the dijet category, especially in the 2 b-tag sub-category. In the dijet and merged jet categories it is estimated from the data using $e^{\pm}\mu^{\mp}$ events passing the same selection as the signal for the two categories that use b-tag subcategories (i.e. dijet and merged jet). This method accounts for other small backgrounds (as WW + jets, $Z \rightarrow \tau^+\tau^- +$ jets, single top, fakes) where the lepton flavor symmetry can be invoked as well. The kinematics of the $e^{\pm}\mu^{\mp}$ control sample has been checked to be identical to the $t\bar{t}$ contribution in the signal region. The background shapes obtained this way are validated against a top-enriched data sample. The agreement between the $e^{\pm}e^{\mp} +\mu^{\pm}\mu^{\mp}$ distributions is within statistical uncertainties.

4.2 $Z+$jets background

The main background is real $Z$-bosons produced in association with jets. We use large exclusive simulated samples of $Z + n$ jets ($n = 1$ to $4$) to study the shape of the corresponding $M_{\ell\ell jj}$ distribution. Small discrepancies are found in the $p_T$ spectrum of the $\ell^+\ell^-jj$ system between data and simulations, the mis-modeling of the $p_T$ spectrum of the $\ell^+\ell^-jj$ system in the simulation is the same in the signal and sideband regions. Thus, a weight calculated as the ratio of the $p_T$ distributions in data over simulations in the sideband region is used to correct the $Z+$jets distributions in the signal region on an event-by-event basis. The weights are corrected for contamination of the sideband data by diboson events (obtained from simulation) and $t\bar{t}$ events (from data, as described above). The agreement between the $e^+e^- +\mu^+\mu^-$ and $e^{\pm}\mu^{\mp}$ distributions is within statistical uncertainties.

The normalization of $Z+$jets background in the signal region is constrained to the relative normalization of the data in the sideband, after subtracting the diboson and $t\bar{t}$ backgrounds. This procedure is applied independently to each b-tag category and lepton flavor. A good agreement between the shapes of the background expectations and the data in the sidebands validates the predictions of the $Z+$jets MC used in the signal region. The agreement is preserved both in the left and right sidebands.
Table 4: Selection for the multivariate discriminants in the various categories.

<table>
<thead>
<tr>
<th>Discriminant</th>
<th>dijet</th>
<th>merged jet</th>
<th>VBF</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\lambda(E_T)$ angular likelihood discriminant</td>
<td>$&lt; 10$</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>$\lambda(E_T)$ VBF discriminant</td>
<td>$&gt; 0.5$</td>
<td>$&gt; 0.5$</td>
<td>$&gt; 0.2$</td>
</tr>
</tbody>
</table>

Table 5: Summary of systematic uncertainties on the signal normalization. Most sources give multiplicative errors on the cross-section measurement, except for the expected Higgs boson production cross section, which is relevant for the measurement of the ratio to the SM expectation. Details concerning the cross-section uncertainty can be found in [48].

<table>
<thead>
<tr>
<th>Source</th>
<th>VBF</th>
<th>merged jet</th>
<th>dijet</th>
<th>Comment</th>
</tr>
</thead>
<tbody>
<tr>
<td>Muon trigger &amp; ID</td>
<td>1.8%</td>
<td>1.7%</td>
<td>1.8%</td>
<td>Tag-&amp;-probe study</td>
</tr>
<tr>
<td>Electron trigger &amp; ID</td>
<td>2%</td>
<td>2%</td>
<td>2%</td>
<td>Tag-&amp;-probe study</td>
</tr>
<tr>
<td>Electron energy scale</td>
<td>-</td>
<td>0-1%</td>
<td>0.2%</td>
<td>JES, correlated among categories</td>
</tr>
<tr>
<td>Muon momentum scale</td>
<td>-</td>
<td>0.1%</td>
<td>0.1%</td>
<td></td>
</tr>
<tr>
<td>Jet reconstruction</td>
<td>1-13%</td>
<td>0.2-0.6%</td>
<td>1-4%</td>
<td>Anti-correlated among categories</td>
</tr>
<tr>
<td>$b$-tagging eff. and mistag rate</td>
<td>-</td>
<td>0.5-7.5%</td>
<td>1-6%</td>
<td></td>
</tr>
<tr>
<td>$Z$-tag</td>
<td>0-10%</td>
<td>10%</td>
<td>-</td>
<td>merged jets only</td>
</tr>
<tr>
<td>Production mechanism (PDF)</td>
<td>1-9%</td>
<td>0.1-1%</td>
<td>1.5%</td>
<td>Correlated between categories</td>
</tr>
<tr>
<td>Production mechanism (lineshape)</td>
<td>0-7%</td>
<td>0-7%</td>
<td>0-7%</td>
<td>Only for $M_H &gt; 400$ GeV</td>
</tr>
<tr>
<td>MC statistics</td>
<td>3-6%</td>
<td>-</td>
<td>-</td>
<td>Limited by ggH + 2 jets</td>
</tr>
<tr>
<td>Luminosity</td>
<td>2.6%</td>
<td>2.6%</td>
<td>2.6%</td>
<td>Same for all analyses</td>
</tr>
<tr>
<td>Higgs cross-section (for $R$)</td>
<td>13-18%</td>
<td></td>
<td></td>
<td>For details see Ref. [48]</td>
</tr>
</tbody>
</table>

5 Results

Taking into account systematic uncertainties, the $m_{ZZ}$ distributions of the selected events split into the aforementioned 14 categories are examined for two signal hypothesis: a standard model like Higgs boson and a mixture between the observed 125 GeV Higgs boson and a yet undiscovered scalar particle.

5.1 Systematic uncertainties

The systematic uncertainties influence both shape and normalization of the background and signal. They are summarized in Tables 5 and 6. We consider effects from leptons (energy scale, resolution, selection, and trigger), hadrons (jet energy scale, pile-up, $E_T$ requirements, heavy-quark flavor tagging), theory (Higgs boson production mechanism, cross section, and branching fractions), and LHC luminosity [47]. In the VBF category we additionally consider theoretical uncertainties related to the production of Higgs bosons in gluon fusion in combination with two jets.

The theoretical uncertainty on the shape of the resonance due to the higher-order terms neglected in the line-shape calculation as well as the missing interference between background and signal are included, as are the uncertainties due to electroweak corrections [22, 24].
Table 6: Summary of systematic uncertainties on the normalization of the background determination. Additional shape uncertainties are associated to the Jet energy scale, $p_T^{\ell\ell jj}$ weighting and residual differences between data and simulation in the sideband region.

<table>
<thead>
<tr>
<th>Source</th>
<th>VBF</th>
<th>merged jet</th>
<th>dijet</th>
</tr>
</thead>
<tbody>
<tr>
<td>Muon trigger &amp; ID</td>
<td>1.8%</td>
<td>1.7%</td>
<td>1.8%</td>
</tr>
<tr>
<td>Muon momentum scale</td>
<td>-</td>
<td>0.2%</td>
<td>0.1%</td>
</tr>
<tr>
<td>Electron trigger &amp; ID</td>
<td>2.0%</td>
<td>2.0%</td>
<td>2.0%</td>
</tr>
<tr>
<td>Electron energy scale</td>
<td>-</td>
<td>0.8%</td>
<td>0.1%</td>
</tr>
<tr>
<td>Jet energy scale</td>
<td>12-19%</td>
<td>0.9%</td>
<td>3.6%</td>
</tr>
<tr>
<td>$b$-tagging efficiency SF 0-tag</td>
<td>-</td>
<td>+0.3%</td>
<td>+0.4%</td>
</tr>
<tr>
<td>$b$-tagging efficiency SF 1-tag</td>
<td>-</td>
<td>-0.7%</td>
<td>-0.8%</td>
</tr>
<tr>
<td>$b$-tagging efficiency SF 2-tag</td>
<td>-</td>
<td>-4.6%</td>
<td>-4.6%</td>
</tr>
<tr>
<td>Mistag SF 0-tag</td>
<td>-</td>
<td>-1.4%</td>
<td>-1.1%</td>
</tr>
<tr>
<td>Mistag SF 1-tag</td>
<td>-</td>
<td>+5.0%</td>
<td>+3.7%</td>
</tr>
<tr>
<td>Mistag SF 2-tag</td>
<td>-</td>
<td>+6.7%</td>
<td>+3.6%</td>
</tr>
<tr>
<td>$E_T$ significance</td>
<td>-</td>
<td>-</td>
<td>0.3%</td>
</tr>
<tr>
<td>Pile-up</td>
<td>0-1%</td>
<td>0.7%</td>
<td>0.1%</td>
</tr>
<tr>
<td>Diboson cross section</td>
<td>15%</td>
<td>12%</td>
<td>12%</td>
</tr>
<tr>
<td>tt cross section</td>
<td>15%</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>Luminosity</td>
<td>-</td>
<td>2.6%</td>
<td>-</td>
</tr>
</tbody>
</table>

Shape variations are introduced for uncertainties that have an impact on the $M_{\ell\ell jj}$ distribution. The largest contribution comes from the jet energy scale which introduce a variation on $M_{\ell\ell jj}$ that is negligible at low mass but grows up to 4% at 600 GeV. To account for additional effects that are not explicitly evaluated, we compare data and simulation in the sideband region and treat the residual difference as another uncertainty.

5.2 Standard-model-like interpretation

The $m_{ZZ}$ spectra are studied for 16 Higgs boson mass hypotheses in the range between 230 and 1000 GeV, where the mass steps are optimized to account for the expected width and resolution for measurement of $m_H$ [49].

Based on the normalization and shape of the $m_{ZZ}$ distribution for signal and background (Fig. 6 to Fig. 8), we extract exclusion limits using the formalism developed by the CMS and ATLAS Collaborations in the context of the LHC Higgs Combination Group [50]. The 14 data categories are used simultaneously in a maximum likelihood fit of the binned histograms for each mass hypothesis with a background-only and a signal-plus-background models, and the likelihood ratio of the two fits serves as test statistic for the modified frequentist criterion $\text{CL}_s$ [51, 52]. Systematic uncertainties are incorporated as nuisance parameters and are treated according to the frequentist paradigm. The results presented here are obtained using asymptotic formulae [53], including a few updates recently introduced in the ROOSTATS package [54]. The resulting post-fit background estimates compared to the observed number of events are shown in Table 7 compared to expected signal yields.

We present exclusion limits on the ratio of the production cross section for the Higgs boson compared to the SM expectation for the exclusive analyses, Fig. 9, and their combination, Fig. 10 (left). Limits on the SM Higgs boson production cross section times branching fraction of $H \rightarrow ZZ$ are presented in Fig. 10 (right). For comparison, expectations are shown for a SM-like Higgs boson. This analysis excludes the existence of a resonance with properties of the SM Higgs boson in the mass range between 305 and 744 GeV. The merged jet category has a
Figure 6: The $m_{ZZ}$ invariant mass distribution for the dijet category after final selection in three categories: 0 b-tag (top), 1 b-tag (middle), and 2 b-tag (bottom). The electron channel is shown on the left and the muon channel is shown on the right. The points show the data and the histograms the background contributions. Also shown is a SM Higgs boson signal ($M_H = 400$ GeV). The gray band indicates the statistical and systematic uncertainty on the normalization and shape of the background. Background normalizations are derived from the limit-setting fit.
5.3 Electroweak singlet interference interpretation

In addition, we interpret the results in terms of an electroweak singlet scalar mixing with the 125 GeV Higgs boson, similar to other analysis in CMS [5]. Phenomenologically the couplings
Table 7: Expected background and observed yields in with 19.7 fb\(^{-1}\) of data. The expected background is composed of \(p_T\)-weighted simulated Z+jets, data-driven top+X and diboson MC, except for the VBF category, where all backgrounds are derived from simulation. The background is in the limit-setting fit using \(m_{jj}\) sideband and takes into account statistical and systematic uncertainties. The uncertainties on the signal expectation are small.

<table>
<thead>
<tr>
<th>(M_H) (GeV)</th>
<th>VBF</th>
<th>merged jet</th>
<th>dijet</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>expected background</td>
<td>observed data</td>
<td>signal expectation</td>
</tr>
<tr>
<td>300</td>
<td>446 ± 16</td>
<td>449</td>
<td>19.9</td>
</tr>
<tr>
<td>600</td>
<td>377 ± 15</td>
<td>384</td>
<td>negligible</td>
</tr>
<tr>
<td>900</td>
<td>132 ± 9</td>
<td>150</td>
<td>24.1</td>
</tr>
</tbody>
</table>

Figure 9: Observed (solid) and expected (dashed) 95% CL upper limits for the (left) dijet, (middle) merged jet, and (right) VBF analyses, obtained with the CL\(_s\) technique, on the ratio of the production cross section to the SM expectation for the Higgs boson. The 68% and 95% ranges of expectation for the background-only model are also shown with green and yellow bands, respectively. The solid horizontal lines at 1 indicate the expectation for a SM-Higgs-like boson.

of the two gauge eigenstates (SM and singlet) become inter-related by unitarity and the original coupling strength of the light Higgs boson is therefore reduced with respect to the SM case. If we define \(C\) (\(C'\)) as the scale factor of the couplings of the low (high) mass with respect to the SM, one can write \(C^2 + C'^2 = 1\) as the unitarity condition to be preserved. The EWK singlet cross-section is also modified by a factor, \(\mu'\), and the modified width is \(\Gamma'\); they are defined as:

\[
\mu' = C'^2 \cdot (1 - BR_{\text{new}}) \tag{1}
\]

\[
\Gamma' = \Gamma_{\text{SM}} \cdot \frac{C^2}{1 - BR_{\text{new}}} \tag{2}
\]

where \(BR_{\text{new}}\) is the branching fraction of the electroweak singlet to non-SM-like decay modes. Indirectly we can set an upper limit at 95% CL on \(C^2 < 0.446\) using the signal strength fits to the H(125) candidate as obtained in [55].

The analysis is performed similarly to the exclusion limits for a SM-like Higgs boson, but weighting the generated signal shapes to a relativistic Breit-Wigner function of width \(\Gamma'\) in
the VBF channel, and to the corresponding CPS shape (including interference effects) for the dijet and merged jet channel. Deviations of the true signal shape of the new resonance are expected to be small compared the experimental resolution. To study the case where \( BR_{\text{new}} \neq 0 \) we limit ourselves to the parameter space, where \( C' \leq \sqrt{1 - BR_{\text{new}}} \), for a given value of \( BR_{\text{new}} \), to remain in the region where \( \Gamma' \leq \Gamma_{\text{SM}} \). This allows us to use the method described above also for \( BR_{\text{new}} \neq 0 \). Under this restriction, our signal simulation can still be weighted to correspond to a given combination of \( C' \) and \( BR_{\text{new}} \).

Expected and observed exclusion limits as a function of the mass for different \( C'^2 \) hypotheses (for \( BR_{\text{new}}=0 \)) are presented in Fig. 11, combining the gluon fusion and VBF analyses. Additionally, exclusion limits for different \( BR_{\text{new}} \) and \( C'^2=0.5 \) are shown in Fig. 12.

Expected and observed exclusion limits in the \( C'^2 - m_H \) plane, for \( BR_{\text{new}}=0 \), are displayed in Fig. 13 for the ggH and VBF analyses, and their combination. The excluded regions for the latter are presented in Fig. 14. Similarly, expected and observed exclusion limits in the \( BR_{\text{new}} - C'^2 \) plane are depicted in Fig. 15, for the most sensitive mass hypothesis (\( M_H = 400 \text{ GeV} \)) and for a high mass hypothesis (\( M_H = 900 \text{ GeV} \)). CMS data disfavor EW-singlet models with high \( C'^2 \) and low \( BR_{\text{new}} \) values, excluding some regions in the parameter space for \( M_H \) in the range 300 to 500 GeV.

6 Summary

A search for a standard model like Higgs boson decaying into two Z bosons which subsequently decay to two leptons and two quarks, \( H \rightarrow ZZ \rightarrow \ell^+\ell^-q\bar{q} \), has been presented. Data corresponding to an integrated luminosity of 19.7 fb\(^{-1} \) of proton-proton collisions at center-of-mass energy of 8 TeV have been analyzed by the CMS Collaboration at the LHC. No evidence
Figure 11: Expected (left) and observed (right) 95% CL upper limits on the ratio of the production cross section to the EW-singlet model expectations as a function of the mass for the combination of the gluon fusion and VBF analyses under the scenario of $BR_{\text{new}} = 0$. Each line represent a different $C'^2$ hypotheses.

Figure 12: Expected (left) and observed (right) 95% CL upper limits on the ratio of the production cross section to the EW-singlet model expectations as a function of the mass for the combination of the gluon fusion and VBF analyses. Each line represent a different $BR_{\text{new}}$ with a fixed value of the modified strength of the signal ($C'^2 = 0.5$).
Figure 13: Contours of the observed limits at the 95% CL in the $C'^2 - m_H$ plane ($BR_{new} = 0$) for the gluon fusion analyses combined (upper left), for the VBF analysis (upper right), and their combination (lower). The solid (dashed) lines depict the observed (expected) exclusion limits, corresponding to $\sigma/\sigma_{th} = 1$. No region is excluded in the $C'^2 - m_H$ parameter space for the VBF analysis alone.
Figure 14: Expected (dashed black) and observed (solid red) lower limits at 95% CL set on $C'^2$ parameter (under $BR_{\text{new}} = 0$ hypothesis) as a function of the mass, for the combination of the gluon fusion and VBF analyses.
Figure 15: Contours of the observed limits at the 95% CL in the $\text{BR}_{\text{new}} - C^2$ plane, for $M_H = 400\text{ GeV}$ (left), and $M_H = 900\text{ GeV}$ (right), for the gluon fusion (upper), VBF (middle) and combined analyses (lower). The solid (dashed) lines depict the observed (expected) exclusion limits, corresponding to $\sigma/\sigma_{\text{th}} = 1$. The white areas in the graphs correspond to model parameters not probed in these analyses.
for a SM-like Higgs boson has been found and upper limits on the production cross section for the SM Higgs boson have been set in the range of masses between 230 and 1000 GeV. We have excluded the existence of a resonance with properties of the SM Higgs boson in the mass range between 305 and 744 GeV. Additionally, we search for an electroweak singlet which becomes visible through interference with the recently discovered 125 GeV boson with similarly negative results. The most stringent limit on the mixing parameter is $C' < 0.41$ for a new resonance of mass 500 GeV without additional invisible decays to new particles.

Acknowledgements

We congratulate our colleagues in the CERN accelerator departments for the excellent performance of the LHC machine. We thank the technical and administrative staff at CERN and other CMS institutes, and acknowledge support from: FMSR (Austria); FNRS and FWO (Belgium); CNFq, CAPES, FAPERJ, and FAPESP (Brazil); MES (Bulgaria); CERN; CAS, MoST, and NSFC (China); COCIENTIAS (Colombia); MSES (Croatia); RPF (Cyprus); MoER, SF0690030s09 and ERDF (Estonia); Academy of Finland, MEC, and HIP (Finland); CEA and CNRS/IN2P3 (France); BMHE, DFG, and HGF (Germany); GSRT (Greece); OTKA and NKTH (Hungary); DAE and DST (India); IPM (Iran); SFI (Ireland); INFN (Italy); NRF and WCU (Korea); LAS (Lithuania); CINEVESTAV, CONACYT, SEP, and UASLP-FAI (Mexico); MSI (New Zealand); PAEC (Pakistan); MSHE and NSC (Poland); FCT (Portugal); JINR (Armenia, Belarus, Georgia, Ukraine, Uzbekistan); MON, RosAtom, RAS and RFBR (Russia); MSTD (Serbia); IN2P3 (Spain); Swiss Funding Agencies (Switzerland); NSC (Taipei); TUBITAK and TAEK (Turkey); STFC (United Kingdom); DOE and NSF (USA).

Individuals have received support from the Marie-Curie programme and the European Research Council (European Union); the Leventis Foundation; the A. P. Sloan Foundation; the Alexander von Humboldt Foundation; the Belgian Federal Science Policy Office; the Fonds pour la Formation à la Recherche dans l’Industrie et dans l’Agriculture (FRIA-Belgium); the Agentschap voor Innovatie door Wetenschap en Technologie (IWT-Belgium); the Council of Science and Industrial Research, India; the Compagnia di San Paolo (Torino); and the HOMING PLUS programme of Foundation for Polish Science, cofinanced from European Union, Regional Development Fund.

References


