Search for Higgs decaying to WW at CMS

Leonardo Di Matteo[1] *on behalf of the CMS Collaboration*

[1] University and INFN of Milano-Bicocca

E-mail: `leonardo.di.matteo@cern.ch`

Abstract. A search for a standard model Higgs boson decaying into a W^+W^- pair is presented, using CMS proton-proton collision data samples both at 7 TeV and 8 TeV center-of-mass energy, corresponding to an integrated luminosity of 4.9 ± 0.2 fb^{-1} and 12.1 ± 0.5 fb^{-1} respectively. Both fully-leptonic and semi-leptonic W decay final states are considered. Upper limits are set on the Higgs boson production relative to the standard model expectation. The standard model Higgs boson is excluded at 95% confidence level in the mass ranges $[128 - 600]$ GeV and $[215 - 490] + [525 - 600]$ GeV using the fully-leptonic and semi-leptonic channels respectively. In the fully-leptonic final state, an excess of events is observed above background which is consistent with the expectations from a standard model Higgs boson of mass 125 GeV and has a statistical significance of 3.1 standard deviations.

1. Introduction

In the context of the Standard Model (SM) of particle physics, the origin of the masses of fundamental particles remains an open issue. The electroweak symmetry breaking mechanism, which causes vector bosons to acquire masses, rely on the existence of the Higgs field and its quantum, the Higgs boson, which is yet to be proven experimentally. One of the main goals of the CERN Large Hadron Collider (LHC) physics program is to investigate on the existence of the SM Higgs boson.

In particular, searches of the Higgs boson are being carried on in the two multipurpose experiments located on the LHC ring, ATLAS and CMS. In these proceedings, I report about two particular searches performed at CMS, both exploiting the Higgs decay into two W bosons: the first exploits the fully-leptonic final state, where each W boson decays in a lepton and a neutrino; the second explores the semi-leptonic final state, where one W decays leptonically and the other hadronically.

The fully-leptonic analysis, $H \rightarrow W^+W^- \rightarrow \ell\nu_\ell\,\ell\nu_\ell$ [1], is performed in the $110 - 600$ GeV Higgs mass range, while the semi-leptonic one, $H \rightarrow W^+W^- \rightarrow \ell\nu_\ell\,q\bar{q}$ [2], is restricted to the $170 - 600$ GeV range, due to its dependency on single lepton high p_T triggers. Both the analyses exploit the full 2011-2012 LHC pp collision dataset, corresponding to an integrated luminosity of 12.1 ± 0.5 fb^{-1}, at $\sqrt{s} = 8$ TeV, and 4.9 ± 0.2 fb^{-1}, at $\sqrt{s} = 7$ TeV.

2. Event reconstruction and physics objects

The CMS detector [3] is used to reconstruct proton-proton collisions. A full event reconstruction aimed at characterizing and identifying all stable particles in the event is obtained via a particle-flow (PF) technique [4]. This approach combines the information from all CMS sub-detectors to identify and reconstruct individual particles, classifying them into mutually exclusive categories:

charged hadrons, neutral hadrons, photons, electrons, and muons.

The electron reconstruction algorithm links clusters of energy deposits in the ECAL to trajectories in the inner tracker. To reconstruct the trajectories in the tracker, a dedicated model of electron energy loss is developed, and series of hits are fitted with a Gaussian sum filter. The electron candidates are filtered with an identification algorithm, relying on a multivariate technique that combines observables sensitive to the amount of bremsstrahlung along the electron trajectory, the geometrical and momentum matching between the electron trajectory and the associated clusters, and shower-shape observables.

The muon reconstruction algorithm exploits both the silicon tracker and the muon system. Segments in the muon spectrometer are matched with energy deposits in the calorimeters and tracks in the silicon detector, and muon candidates are built and filtered with quality criteria.

PF candidates are clustered together into Jets, using the anti-k_T algorithm [5] with distance parameter $R = 0.5$. Then, to account for the non-linear response of the calorimeters to the particle energies, and other instrumental effects, jet energy corrections are applied. The corrections are based on in-situ calibration using dijet and $\gamma + $ jet data samples. The pileup energy, computed as the median energy density in each event, is subtracted from each jet. Jets are additionally required to originate from the primary vertex, which is identified as the vertex with the highest summed p_T^2 of its associated tracks. Out of this collection, b-jets are tagged by looking at the jet displacement with respect to the primary vertex in the transverse direction. In addition, a multivariate selection is applied to separate jets from the primary interaction from those reconstructed due to energy deposits associated with pile-up, based on the differences in the jet shapes, in the relative multiplicity of charged and neutral components and in the different fraction of transverse momentum which is carried by the hardest components.

At the end of the reconstruction sequence, the missing transverse energy E_T^{miss} is found as the modulus of the negative vector sum of the transverse momenta of all reconstructed particles in an event.

Since charged leptons from W and Z boson decays are typically expected to be isolated from other activity in the event, lepton isolation is ensured by computing the sum of the transverse energies of all reconstructed particles, charged or neutral, within a cone of $\Delta R = \sqrt{(\Delta\eta)^2 + (\Delta\phi)^2} < 0.4$ around the lepton direction. In the computation, pileup is taken into account by subtracting the median energy density estimated on an event-by-event basis.

At trigger level, depending on the decay mode, events have to contain a pair of electrons or muons, one lepton with $p_T > 17$ GeV and the other with $p_T > 8$ GeV, or a single electron (muon) with $p_T > 27(24)$ GeV.

Efficiencies for trigger selection, reconstruction, identification, and isolation of leptons are measured from data, using the tag-and-probe technique [6], based on a pure sample of Z events. These measurements are performed in bins of p_T^ℓ and $|\eta^\ell|$. The overall efficiency for selecting electrons in the ECAL barrel (endcaps) vary from around 82% (73%) at $p_T^e \simeq 10$ GeV to 90% (89%) for $p_T^e \simeq 20$ GeV. It drops to about 85% in the ECAL barrel-endcap transition region, at $1.44 < |\eta^e| < 1.57$. Muon offline efficiencies are greater than $\sim 98\%$ in the whole $|\eta^\mu| < 2.4$ range. The muon trigger efficiency for events selected for this analysis ranges from 96% to 99%.

3. The H $\to$ W$^+$W$^- \to \ell\nu_\ell\,\ell\nu_\ell$ search

The search strategy for the H $\to$ W$^+$W$^- \to \ell\nu_\ell\,\ell\nu_\ell$ analysis is based on a signature with two isolated, oppositely charged, high p_T leptons (electrons or muons) and large missing transverse momentum E_T^{miss}, due to the undetected neutrinos.

To improve the signal sensitivity, events are classified by jet multiplicity into three mutually exclusive categories, characterized by different signal yields and signal-to-background ratios. In

the following, these will be referred to as the 0-jet, 1-jet and 2-jet samples, where the selected jets are required to have $E_{\mathrm{T}} > 30$ GeV and $|\eta| < 4.7$. Events with more than three jets are discarded, while events with exactly three jets are incorporated in the 2-jet category if the lowest-p_{T} jet is not between the two other ones in pseudo-rapidity. Additionally, the signal candidates are splits into three final states: e^+e^-, $\mu^+\mu^-$, $e^{\pm}\mu^{\mp}$.

Tag leptons are required to have opposite charge, p_{T} greater than 20 GeV for the leading lepton ($p_{\mathrm{T}}^{\ell,\mathrm{max}}$) and p_{T} greater than 10 GeV for the trailing lepton ($p_{\mathrm{T}}^{\ell,\mathrm{min}}$), and to originate from the primary vertex of the event (chosen as the vertex with the highest $\sum p_{\mathrm{T}}^2$). Only electrons (muons) with $|\eta| < 2.5(2.4)$ are considered in the analysis.

In addition to high momentum isolated leptons and low jet activity, missing transverse momentum is present in signal events but generally not in background. In this analysis, a *projected* $E_{\mathrm{T}}^{\mathrm{miss}}$ variable is devised. It is defined as the component of $E_{\mathrm{T}}^{\mathrm{miss}}$ transverse to the nearest lepton if the difference in azimuth (ϕ) between this lepton and the $E_{\mathrm{T}}^{\mathrm{miss}}$ vector is less than $\pi/2$. If no leptons closer than $\pi/2$ with respect to the direction of $E_{\mathrm{T}}^{\mathrm{miss}}$ in ϕ are present, then the *projected* $E_{\mathrm{T}}^{\mathrm{miss}}$ is equal to $E_{\mathrm{T}}^{\mathrm{miss}}$. Since the *projected* $E_{\mathrm{T}}^{\mathrm{miss}}$ resolution is degraded by pile-up, the minimum of two $E_{\mathrm{T}}^{\mathrm{miss}}$ observables is used: the first includes all reconstructed particles in the event, while the second uses only the charged particles associated with the primary vertex. Events with *projected* $E_{\mathrm{T}}^{\mathrm{miss}}$ above 20 GeV are selected for the analysis.

Top-quark background is suppressed by means of a top tagging technique based on soft-muon and b-jet tagging. The first method is designed to veto events containing soft-muons produced by b-quarks resulting from top-quark decays. The second one exploits the b-jet tagging algorithm, which searches for jet tracks with large impact parameter.

W+jets background is reduced by requiring a minimum dilepton transverse momentum ($p_{\mathrm{T}}^{\ell\ell}$) of 45 GeV.

To reduce the background from WZ production, any event with a third lepton passing the identification and isolation requirements is rejected.

The contamination from Wγ production, where the photon is misidentified as an electron, is reduced by about 90% in the e^+e^- state by γ conversion rejection requirements.

The background from low mass resonances is rejected by requiring a dilepton mass ($m_{\ell\ell}$) greater than 12 GeV.

The huge Drell-Yan process, which greatly affects same-flavour final states, is dealt with a more complex strategy. First of all, the resonant component of the Drell-Yan production is rejected by requiring a dilepton mass outside a 30 GeV window centered on the Z mass pole. Then, the remaining off-peak contribution is suppressed by exploiting various $E_{\mathrm{T}}^{\mathrm{miss}}$-based approaches depending on the number of jets. In the 0/1-jet categories, Drell-Yan events are more difficult to separate from the signal, therefore, in this case, a dedicated multivariate selection, combining missing transverse momentum with kinematic and topological variables, is used to reject background events and maximize the surviving signal yield. In the 2-jet category, the dominant source of fake $E_{\mathrm{T}}^{\mathrm{miss}}$ is the mis-measurement of the hadronic recoil, therefore the optimal performance is obtained by requiring $E_{\mathrm{T}}^{\mathrm{miss}}$ greater than 45 GeV. Additionally, for the 2-jet category, the momenta of the dilepton system and of the most energetic jet must not be back-to-back in the transverse plane. The combination of these selections reduce the Drell-Yan background by three orders of magnitude, while rejecting less than 50% of signal events.

Furthermore, to enhance the sensitivity to a Higgs boson signal, a cut-based approach is devised for the final Higgs selection in all categories. Due to different signal kinematics for different Higgs mass hypotheses, separate optimizations are performed for different values of m_{H} in a cut-based analysis. In addition, a two-dimensional shape analysis technique is also pursued for

the $e^{\pm}\mu^{\mp}$ final state in the 0/1-jet categories, since it allows for a direct physical interpretation of the observed data with a sensitivity comparable to other more complex techniques.

The optimized cut-based approach consists in extra requirements placed on $p_T^{\ell,\max}$, $p_T^{\ell,\min}$, $m_{\ell\ell}$, $\Delta\phi_{\ell\ell}$ and the transverse mass m_T, defined as $\sqrt{2p_T^{\ell\ell}E_T^{\mathrm{miss}}(1-\cos\Delta\phi_{E_T^{\mathrm{miss}}\ell\ell})}$, where $\Delta\phi_{E_T^{\mathrm{miss}}\ell\ell}$ is the azimuthal difference between E_T^{miss} and the dilepton system.

Also for the 2-d analysis, which exploits the independent variables m_T and $m_{\ell\ell}$, an additional loose set of requirements is applied: m_T must be greater than 80 GeV and smaller than 280 (600) GeV for m_H hypotheses smaller or equal than 250 GeV (greater than 250 GeV), while $m_{\ell\ell}$ must be smaller than 200 (600) GeV for m_H hypotheses smaller or equal than 250 GeV (greater than 250 GeV). Finally, $p_T^{\ell,\max}$ is required to be larger than 50 GeV for m_H hypotheses greater than 250 GeV. The two-dimensional distributions for the $m_H = 125$ GeV and $m_H = 200$ GeV Higgs signal hypotheses, the background processes and the data, are shown in Figure 1 for the 0-jet sample. This specific binning size is chosen to avoid empty bins for the overall background contribution for the present Monte Carlo statistics, while still keeping enough cells to differentiate the signal shape from the background shape. All the 80 bins enter in a binned likelihood fit of the data to the signal and background hypotheses in this two-dimensional shape analysis.

The 2-jet category is mainly sensitive to the vector boson fusion (VBF) production mode, whose cross section is roughly ten times smaller than that of the gluon-gluon fusion mode, but it is characterized by a relatively low background environment and it probes a different production mechanism to test the compatibility of a signal with the SM Higgs boson hypothesis. The $H \to W^+W^-$ events from VBF production are characterized by a pair of energetic forward-backward jets and very little hadronic activity in the rest of the event. Therefore, 2-jets events are further required to have zero hard jets and both the tag leptons present in the pseudorapidity region between the two leading jets. Additionally, to reject the main background from top-quark decays, two additional requirements are applied to the two jets, j_1 and j_2: $\Delta\eta(j_1,j_2) > 3.5$ and $m_{j_1j_2} > 500$ GeV. Finally, the cut-based approach in the 2-jet category makes use of the same m_H dependent requirements, with the exception that m_T is required to be larger than 30 GeV.

3.1. Background determination

To evaluate the background contamination in the signal region after the Higgs selection, a combination of techniques is used. When possible, data are used to estimate directly background contributions, thus avoiding large uncertainties related to the limited Monte Carlo simulation. The remaining contributions taken from simulation are small.

The W + jets and QCD multi-jet backgrounds are due to leptonic decays of heavy quarks, hadrons misidentified as leptons, and electrons from photon conversion. Hence, their estimation is derived directly from data using a control sample of events in which one lepton passes the standard analysis criteria and the other does not, but instead satisfies a relaxed set of requirements, resulting in a tight-fail sample. Firstly, the efficiency for a jet selected by the loose selection to pass the tight selection is measured using data from an independent multi-jet event sample dominated by non-prompt leptons. Secondly, this efficiency is used to extrapolate the tight-fail counts into the signal region (tight-tight sample). Systematic uncertainties related to the efficiency measurement dominate the overall uncertainty of this method, which is estimated to be about 36%.

Top-quark background is also estimated from data, by counting the number of top-tagged events and by extrapolating into the signal region using the corresponding top-tagging efficiency, which is measured on data with one b-tagged jet. This method is limited by the statistical uncertainty of the control sample and from the systematic uncertainties related to
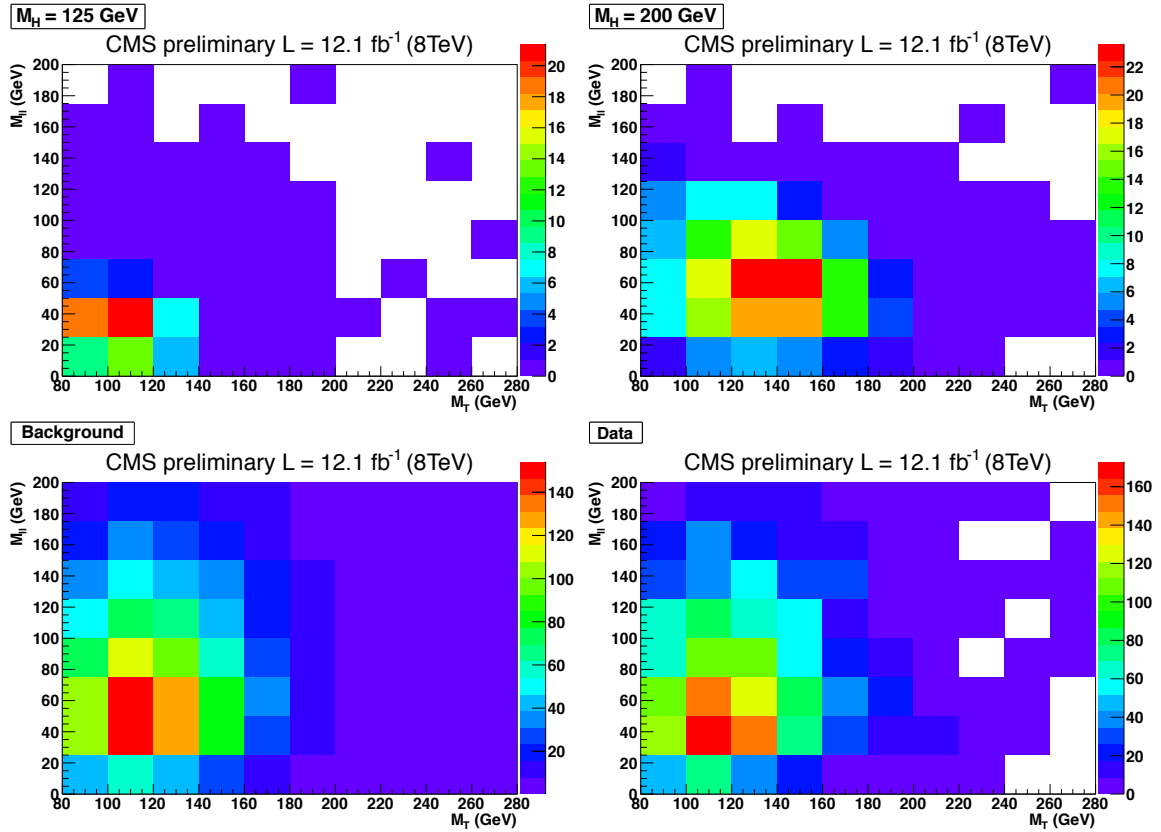
Figure 1. $H \to W^+W^- \to \ell\nu_\ell\,\ell\nu_\ell$ search: $(m_T, m_{\ell\ell})$ distributions in the 0-jet bin for the $m_H = 125$ GeV Higgs signal hypothesis (top left), the $m_H = 200$ GeV Higgs signal hypothesis (top right), the background processes (bottom left), and the data (bottom right).

the measurement of the tagging efficiency (about 20% in the 0-jet category and about 4% in the 1-jet category).

For the low-mass signal region, $m_H \leq 200$ GeV, the non-resonant W^+W^- contribution is estimated from data. Events with a dilepton mass larger than 100 GeV, where the Higgs boson signal contamination is negligible, are used to measure this contribution and the MC simulation is used to extrapolate into the signal region. The total uncertainty is about 10%. For larger Higgs boson masses the overlap between the non-resonant W^+W^- and the Higgs boson signal becomes significant, and the sole simulation is used for the background estimation.

The Drell-Yan contribution to the same flavour final states is based on an extrapolation from the observed number of events with a dilepton mass within ± 7.5 GeV of the Z mass. The residual background in the control region is subtracted using $e^\pm\mu^\mp$ events. This method is limited by the statistical uncertainty of the control sample, which is about 20% to 50%. The $Z/\gamma^* \to \tau\tau$ contamination is estimated using $Z/\gamma^* \to ee$ and $Z/\gamma^* \to \mu\mu$ events selected in data, where the leptons are replaced with simulated τ decays, thus providing a better description of the experimental conditions with respect to the simulation.

The $W\gamma^*$ contamination, from asymmetric virtual photon decays where one lepton escapes detection, is determined using the MADGRAPH generator with dedicated cuts. The normalization scale of the simulated events is estimated through a control sample of high purity $W\gamma^*$ events with three reconstructed leptons. A measured factor of 1.6 ± 0.5 with respect to the

leading order cross-section is found.

The remaining minor backgrounds, from WZ, ZZ (when the two selected leptons come from different bosons) and Wγ, are estimated from simulation.

For the two-dimensional analysis, the simulation is employed to build the templates, which are then cross-checked in data control samples. For the W + jets background the nominal shape is derived from the same control sample used to determine the normalization.

4. The H $\to$ W$^+$W$^- \to \ell\nu_\ell$ q$\bar{\text{q}}$ search

The W$^+$W$^-$ semi-leptonic channel has two main advantages with respect to the fully-leptonic final state: it has a largest branching fraction and a reconstructible Higgs boson mass peak. On the other hand, it is contaminated by a large W + jets background. The precision on the estimation of this background largely determines the sensitivity of the analysis.

The reconstructed electrons (muons) are required to have $p_{\text{T}} > 35(25)$ GeV, and are restricted to $|\eta| < 2.5(2.1)$. The jets are required to have $p_{\text{T}} > 30$ GeV and $|\eta| < 2.4$. Events with the electrons and muons, and with exactly two or three jets are analyzed separately, giving four categories in total. In addition, the electron(muon) categories are required to have $E_{\text{T}}^{\text{miss}} > 30(25)$ GeV.

The two highest-p_{T} jets are assumed to be the hadronic W decay products. According to the simulation, in the case of 2(3)-jet event categories, this criteria correctly identifies the jet pair at a rate which varies from 68(26)% for $m_{\text{H}} = 200$ GeV to 88(84)% for $m_{\text{H}} = 600$ GeV. Events with an incorrect dijet combination result in a broad non-peaking background in the m_{WW} spectrum.

The $(\ell, E_{\text{T}}^{\text{miss}})$ system is used to reconstruct the leptonic W candidate. To reduce the background from processes that do not contain W $\to \ell\nu$ decays, requirements on $m_{\text{T}}(\ell, E_{\text{T}}^{\text{miss}}) > 30$ GeV and $|\Delta\phi_{\text{leadingjet},E_{\text{T}}^{\text{miss}}}| > 0.8\ (0.4)$ are imposed for electrons (muons). These selection criteria reduce the QCD multi-jet background, for which in many cases the $E_{\text{T}}^{\text{miss}}$ is generated by the mis-measurement of the hadronic recoil.

In addition to the listed cuts, a likelihood discriminant is built to enhance the signal sensitivity. The discriminant input variables comprise five angles between the Higgs decay products, fully describing the Higgs production kinematics [7], the p_{T} and rapidity of the W$^+$W$^-$ system, and the lepton charge. The likelihood discriminating power is optimized independently with dedicated simulation samples for several Higgs mass hypotheses, for each lepton flavor and for each jet multiplicity (2-jet, 3-jet): four different optimizations are therefore obtained per mass hypothesis. The figure of merit used for the optimization is the expected limit for the Higgs cross-section. For each of the event categories, events are retained if they survive a simple selection on the likelihood discriminant.

Finally, the m_{WW} resolution is improved by means of a kinematic fit, where both the W candidates are constrained to the W-boson mass to within its known width, with the longitudinal component of the neutrino momentum, $|p_{\text{z}}|$, treated as a free parameter.

To estimate the normalizations of all background components in the signal region, an unbinned maximum likelihood fit is performed on the invariant mass distribution of the dijet system, m_{jj}. The fit is performed independently for each Higgs boson mass hypothesis. The signal region corresponding to the W mass window, 65 GeV $< m_{jj} < 95$ GeV, is excluded from the fit.

The measured invariant mass of the W$^+$W$^-$ system, $\ell\nu_\ell\, jj$, is used as final discriminating variable. The binned distributions of $\ell\nu_\ell\, jj$ for the total background, signal and data for each mass hypothesis and event category are constructed and used as input for the limit-setting

procedure. The $\ell\nu_\ell jj$ shape for the major background, W + jets, is determined from data as an interpolation of the shapes measured in two signal-free sideband regions of m_{jj} (55 GeV $< m_{jj} < 65$ GeV, 95 GeV $< m_{jj} < 115$ GeV). The relative fraction of the two sidebands into the signal region is found through simulation, separately for each Higgs boson mass hypothesis, by minimizing the χ^2 between the interpolated shape in the signal region and the expected one. In addition, to avoid statistical fluctuations due to the low event count, the interpolated distribution is fit with an exponential function. This function is then used for the limit-setting.

The m_{jj} and $m_{\ell\nu jj}$ distributions with final background estimates are shown in Figure 2, with selections optimized for a 500 GeV Higgs boson mass hypothesis, for the muon 2-jets category. All uncertainties arising from interpolation and fit procedures are propagated to the limit calculation as systematic uncertainties.
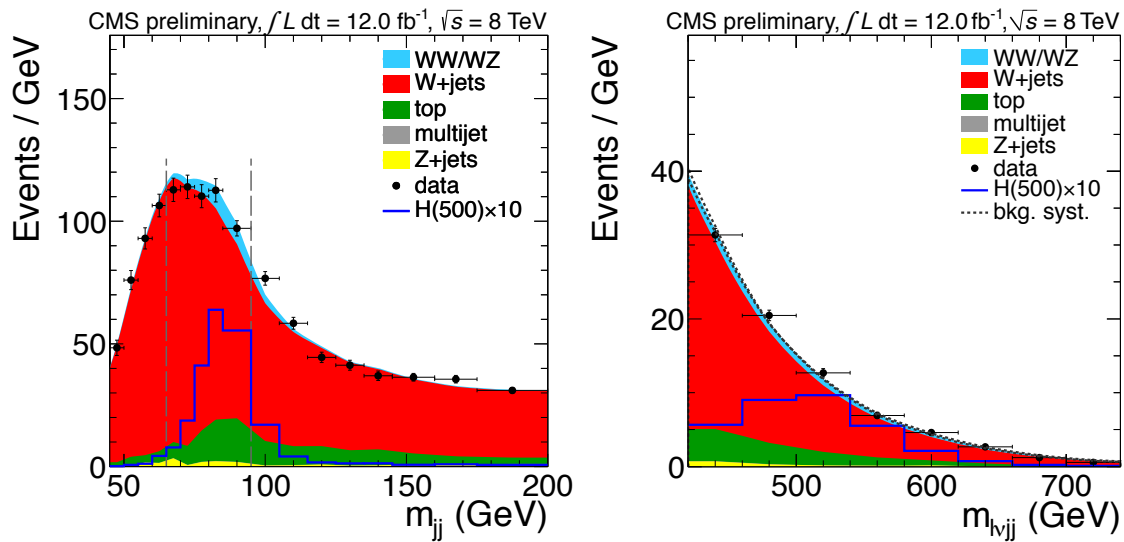
Figure 2. $H \to W^+W^- \to \ell\nu_\ell q\bar{q}$ search: Invariant mass distributions for the $m_H = 500$ GeV mass hypothesis, muon 2-jets category. On the left, the dijet invariant mass distribution with the major background contributions. The vertical lines corresponds to the signal region of this analysis 65 GeV $< m_{jj} < 95$ GeV. On the right, the WW invariant mass distribution with the major background contributions in the signal region.

5. Results

An excess of data with respect to the total background prediction is observed in the $H \to W^+W^- \to \ell\nu_\ell \ell\nu_\ell$ analysis for low Higgs mass selections. The significance of the excess with respect to the background only hypothesis is 3.1 standard deviations, while 4.1 standard deviations are expected assuming the SM prediction.

No other significant deviations are observed, and therefore upper limits on the Higgs boson production relative to the standard model Higgs expectation are derived. To compute the upper limits the modified frequentist construction CL_S [8] is used. The standard model Higgs boson is excluded in the mass range $128 - 600$ GeV at 95% confidence level (CL), while the expected exclusion limit in the hypothesis of background only is $118 - 565$ GeV.

With the $H \to W^+W^- \to \ell\nu_\ell q\bar{q}$ analysis, the presence of a SM Higgs boson in the mass range $215 - 490$ and $525 - 600$ GeV is excluded at 95%, while the expected exclusion limit in

the hypothesis of background only is $170 - 585$ GeV.

Both the results are graphically summarized in Figure 3.
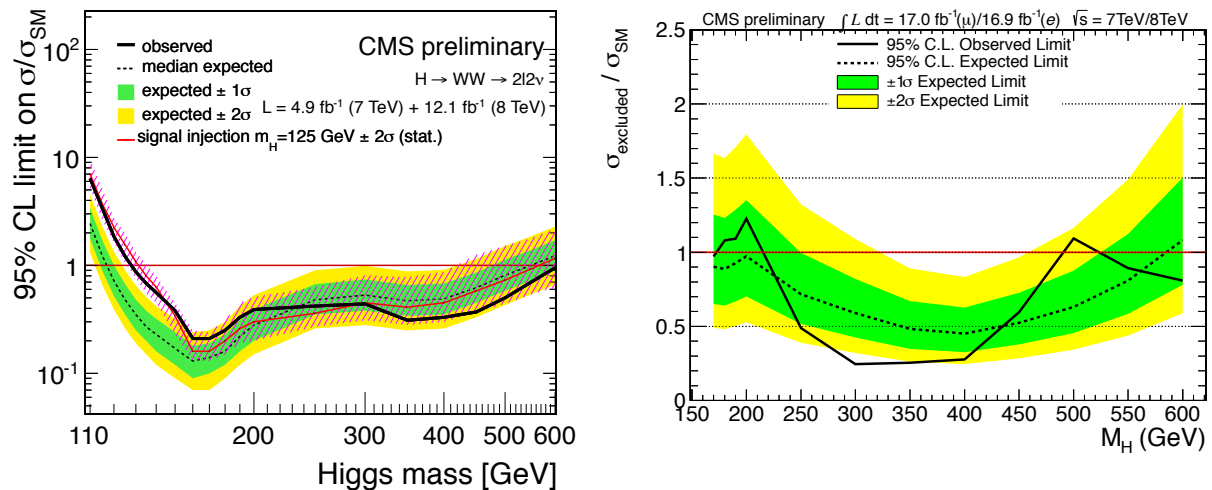
Figure 3. Expected and observed 95% CL upper limits on the cross section times branching fraction, relative to the SM Higgs expectation, for the $H \rightarrow W^+W^- \rightarrow \ell\nu_\ell\,\ell\nu_\ell$ (left) and the $H \rightarrow W^+W^- \rightarrow \ell\nu_\ell\,q\bar{q}$ (right) analyses. For the $H \rightarrow W^+W^- \rightarrow \ell\nu_\ell\,\ell\nu_\ell$ search, the expected limits in the presence of the Higgs with $m_H = 125$ GeV and its associated uncertainty are also shown.

[1] The CMS Collaboration 2012 CMS-PAS-HIG-12-042
[2] The CMS Collaboration 2012 CMS-PAS-HIG-12-046
[3] Chatrchyan S *et al.* (CMS) 2008 *JINST* **3** S08004
[4] The CMS Collaboration 2012 CMS-PAS-PFT-09-001
[5] Cacciari M, Salam G P and Soyez G 2008 *JHEP* **04** 063 (*Preprint* 0802.1189)
[6] Chatrchyan S *et al.* (CMS) 2011 *JHEP* **10** 132
[7] Gao Y *et al.* 2010 *Phys. Rev. D* **81** 075022 (*Preprint* 1001.3396)
[8] Read A L 2002 *Journal of Physics G: Nuclear and Particle Physics* **28** 2693 URL http://stacks.iop.org/0954-3899/28/i=10/a=313