Standard Model Higgs Combination from CMS with up to 1.7 fb^{-1} of data

Michał Bluj[a,b] for the CMS Collaboration

CNRS-IN2P3/LLR-École Polytechnique, Palaiseau, France

Abstract. The combination is presented of searches for a standard model (SM) Higgs boson in eight decay modes: $H \to \gamma\gamma$, $H \to \tau\tau$, $H \to bb$, $H \to WW^* \to 2\ell2\nu$, $H \to ZZ^* \to 4\ell$, $H \to ZZ \to 2\ell2\tau$, $H \to ZZ \to 2\ell2\nu$, and $H \to ZZ \to 2\ell2q$. The searches were performed by the CMS Collaboration using 1.1–1.7 fb^{-1} of integrated luminosity, depending on the analysis. No excess compatible with a SM Higgs signal has been observed; the largest excursion of the observed data from the expected background has a probability of 0.4 after taking into account the look-elsewhere effect. The SM Higgs boson is excluded at 95% C.L. in three mass ranges 145–216, 226–288, and 310–400 GeV/c^2, while the expected exclusion range is 130–440 GeV/c^2.

1 Introduction

Understanding of the mechanism of electroweak symmetry breaking is one of the main goals of the CMS physics program. In the standard model (SM) the electroweak symmetry breaking is described by so-called Higgs mechanism which leads to prediction of one scalar particle - a Higgs boson (H) [1–6]. So far, experimental searches for this particle have given negative results allowing to exclude its mass below $m_H < 114.4 \text{ GeV/c}^2$ (LEP experiments [7]) and for $m_H \in 100\text{–}108 \text{ GeV/c}^2$ and $m_H \in 156\text{–}177 \text{ GeV/c}^2$ (Tevatron experiments [8]). Indirect searches using fits to precise measurements predict relatively light Higgs boson with $m_H < 158 \text{ GeV/c}^2$ [9].

In this report combination is presented of searches for the Higgs boson in eight decay modes: $H \to \gamma\gamma$, $H \to \tau\tau$, $H \to bb$, $H \to WW^* \to 2\ell2\nu$, $H \to ZZ^* \to 4\ell$, $H \to ZZ \to 2\ell2\tau$, $H \to ZZ \to 2\ell2\nu$, and $H \to ZZ \to 2\ell2q$.

The cross sections for Higgs boson production, its decay branching fractions, and their uncertainties are taken from the report prepared by LHC Higgs Cross Section group [10].

In Section 2 an overview of all eight analysis used in the combination is provided, then in Section 3 a statistical methodology used in this work is briefly described, and finally in Section 4 the combined result is presented.

2 Channels used in the combined search

The combination presented in this report bases on eight analyses corresponding to main decay modes of the Higgs boson as summarised in Table 1. In the following subsections a brief description is provided of each individual analysis.

2.1 $H \to \gamma\gamma$ channel [11]

The $H \to \gamma\gamma$ analysis is a search for a narrow peak in the diphoton mass $m_{\gamma\gamma}$ distribution on top of a large falling back-

---

[a] On leave from NCBJ, Warsaw, Poland
[b] e-mail: `michal.bluj@cern.ch`

Table 1. Summary information on the analyses included in the combination.

channel	mass range [GeV/c^2]	lumi. [fb^{-1}]	no. of sub-chan.	type of analysis
$H \to \gamma\gamma$	110–150	1.7	8	mass shape (unbinned)
$H \to \tau\tau$	110–140	1.1	6	mass shape (binned)
$H \to bb$	110–135	1.1	5	cut & count
$H \to WW^* \to 2\ell2\nu$	110–600	1.5	5	cut & count
$H \to ZZ^* \to 4\ell$	110–600	1.7	3	mass shape (unbinned)
$H \to ZZ \to 2\ell2\tau$	180–600	1.1	8	mass shape (unbinned)
$H \to ZZ \to 2\ell2\nu$	250–600	1.6	2	cut & count
$H \to ZZ \to 2\ell2q$	226–600	1.6	6	mass shape (unbinned)

ground. It is the most sensitive channel at low masses despite a small branching fraction ($Br(H \to \gamma\gamma) \sim 1\text{–}2 \times 10^{-3}$).

All events are divided into eight categories based on whether the transverse momentum of the di-photon system $p_T^{\gamma\gamma} > 40 \text{ GeV/c}$, whether both photons are in the central part of the CMS detector, and whether both photons are unconverted (have compact electromagnetic showers). The categorisation is motivated by different resolution in each category (1-3%).

The background under the expected signal peak is derived from sidebands without use of Monte Carlo simulation.

2.2 $H \to \tau\tau$ channel [12]

In this analysis a broad excess in the visible di-tau mass $m_{\tau\tau}^{\text{vis}}$ distribution is looked for (resolution of $m_{\tau\tau}^{\text{vis}} \sim 20\%$).

Three di-tau final states are used: $e\mu$, $e\tau_h$, $\mu\tau_h$ (τ_h stands for a τ decaying hadronically). Each of of these three categories is further divided into two mutually exclusive subcategories: events with the Vector Boson Fusion (VBF) signature (two jets separated in pseudorapidity with no additional jets in between), and events with less than two jets or with exactly two jets that fail VBF requirements.

The main irreducible background is $Z \to \tau\tau$ with normalisation taken from the $Z \to \ell\ell$ cross section measurement and shape of $m_{\tau\tau}^{\text{vis}}$ modelled using simulated events. The

reducible backgrounds (W+jets, QCD, t͞t, Z→ℓℓ) are evaluated basing on data.

2.3 H → bb **channel [13]**

The H → bb search exploits the Higgs boson production in association with W or Z bosons (V). The following W and Z boson decay modes are considered: W → ℓν and Z → ℓℓ, νν (ℓ = e, μ). It is required that the system of two b-tagged jets (a Higgs boson decay candidate) is boosted in the transverse plane, which reduces background and improves the di-jet mass resolution.

The result of the analysis bases on event count in signal regions defined by the output of a multivariate analysis classifier (MVA). The classifier was trained for a number of Higgs boson masses.

The rates of the main backgrounds V+jets, Vbb̄, and t͞t are estimated from control samples and then applied to simulation. The WZ and ZZ backgrounds with Z→bb̄, and the single-top background are modelled with Monte Carlo simulation.

2.4 H → WW* → 2ℓ2ν **channel [14]**

The signature of the H→WW*→2ℓ2ν signal is the presence of exactly two opposite sign, isolated leptons and significant E_T^{miss}. There is no signal mass pick due to escaping energy due to presence of two neutrinos from W decays. The search is based on event counting.

Events are split into three categories basing on a jet multiplicity in the event (0, 1 or 2 jets) with different signal-to-background ratios. For the 0-jet category the main background is the electroweak WW production; for the 1-jet category the WW and t͞t processes. Both the 0- and 1-jet categories are further split into same-flavour and opposite-flavour di-lepton sub-channels, since different contribution of the Drell-Yan background. The 2-jet category is optimised to take advantage of the VBF production signature (jets separation in pseudorapidity). The main background for this category is t͞t.

To separate the H→WW signal from the electroweak WW background the scalar nature of the Higgs boson is explored.

Contributions from all main backgrounds are estimated basing on data.

2.5 H → ZZ* → 4ℓ **channel [15]**

The H→ZZ*→4ℓ analysis is a search for a four-lepton mass peak over the continuum background. Three final states 4e, 4μ, 2e2μ are considered separately, as each of them has a different resolution of m₄ℓ and a different composition of background with jets faking leptons.

The dominant irreducible background is the electroweak ZZ production. Its contribution is modelled using simulation and normalised using the measured yield of Z → ℓℓ events scaled by the ratio of ZZ and Z cross sections. The reducible backgrounds with jets faking leptons (Z+jets, Zbb̄, t͞t) are evaluated using control regions. Their contribution was found to be small.

2.6 H → ZZ → 2ℓ2τ **channel [16]**

In the H→ZZ→2ℓ2τ search, the presence is required of one di-lepton pair (ee or μμ) forming an on-shell Z boson. Then a second Z boson is required to decay to τ-pair, with one of the following four decay modes: eμ, eτ_h, μτ_h, τ_hτ_h (τ_h stands for a τ decaying hadronically). It makes eight exclusive sub-channels. In the analysis, the mass of two leptons and visible products of two tau decays (without accounting for missing neutrinos) is a final observable.

The dominant background is the electroweak ZZ production which is taken from simulation and normalised using the measured yield of Z→ℓℓ events scaled by the ratio of ZZ and Z cross sections. The sub-leading backgrounds with jets faking tau come from Z+jets (including ZW) and t͞t, are evaluated from data using fake-rate method.

2.7 H → ZZ → 2ℓ2ν **channel [17]**

In the H→ZZ→2ℓ2ν analysis, events with one di-lepton pair (ee or μμ) consistent with an on-shell Z boson, and significant E_T^{miss} are selected. Then the transverse mass m_T from the di-lepton pair momenta and E_T^{miss} is constructed assuming that E_T^{miss} arises from the Z→νν decay[1]. Finally, events are counted in a m_H dependent window in the m_T distribution.

The main ZZ and WZ backgrounds are taken from simulation, while all other backgrounds are evaluated from control samples.

2.8 H → ZZ → 2ℓ2q **channel [18]**

The H→ZZ→2ℓ2q analysis is a search for a peak in the mass of the di-lepton plus di-jet system (m₂ℓ₂ⱼ). There are six exclusive final states used in the search with the lepton pair in one of two possible flavours (ee or μμ) and the jet pair with 0, 1 or 2 b-tags. Both lepton and jet pairs are required to be consistent with the Z boson mass. Background is further suppressed by employing a multivariate angular likelihood constructed from the kinematic variables of the two leptons and the two jets.

The background m₂ℓ₂ⱼ distribution is obtained using control regions in data.

3 Statistical methodology

The modified frequentist construction CLs [19, 20] is used for calculations of exclusion limits presented in this report. To completely define the method, one needs to make a choice of test statistic and of treatment of the systematic uncertainties in the construction of the test statistic and in generating pseudo-data. Here the LHC Higgs Combination Group prescription [21] is followed.

The likelihood $\mathcal{L}(data \,|\, \mu, \theta)$ used to construct the test statistic is defined as follows:

$$\mathcal{L}(data|\mu, \theta) = \mathrm{Poisson}\left(data|\mu \cdot s(\theta) + b(\theta)\right) \cdot p(\tilde{\theta}|\theta), \quad (1)$$

---

[1] $m_T^2 = \left(\sqrt{p_{T,Z}^2 + M_Z^2} + \sqrt{(E_T^{miss})^2 + M_Z^2} \right)^2 - \left(\overrightarrow{p}_{T,Z} + \overrightarrow{E}_T^{\mathrm{miss}} \right)^2$

where Poisson $(data|\mu \cdot s(\theta) + b(\theta))$ is the Poisson probability to observe "data", assuming signal and background models, $s(\theta)$ and $b(\theta)$, that depend on some nuisance parameters θ. The free parameter μ is a common signal strength modifier affecting signal event yields in all production modes (σ/σ_{SM}). Nuisance parameters θ correspond with independent sources of systematic uncertainties. The probability of "measuring" $\tilde{\theta}$ which is best known estimate of true value θ is $p(\tilde{\theta}|\theta)$ and describes the scale of the systematic uncertainty.

Then, the test statistics is definied as:

$$q_\mu = -2\ln\frac{\mathcal{L}(data|\mu, \hat{\theta}_\mu)}{\mathcal{L}(data|\hat{\mu}, \hat{\theta})}, \quad \text{where} 0 \le \hat{\mu} \le \mu, \quad (2)$$

where "data" stands either for the real observation or pseudo-data. Both the denominator and numerator are maximized. In the numerator, μ is fixed and only the nuisance parameters θ can float. Their values at which $\mathcal{L}$ reaches the maximum are denoted as $\hat{\theta}_\mu$. In the denominator, both μ and θ are allowed to float in the fit, and $\hat{\mu}$ and $\hat{\theta}$ are parameters at which $\mathcal{L}$ reaches its global maximum. The lower constraint on $\hat{\mu}$ $(0 \le \hat{\mu})$ is imposed as the signal rate cannot be negative; the upper constraint $(\hat{\mu} \le \mu)$ forces the limit to be one-sided. The value of the test statistic for the real observation is denoted as q_{obs}.

In the next step, the values of the nuisance parameters $\hat{\theta}_0^{obs}$ and $\hat{\theta}_\mu^{obs}$ best describing the observed data (maximizing $\mathcal{L}$) are obtained for the background-only and signal+background hypotheses, respectively. Using these best-fit values of the nuisance parameters, toy Monte Carlo pseudo-data is generated to construct the test statistic sampling distributions for the both signal+background hypothesis (with signal strength μ) and for the background-only hypothesis (μ=0). The "measurements" $\tilde{\theta}$ are also randomized in each pseudo-data.

With signal+background and background-only sampling distributions for the test statistic q_μ one can find the probability to obtain a test statistic value as high as, or higher than, the one observed in data, under the signal+background hypothesis, and obtain $CLs(\mu)$ from the ratio

$$CL_s(\mu) = \frac{P\left(q_\mu \ge q_\mu^{obs}|\mu \cdot s(\hat{\theta}_\mu^{obs}) + b(\hat{\theta}_\mu^{obs})\right)}{P\left(q_\mu \ge q_\mu^{obs}|b(\hat{\theta}_0^{obs})\right)}. \quad (3)$$

The $CLs \le \alpha$ for a given μ means that the signal with strenth μ is excluded at the $(1 - \alpha)$ confidence level (C.L.). To quote the 95% C.L. upper limit on μ, we adjust μ until we reach $CLs = 1 - 0.95$.

To quantify an excess of events the alternative test statistic q_0 is used:

$$q_0 = -2\ln\frac{\mathcal{L}(data\,|\,0, \hat{\theta}_0)}{\mathcal{L}(data\,|\,\hat{\mu}, \hat{\theta})} \quad \text{and } \hat{\mu} \ge 0. \quad (4)$$

This test statistic allows to evaluate significances (Z) and p-values (p_0) from the asymptotic formula:

$$Z = \sqrt{q_0^{obs}}, \quad (5)$$

$$p_0 = P(q_0 \ge q_0^{obs}) = \int_Z^\infty \frac{e^{-x^2/2}}{\sqrt{2\pi}}\,dx = \frac{1}{2}\left[1 - \text{erf}\left(Z/\sqrt{2}\right)\right], \quad (6)$$

where q_0^{obs} is the observed test statistic for $\mu = 0$.

4 Result of combined search

The combined result of the search for the SM boson is presented in Figure 1 which shows the combined 95% C.L. upper limits on the signal strength modifier (σ/σ_{SM}) as a function of the Higgs boson mass. The dashed line stands for the median expected results for the background-only hypothesis, while the green (yellow) bands indicate the range of 68% (95%) deviation from the median. The observed limit is indicated by the solid line with points. The SM Higgs boson is excluded at 95% C.L. in three mass ranges 145–216, 226–288, and 310–400 GeV/c², while the expected exclusion range is 130–440 GeV/c².
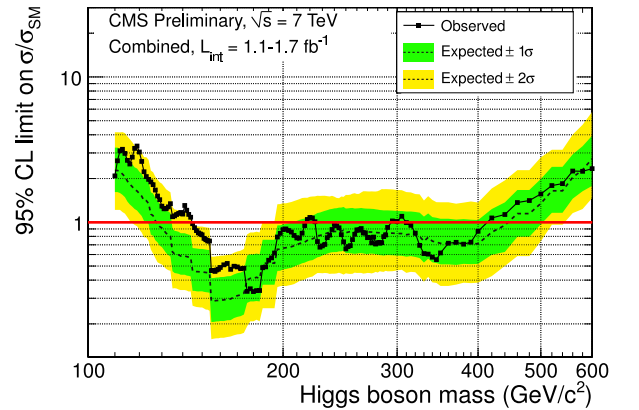

Fig. 1. The 95% C.L. upper limit on the signal cross section σ/σ_{SM} for the SM Higgs boson hypothesis as a function of the SM Higgs boson mass. The observed values are shown by the solid line. The dashed black line indicates the median expected results for the background-only hypothesis, while the green (yellow) bands indicate the range of 68% (95%) deviation from the median.

A contribution of each individual Higgs decay mode to the combined result is illustrated in Figure 2. For high masses ($m_H > 200$ GeV/c²) the exclusion is driven by the H→ZZ decay channels, for an intermediate range 130–200 GeV/c² the exclusion is dominated by the H→WW*→2ℓ2ν mode, while for low masses ($m_H < 130$ GeV/c²) H→γγ is the most significant contributor.

The differences between the observed and expected limits are consistent with statistical fluctuations, as the observed limits lie within the 68% and 95% bands. For the low Higgs boson mass range, we observe an excess of events which leads to weaker observed limit than expected in the absence of the SM Higgs boson. The observed local p-value p_0 which quantifies the consistency of the observed excesses with the background-only hypothesis, is shown in Figure 3 (top panel) for the combined search, and split into individual decay modes in Figure 4. A broad offset for low masses, of about 1σ, corresponds to the excesses seen in the H→WW*→2ℓ2ν channel characterised by poor mass resolution. The excesses observed in the H→ZZ→4ℓ and the H→γγ channels result two narrow minima of the p-value. The minimal p-value is $p_{min} \sim 0.01$, but after accounting for the look-elsewhere effect which is important for this study, it is reduced to a global probability $p_{global} \sim 0.4$. The look-elsewhere effect was esimeted for the whole explored mass range 110–600 GeV/c².
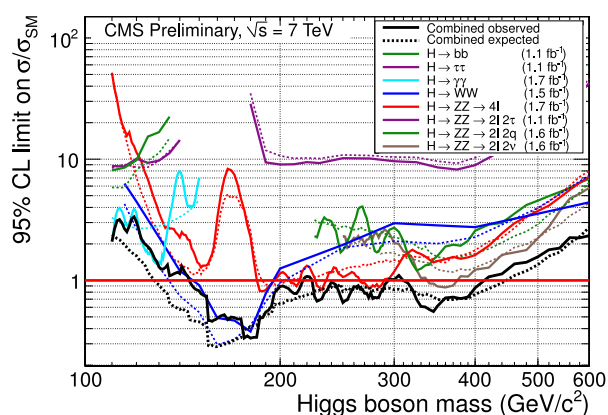
Fig. 2. The observed 95% C.L. upper limit on the signal cross section σ/σ_{SM} for the SM Higgs boson hypothesis as a function of the SM Higgs boson mass for eight decay modes and their combination.

The best-fit value of σ/σ_{SM} is also presented in Figure 3 (bottom panel). The best-fit value is a factor by which the SM Higgs boson cross section has to be rescaled to best describe observed data. In the mass range between 115 and 125 GeV/c^2 the best-fit value agrees within uncertainty with 1 ($\sigma = \sigma_{SM}$), but as discussed above, with the analysed amount of data the excess is not significant. More data will increase the statistical accuracy of the search thus improve its sensitivity.
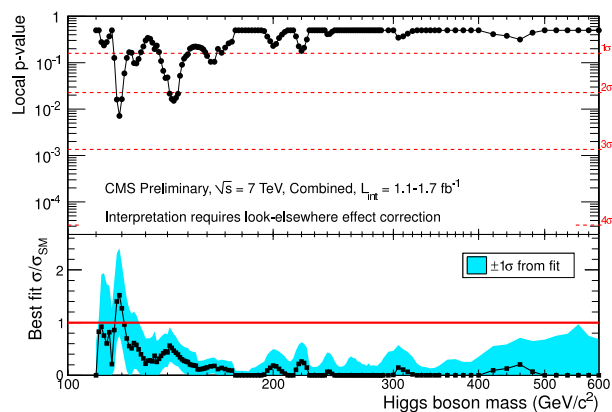

Fig. 3. The observed local p-value p_0 (top panel) and σ/σ_{SM} of the best-fit (bottom panel) as a function of the SM Higgs boson mass. The maximal observed excess (minimal local p-value) corresponds with a global probability for background-only hypothesis equal to 0.4 after accounting for the look-elsewhere effect for $m_H = 110$–600 GeV/c^2.

5 Conclusions

The combined search for the standard model Higgs boson performed by the CMS Collaboration with up to 1.7 fb^{-1} of data was presented. The expected exclusion mass range is 130–440 GeV/c^2. The observed data allowed to exclude the SM Higgs boson at 95% C.L. in three mass ranges 145–216, 226–288, and 310–400 GeV/c^2. The largest excursion of the observed data from the expected background
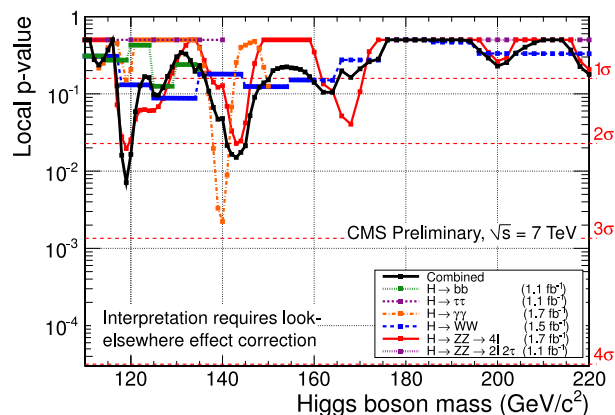

Fig. 4. The observed local p-value as a function of the SM Higgs boson mass for eight decay modes and their combination.

has a probability of 0.4 after taking into account the look-elsewhere effect for the whole explored mass range (110–600 GeV/c^2).

References

1. F. Englert and R. Brout, Phys. Rev. Lett. **13** (1964) 321.
2. P. W. Higgs, Phys. Lett. **12** (1964) 132.
3. P. W. Higgs, Phys. Rev. Lett. **13** (1964) 508.
4. G. S. Guralnik, C. R. Hagen and T. W. B. Kibble, Phys. Rev. Lett. **13** (1964) 585.
5. P. W. Higgs, Phys. Rev. **145** (1966) 1156.
6. T. W. B. Kibble, Phys. Rev. **155** (1967) 1554.
7. R. Barate *et al.* [LEP Working Group for Higgs boson searches and ALEPH and DELPHI and L3 and OPAL Collaborations], Phys. Lett. B **565** (2003) 61
8. [TEVNPH (Tevatron New Phenomina and Higgs Working Group) and CDF and D0 Collaborations], arXiv:1107.5518 [hep-ex].
9. [ALEPH and CDF and D0 and DELPHI and L3 and OPAL and SLD and LEP Electroweak Working Group and Tevatron Electroweak Working Group and SLD Electroweak and Heavy Flavour Groups Collaborations], arXiv:1012.2367 [hep-ex]
10. S. Dittmaier *et al.* [LHC Higgs Cross Section Working Group Collaboration], arXiv:1101.0593 [hep-ph].
11. CMS Collaboration, CMS PAS HIG-11-021 (2011)
12. CMS Collaboration, CMS PAS HIG-11-009 (2011)
13. CMS Collaboration, CMS PAS HIG-11-012 (2011)
14. CMS Collaboration, CMS PAS HIG-11-014 (2011)
15. CMS Collaboration, CMS PAS HIG-11-015 (2011)
16. CMS Collaboration, CMS PAS HIG-11-013 (2011)
17. CMS Collaboration, CMS PAS HIG-11-016 (2011)
18. CMS Collaboration, CMS PAS HIG-11-017 (2011)
19. A. L. Read, In *Geneva 2000, Confidence limits* 81-101
20. T. Junk, Nucl. Instrum. Meth. A **434** (1999) 435 [hep-ex/9902006]
21. ATLAS and CMS Collaborations and LHC Higgs Combination Group, ATL-PHYS-PUB-2011-11, CMS NOTE-2011/005 (2011)