Search for a Standard Model Higgs boson decaying to b quarks and produced in association with Z/W bosons with the CMS detector

Michele De Gruttola[1,2,a]

[1] University of Florida, Institute for High Energy Physics and Astrophysics, Department of Physics, Gainesville, FL USA
[2] Fermi National Accelerator Laboratory, Batavia, IL USA

Abstract. A search for the standard model Higgs boson is performed in a data sample corresponding to an integrated luminosity of 1.1 fb^{-1}, recorded by the CMS detector in proton-proton collisions at the LHC with a 7 TeV center-of-mass energy. The following modes are studied: W($\mu\nu$)H, W(eν)H, Z($\mu\mu$)H, Z(ee)H and Z($\nu\nu$)H, with the Higgs decaying to bb pairs. 95% C.L. upper limits on the VH production cross section are derived for a Higgs mass between 110 and 135 GeV. The expected (observed) upper limit at 115 GeV is found to be 5.7 (8.3) times the standard model expectation.

1 Introduction

The search for the Higgs boson [1] is currently one of the most important undertakings of experimental particle physics.

At the LHC the main Higgs production mechanism is direct production through gluon fusion, with a cross section of $\sim 17 \cdot 10^3$ fb for a Higgs mass $m_H = 120$ GeV [2]. However, in this production mode, the detection of the $H \rightarrow b\bar{b}$ decay is rendered nearly impossible due to overwhelming QCD di-jet production. The same holds true for the next most copious production mode, through vector-boson fusion, with a cross section of $\sim 1,300$ fb. Instead we consider processes in which the Higgs is produced in association with a vector boson which have cross sections of ~ 660 and ~ 360 fb for WH and ZH respectively. Even if the resulting sensitivity of the $H \rightarrow b\bar{b}$ decay is less than other final states such as $H \rightarrow \gamma\gamma$ and $H \rightarrow \tau\tau$ for example, it is paramount to search for the Higgs in these modes given that the observation of the $H \rightarrow b\bar{b}$ decay is key to determine the nature of this particle, if and when discovered.

We summarize a search for the standard model Higgs boson in the $pp \rightarrow VH$ production mode with the CMS detector [4]. The analysis is performed in a data sample corresponding to an integrated luminosity of 1.1 fb^{-1}, collected by the CMS experiment at a 7 TeV center-of-mass energy. The following final states are included: $W(\mu\nu)H$, $W(e\nu)H$, $Z(\mu\mu)H$, $Z(ee)H$ and $Z(\nu\nu)H$ –all with the Higgs decaying to $b\bar{b}$ pairs.

Backgrounds arise from production of W and Z bosons associated with jets (all flavors), singly (ST) and pair-produced top quarks, and di-bosons (VV). Simulated samples of all backgrounds are used to provide guidance in the analysis optimization, and an initial evaluation of their contributions in the search region. For the main backgrounds, high-purity control regions are used to estimate their contribution in the signal region.

An optimization of the event selection, that depends on the Higgs mass, is performed, and 95% C.L. upper limits on the $pp \rightarrow VH$ production cross section are obtained for Higgs masses between 110-135 GeV. These limits are based on the observed event count and background estimate in signal regions defined in either the invariant mass distribution of $H \rightarrow b\bar{b}$ candidates ("M(jj) or cut-and-count analysis"), or in the output discriminant of a boosted decision tree algorithm ("BDT analysis")[5]. The latter enhances the statistical power of the analysis by making full use of correlations between discriminating variables in signal and background events.

For lack of space we will present here only tables and plots for the 115 GeV mass hypothesis, while only the final limits plots will contains all mass range search.

2 Event selection

Candidate $W(\rightarrow \ell\nu)$ decays are identified by requiring the presence of a single, isolated, lepton and additional missing transverse energy(MET). Muons (electrons) are required to have a p_t above 20 (30) GeV. Candidate $Z \rightarrow \ell\ell$ decays are reconstructed by combining isolated, opposite charge pairs of electrons and muons and requiring the dilepton invariant mass to satisfy 75 $<m_{\ell\ell} <$ 105GeV. For Z candidates the electron p_t is lowered to 20 GeV. The identification of $Z \rightarrow \nu\nu$ decays requires MET > 160GeV (the high threshold dictated by the trigger).

The reconstruction of the $H \rightarrow b\bar{b}$ decay is made by requiring the presence of two central ($|\eta| < 2.5$) jets, above a minimum p_t threshold and b-tagged. If more than two such jets are found in the event, the pair with the highest sum of the b-tag outputs for the two jets is chosen (except for the WH analyses, in which the $t\bar{t}$ background is larger, where the pair of jets with highest total p_t is chosen). These combinations are found to yield higher efficiency and rejection of wrong combinations in signal events, as opposed to simply selecting the two highest p_t jets in the event. After b-tagging the fraction of $H \rightarrow b\bar{b}$ candidates that contain the two b-jets from the Higgs decay is near unity.

---

[a] e-mail: michele.de.gruttola@cern.ch

After b-tagging, the background from V+jets and di-bosons is reduced significantly and becomes dominated by the sub-processes where the two jets originate from real b-quarks. Events with additional jets (N_{aj}) or additional leptons (N_{al}) are rejected to further reduce backgrounds from $t\bar{t}$ and WZ.

The topology of VH production is such that the W/Z and the Higgs recoil away from each other with significant p_t. Cuts on the azimuthal opening angle between the vector boson and the reconstructed momenta of the Higgs candidate, $\Delta\phi(V, H)$, on the p_t of the V-boson and on the b-tagged dijet pair achieve significant rejection for most background processes and improve the analysis reach.

For the $Z \rightarrow \nu\nu$ channel, QCD backgrounds are further reduced by a factor of ~ 30 when requiring that the MET does not originate from mismeasured jets.

The training of the BDT is done with simulated samples for signal and background that pass a looser event selection relative to the M(jj) analyses. Several input variables were chosen by iterative optimization. These include the di-jet invariant mass and momentum: M(jj) and p_{tjj}, the V transverse momentum $p_T(V)$, the b-tag value for each of the two jets, the azimuthal angle between the V and the di-jets, $\Delta\phi(V, H)$, and the pseudorapidity separation between the two jets, $\Delta\eta(J1, J2)$. The BDT analysis was expected to improve the sensitivity with respect to the M(jj) analysis by about 10% in every channel.

3 Control regions

Appropriate control regions that are orthogonal to the signal region are identified in data and used to adjust Monte Carlo estimates for the most important background processes: W + jets and Z + jets (with light and heavy-flavor jets), $t\bar{t}$ and QCD multijet and heavy-quark production. Different control regions are found for each of the different search channels by changing the event selection in a way that enriches the content of each specific background. For all cases, control regions with purity ranging from about 20% to nearly 100% have been successfully found. Discrepancies between the expected and observed yields in the data in these control regions are used to obtain a scale factor by which the estimates from the simulation are adjusted. The background from these sources in the signal region are then estimated from the adjusted simulation samples, taking into account the associated systematic uncertainty. The precise construction of all the control regions is involved and outside the scope of this summary. The procedures applied include, for example: reversing the b-tagging requirements to enhance W + jets and Z + jets with light-flavor jets; enforcing a tighter b-tag requirement and requiring extra jets to enhance $t\bar{t}$ and requiring low "boost" in order to enhance $V \rightarrow bb$ over $t\bar{t}$. Table 1 lists the control regions and the corresponding purities and scale factors obtained.

The $Z \rightarrow \nu\nu H$ channel is unique among the five modes analyzed, in that it does not include charged leptons. An important check is to compare the observed pfMET distribution with the predicted distribution from simulation. To accomplish this, muons are removed from the $Z \rightarrow \mu\mu J$ data sample. Reasonably pure samples of $t\bar{t}$ and W + jets events can be obtained by requiring at least one additional isolated lepton in the event, and then either requiring (for

Table 1. Purity and scale factors (Data/MC) derived from background enriched control regions (CR), as described in the text. The scale factors for $W \rightarrow \mu\nu H$ and $W \rightarrow e\nu H$ were averaged together, and the same was done for $Z \rightarrow \mu\mu H$ and $Z \rightarrow ee H$.

| CR | $W \rightarrow \ell\nu H$ | | $Z \rightarrow \ell\ell H$ | |
	Purity	SF	Purity	SF
V + udscg	79.4%	0.84 ± 0.10	92.8%	0.88 ± 0.02
$t\bar{t}$	85.8%	1.01 ± 0.11	97.5%	0.99 ± 0.05
V+$b\bar{b}$	20.2%	1.40 ± 0.29	81.6%	1.16 ± 0.08

Table 2. Z_{inv} Purity and scale factors (SF, Data/MC) derived from background enriched control regions, as described in the text.

CR	Purity	SF
Z + udscg	92.4%	0.97 ± 0.06
W + udscg	94.1%	0.92 ± 0.05
$Z \rightarrow bb$	44.4%	1.00 ± 0.30
$t\bar{t}$	89.9%	0.91 ± 0.09

$t\bar{t}$) or vetoing (for W + jets) b-jets. Table 2 lists the control regions and the corresponding purities and scale factors obtained.

The QCD background in the signal region is also estimated from data using control regions of high and low values of two uncorrelated variables with significant discriminating power towards QCD events. One is the angle between the missing energy vector and the closest jet in azimuth, $\Delta\phi(\text{pfMET}, J)$ and the other is the sum of the CSV values of the two b-tagged jets. The signal region is at high values of both discriminants, while QCD populates regions with low values of either. The method predicts a negligible contamination of this background.

4 Systematics

The following systematic uncertainties on the expected signal and background yields affect the upper limit. The values listed are an approximation of what is actually used in the limit calculation.

The total uncertainty on the signal prediction is taken to be 26% and 28% for ZH and WH production, respectively. Background uncertainties range from 12% to 20% depending on mode and mass point.

Experimental sources of systematics are the b-tag efficiency ($\sim 10\%$), the jet energy resolution ($\sim 10\%$) and scale ($\sim 1\%$) uncertainty, the machine luminosity ($\sim 4.5\%$), the trigger efficiency ($\sim 2\%$). The signal cross section is affected by electroweak corrections for a boost of ~ 150 GeV are 5% for ZH and 10% for WH, and QCD correction, relevant in the comparison NNLO vs. NLO, where an uncertainty of 10% for both ZH and WH is estimated.

5 Results

The final predicted number of events in the signal regions of the BDT and M(jj) analyses are determined with a mix of data-driven estimates based on the control regions, and expectations from simulation. We summarize the final signal

and background estimates in both sets of analyses, including the systematic uncertainties summarized in the previous section, and the expected and observed upper limits using 1.1 fb^{-1} of integrated luminosity. We report in tables 3 and 4 and figures 1 and 2 the results for a single mass point, 115 GeV. While the final limits plots include mass points from 110 to 135 GeV.

Table 3. Predicted backgrounds, signal yields with total uncertainty, and the observed number of events for 115 mass point for the 5 channels M(jj) analysis. We report also the sliding windows on M(jj) for the 115 mass point search.

| Process	W($\to \mu\nu$)H	W($\to e\nu$)H	Z($\to \mu\mu$)H	Z($\to ee$)H	Z($\to \nu\nu$)H
M(jj) cut	100–130	100–130	95–125	95–125	100–130
W + udscg	0.081 ± 0.038	0.01 ± 0.004	-	-	0.023 ± 0.007
W $\to$ bb	0.829 ± 0.221	0.344 ± 0.093	-	-	0.310 ± 0.084
Z + udscg	-	-	0.110 ± 0.065	0.006 ± 0.003	0.180 ± 0.039
Z $\to$ bb	0.184 ± 0.131	0.204 ± 0.146	2.050 ± 0.396	1.545 ± 0.254	1.890 ± 0.578
$t\bar{t}$	1.109 ± 0.287	0.543 ± 0.136	0.090 ± 0.036	0.133 ± 0.073	1.470 ± 0.504
ST	0.24 ± 0.105	0.122 ± 0.049	0.090 ± 0.036	0.009 ± 0.007	0.410 ± 0.156
VV	0.153 ± 0.064	0.065 ± 0.026	0.160 ± 0.064	0.189 ± 0.074	0.460 ± 0.174
B_{exp}	2.596 ± 0.449	1.288 ± 0.242	2.410 ± 0.358	1.883 ± 0.232	4.793 ± 0.938
WH	0.354 ± 0.099	0.296 ± 0.083	-	-	0.091 ± 0.012
ZH	0.006 ± 0.002	0.002 ± 0.001	0.195 ± 0.051	0.193 ± 0.050	0.502 ± 0.104
N_{obs}	4	4	3	2	5

Table 4. Predicted backgrounds, signal yields with total uncertainty, and the observed number of events for 115 mass point for the 5 channels BDT analysis. We report also the BDT cut we choose for the search.

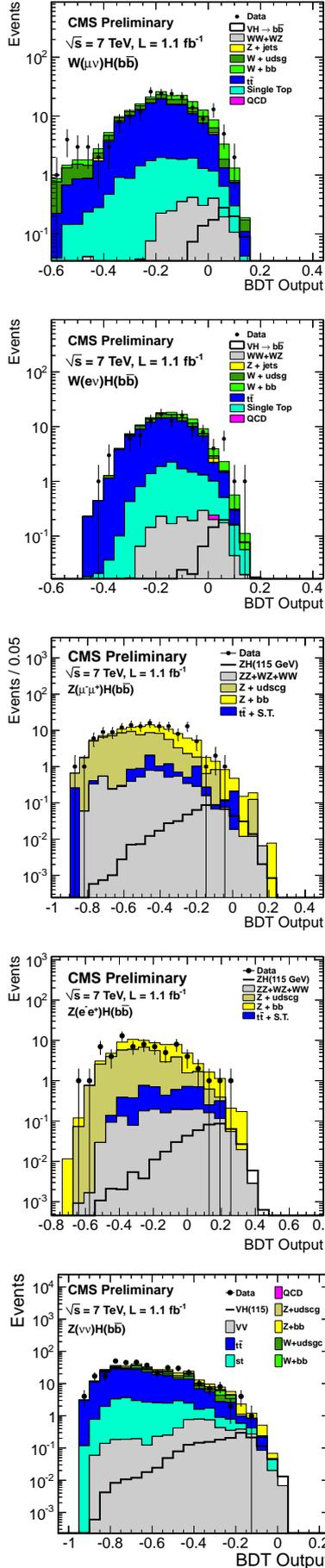| Process	W($\to \mu\nu$)H	W($\to e\nu$)H	Z($\to \mu\mu$)H	Z($\to ee$)H	Z($\to \nu\nu$)H
BDT	> 0.050	> 0.040	> −0.145	> 0.160	> −0.175
W + udscg	0.667 ± 0.192	0.155 ± 0.063	-	-	-
W $\to$ bb	2.035 ± 0.533	1.378 ± 0.374	-	-	0.359 ± 0.097
Z + udscg	-	-	0.110 ± 0.065	0.077 ± 0.036	-
Z $\to$ bb	0.006 ± 0.006	0.198 ± 0.141	2.858 ± 0.645	0.904 ± 0.189	1.108 ± 0.339
$t\bar{t}$	1.173 ± 0.304	1.254 ± 0.315	0.296 ± 0.118	0.179 ± 0.133	0.297 ± 0.077
ST	0.653 ± 0.261	0.653 ± 0.261	0.060 ± 0.014	−	0.453 ± 0.172
VV	0.355 ± 0.140	0.292 ± 0.117	0.334 ± 0.030	0.195 ± 0.078	0.571 ± 0.216
B_{exp}	4.889 ± 0.806	3.930 ± 0.658	4.773 ± 0.641	1.354 ± 0.240	2.901 ± 0.572
WH	0.587 ± 0.164	0.477 ± 0.134	-	-	0.507 ± 0.112
ZH	0.011 ± 0.003	0.004 ± 0.001	0.328 ± 0.085	0.183 ± 0.048	0.049 ± 0.007
N_{obs}	7	9	4	2	1

6 Upper Limits

Preliminary 95% C.L. upper limits on the Higgs production cross section in the VH mode with H $\to b\bar{b}$ were obtained from both the BDT and M(jj) analyses for a dataset corresponding to an integrated luminosity of 1.1 fb^{-1}. For the expected and observed limits, and the 1- and 2-σ bands, the CLs method currently recommended by the LHC Higgs Combination Group was employed [6].
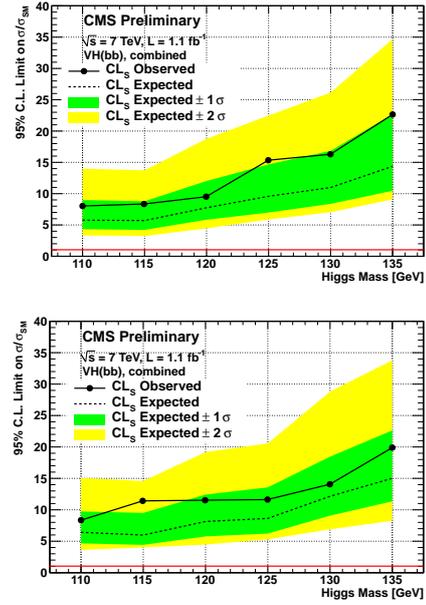
The results of the five BDT analyses are combined to produce limits on Higgs production in the $b\bar{b}$ channel for the assumed masses: 110 − 135 GeV. The identical procedure was applied to the results of the M(jj) analysis. Table 5 summarizes the resulting, expected and observed, upper 95% C.L. cross section limits, with respect to the standard model cross section, for each of the mass points for the BDT and M(jj) analyses. The results are displayed separately in Fig. 3. The primary result is the one from the BDT analysis.
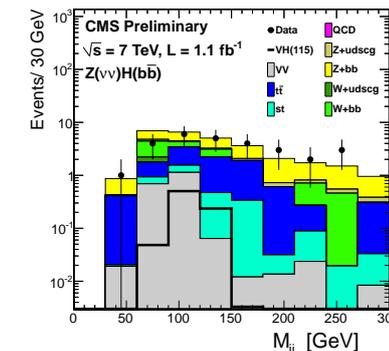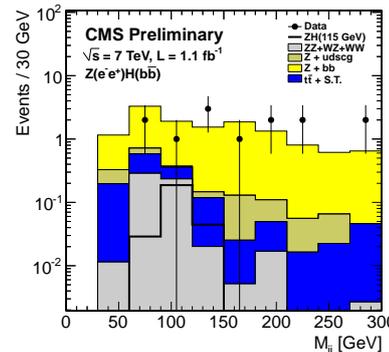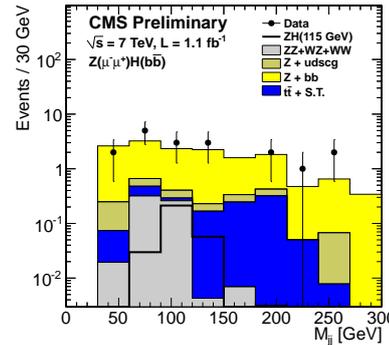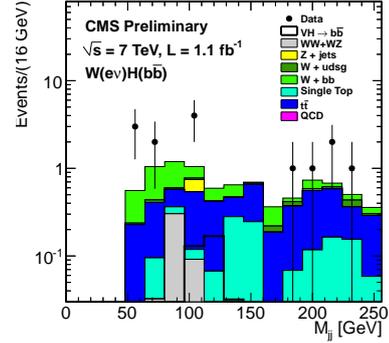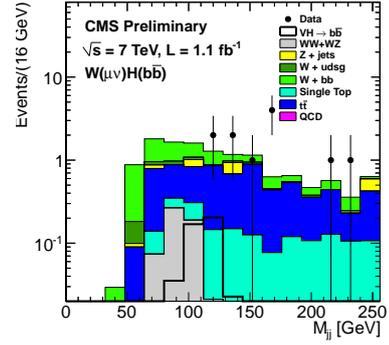
Fig. 1. Distributions of dijet invariant mass after all M(jj) selection criteria have been applied in (from top to bottom: W($\to \mu\nu$)H, W($\to e\nu$)H, Z($\to \mu\mu$)H, Z($\to ee$)H, Z($\Rightarrow \nu\nu$)H).

Fig. 3. Expected and observed 95% C.L. combined upper limits on the ratio of VHbb production for the BDT (top) and M(jj) (bottom) analyses. The median expected limit and the 1- and 2-σ bands are obtained with the LHC CLs method as implemented in LandS, as are the observed limits at each mass point.

Table 5. Expected and observed 95% CL upper limits on the production of a SM Higgs boson in association with W and Z bosons and decaying to b quarks relative to the expected cross section. Limits are listed separately for the BDT and M(jj) analyses.

M_H (GeV)	BDT Expected	BDT Observed	M(jj) Expected	M(jj) Observed
110	5.8	8.0	6.4	8.2
115	5.7	8.3	6.0	11.3
120	7.7	9.5	8.1	11.4
125	9.6	15.3	8.6	11.6
130	11.0	16.3	12.1	14.0
135	14.4	22.5	15.0	19.9

References

1. P. W. Higgs, "Broken symmetry and the mass of gauge vector mesons", *Phys. Rev. Lett.* 13 (1964) 508.
2. ALEPH, CDF, D0, DELPHI, L3, OPAL, SLD Collaborations, the LEP Electroweak Working Group, the Tevatron Electroweak Working Group, and the SLD Electroweak and Heavy Flavour Groups, "Precision electroweak measurements and constraints on the Standard Model", *ArXiv:1012.2367*.
3. CMS Collaboration, "Search for the Standard Model Higgs Boson Decaying to Bottom Quarks and Produced in Association with a W or a Z Boson", CMS Physics Analysis Summary, HIG-11-012 (2011).
4. CMS Collaboration, "The CMS experiment at the CERN LHC", *JINST* 3 (2008) S08004.
5. Byron P. Roe, Hai-Jun Yang, Ji Zhu, Yong Liu, Ion Stancu, Gordon McGregor, " Boosted Decision Trees as an Alternative to Artificial Neural Networks for Particle Identification", *Nucl.Instrum.Meth.*A543 (2005) 577-584.
6. CMS Collaboration,"Search for standard model Higgs boson in pp collisions at ps = 7 TeV", CMS Physics Analysis Summary, HIG-11-011 (2011).

Fig. 2. Distributions of BDT output for data (points with errors) and all backgrounds, from top to bottom: W($\to$ $\mu\nu$)H, W($\to$ $e\nu$)H, Z($\to$ $\mu\mu$)H, Z($\to$ ee)Hand Z($\to$ $\nu\nu$)H.