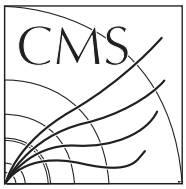
Available on the CMS information server **CMS NOTE AN-13-064**



The Compact Muon Solenoid Experiment

CMS Draft Note

Mailing address: CMS CERN, CH-1211 GENEVA 23, Switzerland

2013/05/24
Head Id:
Archive Id: 187417
Archive Date: 2013/05/23
Archive Tag: trunk

Search for a standard-model-like Higgs boson in the decay channel $H \to ZZ \to 2\ell 2q$ at $\sqrt{s} = 8$ TeV

A. De Cosa (*U. & INFN-Napoli*), F. Fabozzi (*U. Basilicata & INFN-Napoli*), D. Bortoletto, M. Kress, M. Vidal (*Purdue U.*), M. Yalvac (*Middle East Technical U.*), Y.J. Lu, S.S. Yu (*National Central U.*), L. K. Saini (*Panjab U.*), S. Bolognesi, A. Whitbeck (*Johns Hopkins U.*), M. Mozer (*KIT*), G. Codispoti, J. Fernández de Trocóniz (*U. Autonoma de Madrid*), D. Domínguez, O. González, J.M. Hernández, E. Navarro, P. Garcia-Abia (*CIEMAT-Madrid*), and K. Kanishchev (*INFN-Padova*)

Abstract

A search for a standard-model-like Higgs boson decaying into two Z bosons with subsequent decay into two leptons and two quark-jets, $H \to ZZ \to \ell^+\ell^- q\bar{q}$ is performed. The analysis uses 19.6 fb^{-1} of data collected by the CMS experiment from proton-proton collisions produced in LHC at $\sqrt{s} = 8$ TeV. The analysis exploits the kinematic information and the flavor tagging of the leading particles of the event in order to isolate hypothetical Higgs boson signals with mass values in the range from 230 GeV/c^2 to 650 GeV/c^2. No evidence of a Higgs boson signal is found and upper limits are set on the Higgs boson production cross section in that mass range.

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 the Physics of electroweak symmetry breaking. A suitable Higgs boson candiate, predicted by the Higgs mechanism, has recently been found with a mass around 126 GeV [1, 2]. However, many models predict more than a single boson, so we present here further searches for Higgs like particles using the SM Higgs as a benchmark model.

The CMS collaboration is performing searches for the Higgs boson in several decay modes. This comprehensive effort aims at gaining sensitivity over a large range of Higgs boson masses, M_{H}, by combining many different analyses. We expect that this effort will finally explore the region with $M_{\mathrm{H}} \geq 2m_{\mathrm{Z}}$ which is not probed at the Tevatron [3].

In this note we report a study of the search for the Higgs boson in $\mathrm{H} \to \mathrm{ZZ}$ when one of Z decays as $\mathrm{Z} \to \ell^+\ell^-$ (where ℓ is either electron or muon) and the other as $\mathrm{Z} \to \mathrm{q\bar{q}}$, using LHC proton-proton collision data at $\sqrt{s} = 8$ TeV. The mass of the ZZ system, $M_{\ell\ell jj}$, fully reconstructed from the two jets steming from the $\mathrm{q\bar{q}}$ fragmentation and from the two leptons, is a resonance for hypothetical Higgs boson signals, narrower at lower M_{H} values, significantly increasing at high M_{H}. The dominant background process is Z+jets production from Drell-Yan processes, with smaller contributions from top-quark decays and from diboson events. Unlike for the Higgs boson, the $M_{\ell\ell jj}$ distribution of the background events is not resonant, providing a useful handle to isolate 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 normalize and tune the dominant Drell-Yan plus jets background reducing the dependence on the simulation. The contamination from $\mathrm{t\bar{t}}$ events is directly extracted from data. A similar analysis [4] was performed by CMS using data at $\sqrt{s} = 7$ TeV.

2 Data and simulated samples

The analysis uses data from proton-proton collisions produced in the LHC at a centre-of-mass energy, $\sqrt{s}$, of 8 TeV, with an integrated luminosity of 19.6 fb^{-1}.

The data, collected by the CMS experiment in the data taking campaign of year 2012, are reconstructed using the official CMS software CMSSW, release 5_3_X. Data of the various data taking periods are packed into different primary datasets according to the signatures of particles and jets (physics objects) identified by the high level trigger (HLT). Several of these primary datasets are used in order to cover the final states expected in the $\mathrm{H} \to \mathrm{ZZ} \to \ell^+\ell^-\,\mathrm{q\bar{q}}$ processes and to perform dedicated background studies (details in appendix A).

Only data that pass the strict quality requirements imposed by the CMS central certification team are used in the analysis (more technically, the latest available reprocessings and official JSON files are utilized for each data taking period).

Official samples of simulated events (MC), produced in the Summer12 simulation campaign, are used in order to study the properties of the Higgs boson events (for M_{H} values in the range 230 GeV/c^2 to 650 GeV/c^2) and of the background processes. These samples are also reconstructed with the official software CMSSW, release 5_3_X.

The available signal samples of different M_{H} values are listed in appendix A along with the cross-sections times branching fraction of the process $\mathrm{H} \to \mathrm{ZZ} \to \ell^+\ell^-\,\mathrm{q\bar{q}}$. The H production

cross section and the $H \to ZZ$ branching fraction are provided as function of the Higgs boson mass by the LHC Higgs cross section working group [5, 6]. The branching ratio of the Z boson into pairs of leptons (e, μ and τ) and quarks are taken from the PDG [7]. The samples are generated using the POWHEG [8] event generation program.

The samples of background simulated events are displayed in appendix A, together with their cross section (calculated at NLO and NNLO [9] dependending on the process) and equivalent luminosity.

To characterize the dominant Drell-Yan background both inclusive and parton-exclusive Z+jets samples are used, produced with the MadGraph [10] event generator imposing a high mass of the dilepton pair ($M_{\ell\ell} > 50\,\text{GeV}$). The background from top events is due mainly to $t\bar{t}$ events, for whose study a $t\bar{t} \to 2l2\nu2b$ sample is produced using POWHEG. Diboson events from standard model processes are simulated with inclusive ZZ, WZ, and WW samples using Pythia [11].

All the events in the analysis, either data or simulated, are processed with the official CMS tools for analysis (PAT) to ensure standard object definitions (particle flow objects). The common skim of the CMS H2l2q analysis team is described in [12].

3 Event reconstruction

The signature of Higgs boson signal events is a lepton pair and a quark pair, each with an invariant mass around the Z boson mass. The invariant mass of the $\ell^+\ell^-q\bar{q}$ system, $M_{\ell\ell jj}$, is consistent with the mass of a hypothetical Higgs boson and is used as the main observable to discriminate signal events from background events.

Particles are reconstructed in CMS using a particle flow algorithm from the electronic signals they leave in the detector. Leptons (electrons [13] and muons [14]) and jets [15, 16] are selected imposing quality criteria to ensure high efficiency in their reconstruction and identification, and good momentum and mass resolutions.

Data events used in the analysis belong to the *DoubleMu* and *DoubleElectron* primary datasets, which are built from the un-prescaled triggers `HLT_Mu17_Mu8` (*DoubleMu*) and `HLT_Ele17_CaloIdT_TrkIdVL_CaloIsoVL_TrkIsoVL_Ele8_CaloIdT_TrkIdVL_CaloIsoVL_TrkIsoVL` (*DoubleElectron*), among other un-prescaled and prescaled triggers. The event selection requirements of the analysis are tighter than those of the trigger. More details on the trigger strategies for Higgs boson searches are in [17].

To avoid dependence on the details of the trigger emulation no trigger requirements are imposed on simulated events. Instead, event weights are assigned to simulated events according to the probabilities of leptons to pass the trigger (detailed in appendix B). The trigger efficiency tables for leptons are computed in bins of (p_T, η) from data using a tag-&-probe technique [18]. We have checked that the trigger simulation yields similar efficiencies. See discussion in appendix B.

The identification of $Z \to e^+e^-$, $Z \to \mu^+\mu^-$, and $Z \to q\bar{q}$, is a crucial step of the analysis. Leptonic decays of Z bosons are built from same-flavour opposite-charge lepton pairs, which satisfy lepton identification criteria. Hadronic Z boson decays are made of pairs of jets. Leptons and jets are required to fulfill additional kinematic constraints, described later.

Electrons are reconstructed with the GSF algorithm [19]. In order to ensure good electron reconstruction, the electron supercluster is required to be inside the ECAL acceptance volume

($|\eta| < 2.5$) and outside the ECAL barrel-endcap overlap region ($1.4442 < |\eta| < 1.566$). They must satisfy the standard *loose working point* of the cut-based electron identification for 2012 analyses [20]. Muons are reconstructed by both the GlobalMuon and the particle flow muon reconstruction algorithms and are required to lie in the acceptance region $|\eta| < 2.4$. They must satisfy the standard *tight working point* of the cut-based muon identification for 2012 analyses [21]. The precise electron and muon identification cuts are detailed in appendix C. They comprise identification and isolation criteria and, specifically for electrons, conversions rejection requirements.

In the particle flow algorithm, lepton isolation is defined as the sum of p_{T} or E_{T} deposits of charged and neutral hadrons, and photons, computed in a ΔR cone around the lepton direction. To ensure independence of the isolation from the number of pile-up interactions and to reduce the probability of jets to overlap with the cone, a standard recipe recommended by the e-gamma and muon POGs is used. For electrons, the overall pile-up energy contribution is estimated as the average energy density in the event multiplied by an effective area, using the recommended cone size $\Delta R < 0.3$. For muons, the recommended DeltaBeta correction with $\Delta R < 0.4$ is applied.

The lepton identification efficiencies are evaluated from data using a standard tag-&-probe method, which requires the reconstruction of the dilepton system with invariant mass in the range [60-120] GeV/c^2. Appendix C provides details of the tag-&-probe method and the scale factors for electron and muon selection requirements, which are very close to 1.

The $\mathrm{Z} \to \mathrm{e}^+\mathrm{e}^-$ and $\mathrm{Z} \to \mu^+\mu^-$ candidates are constructed from leptons with momenta $p_{\mathrm{T}} > 40$ GeV/c (leading lepton) and $p_{\mathrm{T}} > 20$ GeV/c (subleading lepton). Z decays into leptons are accepted for subsequent analysis if the $\ell^+\ell^-$ invariant mass lies within $[76, 106]$ GeV/c^2 (Figure 1). Not being included the state-of-the-art lepton calibrations originates visible differences in the $M_{\ell^+\ell^-}$ distributions in data and simulations. Reweighting the simulated events to match the peak position and resolution in data has a tiny impact in the signal efficiency, less than 0.2%. The $M_{\ell^+\ell^-}$ resolution is around 10% better in the simulation than in data. The correction of this difference has a negligible impact in the $M_{\ell\ell jj}$ distributions, largely dominated by the dijet resolution.

The particle flow jets are reconstructed with the anti-k_T algorithm [22] with radius parameter $R = 0.5$. In order to obtain high reconstruction efficiency and precise energy measurements, jets are required to be inside the tracker acceptance, $|\eta| < 2.4$. Jet energy corrections are applied to data and simulated events [23]. The so-called Fastjet algorithm [24] applies L1 energy corrections to account for pile-up. Jets originating from pile-up interactions are removed requiring jets to be taggable [25] and $\beta \geq 0.2$, where β is the sum of transverse momenta of the charged particles in the jet coming from the primary vertex normalized to the sum of transverse momenta of all the charged particles in the jet. Jets must have $p_T > 30\,\mathrm{GeV}/c$ to form $\mathrm{Z} \to \mathrm{q\bar{q}}$ candidates, which rejects fake candidates with low-p_{T} jets from QCD background. The invariant mass distribution of $\mathrm{Z} \to \mathrm{q\bar{q}}$ candidates is displayed in Figure 2.

Appendix D contains the p_{T} distributions of leptons and jets after preselection cuts.

4 Event selection

The $\mathrm{Z} \to \ell^+\ell^-$ and $\mathrm{Z} \to \mathrm{q\bar{q}}$ decays are combined into $\ell^+\ell^- jj$ candidates, labeled Higgs boson candidates. To avoid double counting of the same object reconstructed in different collections (for instance, leptons inside a jet) the angular separation of isolated leptons and jets is required

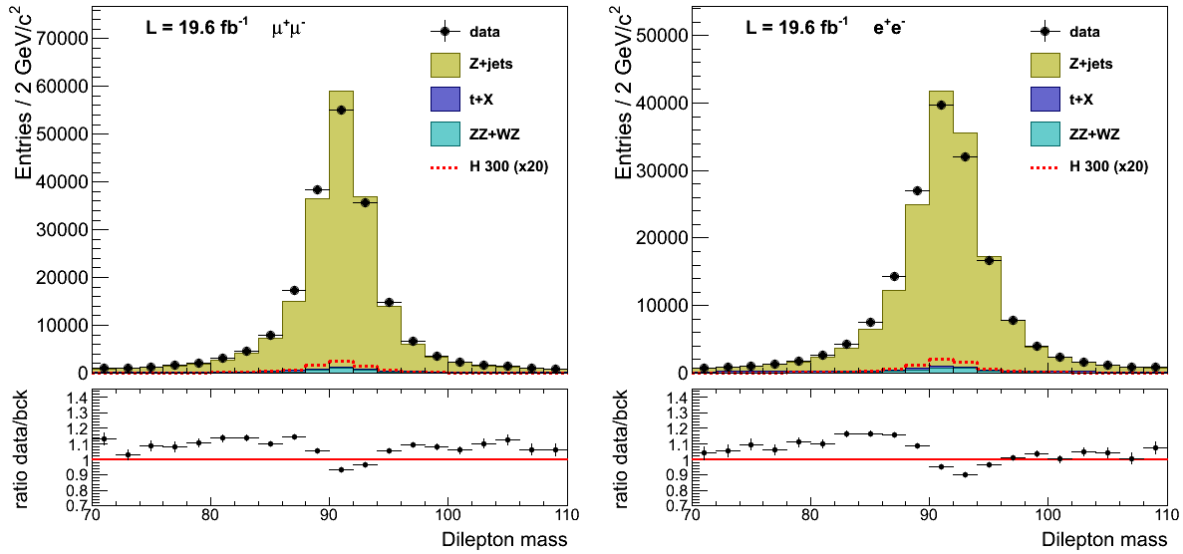


Figure 1: Dilepton invariant mass in data and simulation of $Z \to \ell^+\ell^-$ candidates after lepton selection cuts for muons (left) and electrons (right).

to be $\Delta R = \sqrt{\Delta\phi^2 + \Delta\eta^2} > 0.5$. In the following, the entire selection procedure described above is referred to as "preselection" of $\ell^+\ell^- jj$ candidates.

In order to suppress the dominant Drell-Yan Z+jets background and contamination from $t\bar{t}$ events, $\ell^+\ell^- jj$ candidates are selected in the m_{jj} region $[71, 111]$ GeV/c^2, called signal region. Outside of this signal region, candidates with m_{jj} within $[60, 130]$ GeV/c^2 are used for background determination. This is the so-called sideband region.

Due to the relatively large branching fraction of the Z-bosons decaying into a pair of bottom-antibottom quarks, compared to the abundance of light-quark or gluon jets in Z+jets background events, the *Jet Probability* b-tagging algorithm [26] is used to identify jets originating from heavy-flavour quarks. No selection of candidates is performed based on btagging probabilities. Rather, events are clasified into three exclusive categories: 2-, 1-, and 0-btag. The 2-btag category includes events with one jet tagged using JPM (medium working point of the JP algorithm) and another jet tagged using JPL (loose working point). The 1-btag category includes events with at least one jet tagged using JPL. The untagged events belong to the 0-btag category. The different signal-to-background ratio of each category helps improving the sensitivity of the analysis. The methods used to estimate the backgrounds are common to all the categories.

Using simulated Drell-Yan Z+jets events, the average b-tagging efficiencies $\langle \epsilon_i^{MC} \rangle$ and scale factors $\langle \mathrm{SF}_i \rangle$ are obtained for jets in events passing selection cuts. The scale factors are calculated, using the Moriond13 prescription, for b, c, and light jets separately. The average tagging efficiencies in the data are calculated as the product of the average tagging efficiencies for MC and the corresponding average scale factors: $\langle \epsilon_i \rangle = \langle \epsilon_i^{MC} \rangle \cdot \mathrm{SF}_i$, $i = b,\ c,$ light. For b and c jets these average data tagging effiencies are used. In the case of light jets (ℓ, mistags), data mistag rates were provided by the Btag POG in 2011, as a function of jet p_T and $|\eta|$. To take into account the jet p_T and $|\eta|$ dependence, the average data mistag rate using the 2011 prescription $\langle \epsilon_\ell^{2011} \rangle$ was calculated as well, defining $\epsilon_\ell(p_T, |\eta|) = \epsilon_\ell^{2011}(p_\mathrm{T}, |\eta|)\, \langle \epsilon_\ell \rangle / \langle \epsilon_\ell^{2011} \rangle$. The values of $\langle \epsilon_\ell \rangle / \langle \epsilon_\ell^{2011} \rangle$ are 1.126 (1.089) for JPL (JPM) mistags.

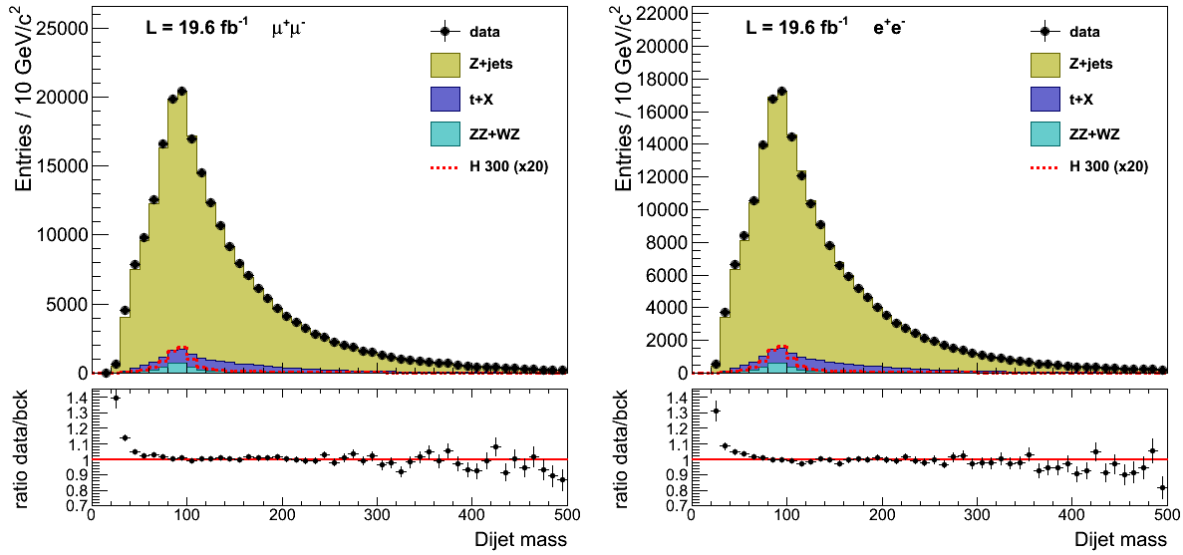
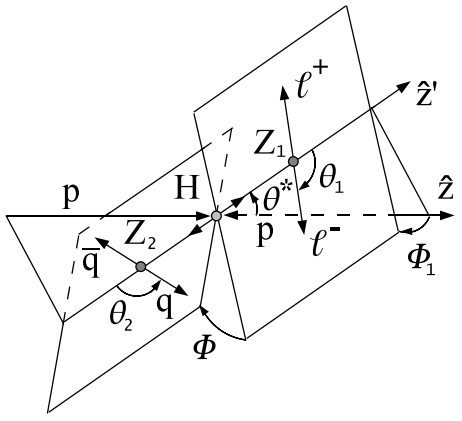


Figure 2: Dijet invariant mass in data and simulations for $Z \to q\bar{q}$ candidates for events with dimuons (left) and dielectrons (right) after preselection cuts. The background distributions are normalized to data.

The Higgs boson candidate is selected among those in the M_{jj} signal region that pass all the selection criteria. When no candidate survives the selection cuts in the signal region, the Higgs boson candidate is chosen from the candidates in the sideband. The event is rejected whenever no candidates are found in either region. In less than 3% of the events after final selection there is more than one Higgs boson candidate per event. In these cases, the candidate in the highest btag category is selected. If more than one candidate remains, the one with minimum $|M_{\ell^+\ell^-} - m_Z| + |M_{jj} - m_Z|$ is chosen.

There are several features in the signal $H \to ZZ \to 2l2j$ decay kinematics which discriminate it against background. These kinematic differences are exploited to optimize the selection and maximize the signal significance. Five helicity-dependent angular observables fully describe kinematics in the decay $2 \to 1 \to 2 \to 4$ as in $ab \to X \to ZZ \to 2l2q$ [27, 28]. They are orthogonal to the three invariant masses of the X and the two Z bosons. Longitudinal and transverse momenta of the X are also additional orthogonal observables and could be used in analyses, but they typically have weaker discrimination power and rely on modeling of the PDFs and process dynamics. The above orthogonal observables are largely uncorrelated and are more attractive to be used in event selection rather than raw kinematic observables.

The diagram on the right illustrates the production and decay chain $ab \to X \to ZZ \to 2\ell 2q$, which is a function of three helicity angles θ_1, θ_2, and Φ, and two production angles θ^* and Φ_1. Parameterization of both signal and background distributions have been derived and implemented in [27, 29]. A linear likelihood discriminant, LD, is constructed from the signal and background probabilities defined using the five angles. Signal sensitivity is improved requiring events with $LD > 0.5$, a selection cut



almost optimal for the three btag categories. A comparison of angular distribution in data and simulation is presented in Figure 3 for events with e^+e^- and $\mu^+\mu^-$ combined. Good agreement is appreciated, except in the low LD region dominated by background.

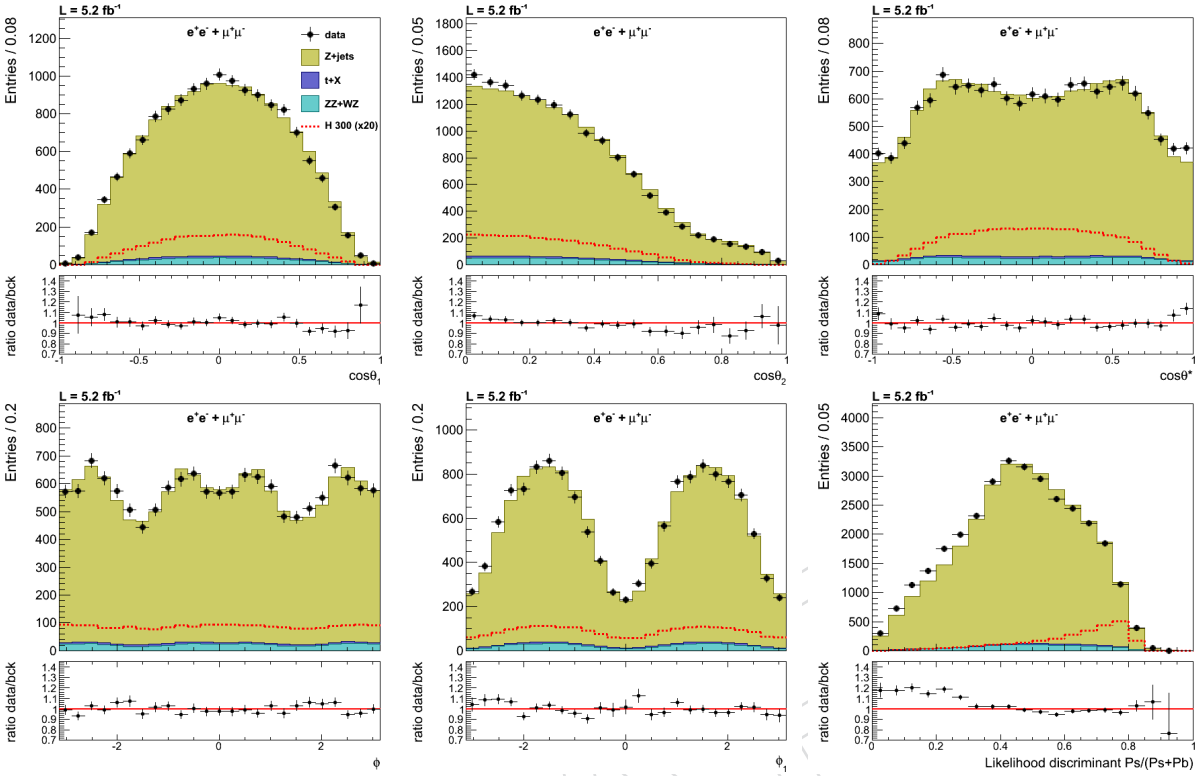


Figure 3: Distributions of the five helicity angles and the likelihood discriminant, LD, after final selection (for the latter, the cut LD > 0.5 is not applied) for unblinded ICHEP 2012 data (points with error bars) and simulated samples (histograms). The red dashed histogram indicates 20 times the expected distribution for a 300 GeV/c^2 mass Higgs boson. The distributions of the background are normalized to the data in the m_{jj} sideband region.

An important background in the 2 b-tag category originates from $t\bar{t}$ decays, which contain two true b-quark jets (Figure 4 (left)). This background is reduced requiring a particle flow missing E_T significance [30] less than 10 (Figure 4 (right)) (applied to the three btag categories). Small differences in the MET significance of the data and the Drell-Yan Z+jets simulation are taken into account studying the region MET significance < 6, where the contamination of the t+X background is reduced. It is observed that a simple multiplicative factor of 1.15 ± 0.01 brings the MC distribution in very good agreement with the data, a factor that is stable for muons and electrons, and as a function of the number of vertices in the event. The validity of this factor has been checked up to larger values of the MET significance in a top-depleated data subsample, requiring $p_T(\ell\ell jj) < 20$ GeV/c. Therefore the MET significance in the Drell-Yan, diboson and Higgs boson MC events is multiplied by a factor 1.15. The effect of this factor on the efficiency of the cut MET significance is minimal (in all cases the efficiency changes by less than 1%).

Selection cuts for the final analysis are listed in table 2, identical for the three b-tag categories. Figure 5 displays the numbers of events after each cut for the data and the background simulation. The efficiency of each cut is listed in Table 1.

The expected numbers of signal and background events in the electron and muon channels, in the $M_{\ell\ell jj}$ range [150, 750] GeV/c^2, are listed for 5.2 fb^{-1} in table 3 (including ICHEP data)

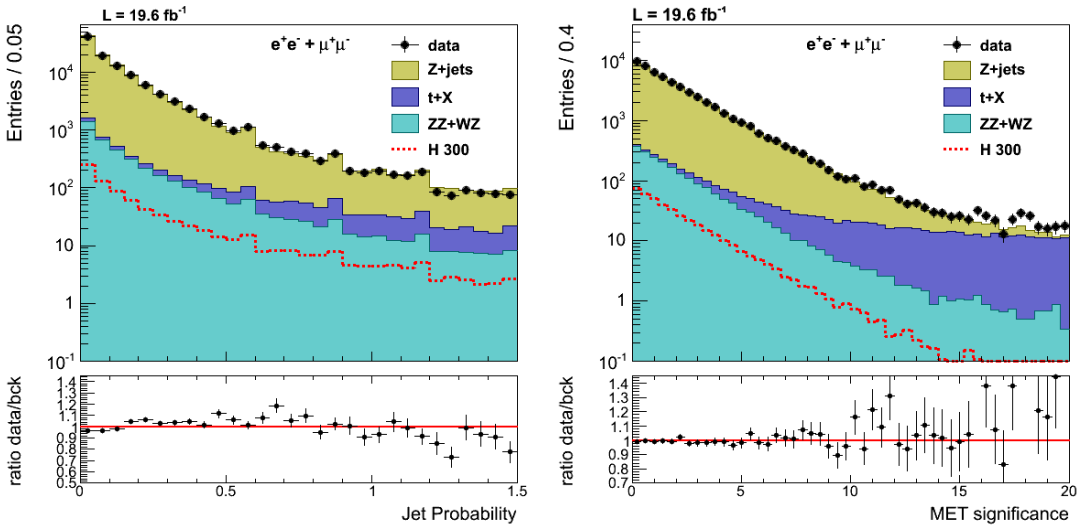


Figure 4: Left: jet probability b-tagger, after all selection cuts, for the three btag categories together. Right: corrected particle flow MET significance in data and simulation for events with electrons and muon combined, after the full selection except the cut on MET significance. The distributions of the background are normalized to the data in the m_{jj} sideband region.

Table 1: Relative and cumulative efficiencies of the cuts for the data and the background simulation.

	data		H (300 GeV/c^2)		Z+jets		Diboson		t+X	
cut	rel.	cumul.	rel.	cumul.	rel.	cumul.	rel.	cumul.	rel.	cumul.
presel.	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00
$m_{\ell\ell}$	0.85	0.85	0.94	0.94	0.93	0.93	0.86	0.86	0.22	0.22
LD	0.53	0.45	0.84	0.79	0.54	0.51	0.58	0.50	0.75	0.17
MET sig.	0.97	0.44	0.99	0.78	0.99	0.50	0.95	0.47	0.40	0.07
m_{jj}	0.29	0.13	0.65	0.51	0.29	0.15	0.57	0.27	0.32	0.02

and for 19.6 fb^{-1} in table 4. Appendix E details of the contributions of the various background sources in the mass range $[M_H - 6\%, M_H + 10\%]$, for different M_H values. ICHEP data referes to the 5.2 fb^{-1} subset of the 2012 data unblinded last summer for the ICHEP conference. These data are used for cross checks in the signal region. The rest of the data, almost 15 fb^{-1}, remain blinded in the signal region after the preselection cuts.

5 Background determination

After the event selection described in section 4, the following SM processes are considered as background in this analysis: diboson, $Z + jets$, and $t\bar{t}$ production.

The diboson production (ZZ and WZ mainly) is simulated using MC while the other two contributions are estimated, either using MC simulation corrected to data in control regions ($Z + jets$), or extracted directly from data in control regions ($t\bar{t}$). The latter two cases are described in detail in the following subsections.

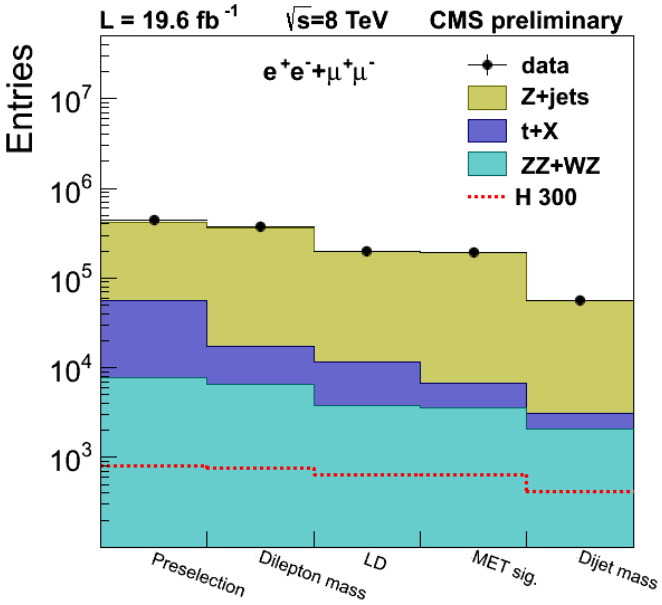


Figure 5: Numbers of events after each cut for the data and the background simulation.

Table 2: Selection requirements in the three btag categories.

observable	0 btag	1 btag	2 btag
btag	none	JPL	JPM & JPL
helicity LD	> 0.5		
missing E_T significance	< 10		
m_{jj}	[71,111] GeV/c^2		
$m_{\ell\ell}$	[76,106] GeV/c^2		
$p_T(l^{\pm})$	$> 40/20$ GeV/c		
p_T(jets)	> 30 GeV/c		
$\|\eta\|(\ell^{\pm})$	$(e^{\pm}) < 2.5$, $(\mu^{\pm}) < 2.4$		
$\|\eta\|$(jets)	< 2.4		
lepton quality	see section 3		
jet quality	see section 3		

Drell-Yan Z+jets contamination

Most events in the selected $\ell^+\ell^-q\bar{q}$ sample come from Drell-Yan Z+jets processes. A real Z boson decaying into $\ell^+\ell^-$ and a number of high p_T jets may easily yield signal-like combinations, even though jets do not stem from a Z boson. The $M_{\ell\ell jj}$ distribution of these events is essentially a falling exponential above 220 GeV/c^2. Below that value, acceptance effects give rise to a steep edge difficult to reproduce with simulations.

To estimate the background from Z+jets and the shape of its $M_{\ell\ell jj}$ distribution, large exclusive simulated samples of Z + n jets ($n = 1$ to 4) events are used, corresponding to integrated luminosities ranging from 36 fb^{-1} to 232 fb^{-1} (appendix A). These samples have been thoroughly verified to reproduce the data in the m_{jj} sideband control region, and in the signal region using the unblinded ICHEP data (for all observables but $M_{\ell\ell jj}$). The small discrepancies found are attributed to a bad modeling of the p_T spectrum of the $\ell^+\ell^-jj$ system, the so-called p_T^H

Table 3: Expected and observed yields with 5.2 fb^{-1} of data. The expected background is composed of p_T-weighted simulated Z+jets, data-driven top+X and diboson MC.

	0 btag		1 btag		2 btag	
	$\mu^+\mu^- jj$	$e^+e^- jj$	$\mu^+\mu^- jj$	$e^+e^- jj$	$\mu^+\mu^- jj$	$e^+e^- jj$
expected background	5658	5143	1971	1784	190	156
observed data	5818	4972	1825	1675	193	172
M_H (GeV/c^2)	signal expectation					
250	31.5	28.6	15.8	14.5	5.1	4.8
300	33.7	29.9	18.0	15.3	6.6	5.6
400	32.6	28.5	18.2	16.1	7.3	6.4
500	15.3	13.8	8.9	7.9	3.7	3.3
600	5.7	5.4	3.5	3.1	1.4	1.3

Table 4: Expected and observed (blinded) yields with 19.6 fb^{-1} of data. The expected background is composed of p_T-weighted simulated Z+jets, data-driven top+X and diboson MC.

	0 btag		1 btag		2 btag	
	$\mu^+\mu^- jj$	$e^+e^- jj$	$\mu^+\mu^- jj$	$e^+e^- jj$	$\mu^+\mu^- jj$	$e^+e^- jj$
expected background	21265	19342	7604	6700	731	666
observed data	–	–	–	–	–	–
M_H (GeV/c^2)	signal expectation					
250	118.0	106.9	58.9	55.2	19.0	17.9
300	126.7	119.1	67.1	57.7	24.7	21.1
400	121.9	107.2	68.5	60.7	27.3	23.9
500	57.2	52.7	33.0	29.9	13.7	12.3
600	21.4	19.3	13.0	12.0	5.3	4.8

observable.

While there are differences between data and simulations (Figure 6 (left)), the mismodeling of the p_T spectrum of the $\ell^+\ell^- jj$ system in the simulation is the same in the signal and sideband regions. A weight calculated as the ratio of the p_T distributions in data over simulations in the sideband region, and fitted to the function

$$f(p_T) = \left(1 + \frac{1}{a + bp_T^2}\right) \frac{1}{e^{-p_T/c} + 1}$$

as function of p_T (Figure 6 (right)), is used to correct the Z+jets distributions in an event-by-event basis. To get a fair weight function, contamination from diboson events (obtained from simulation) and $t\bar{t}$ events (from data, see next section) is subtracted from data. Figure 7 depicts the fitted functions for different b-tag categories and their 1σ bands. The correction calculated with the inclusive sample, without distinguishing among b-tag categories, used in the analysis agrees within errors with the fit results obtained independently for each b-tag category.

In the final analysis, the Z+jets background is extracted from simulation, corrected using the p_T weight. Its normalization is constrained to the relative normalization of diboson+$t\bar{t}$ background subtracted data in the m_{jj} sideband region, independently for each btag category.

Determination of $t\bar{t}$ background from data

The $t\bar{t}$ background is an important source of contamination in the 2-btag category. It is estimated from the data using $e^{\pm}\mu^{\mp}$ events passing the same cuts as the signal. This method

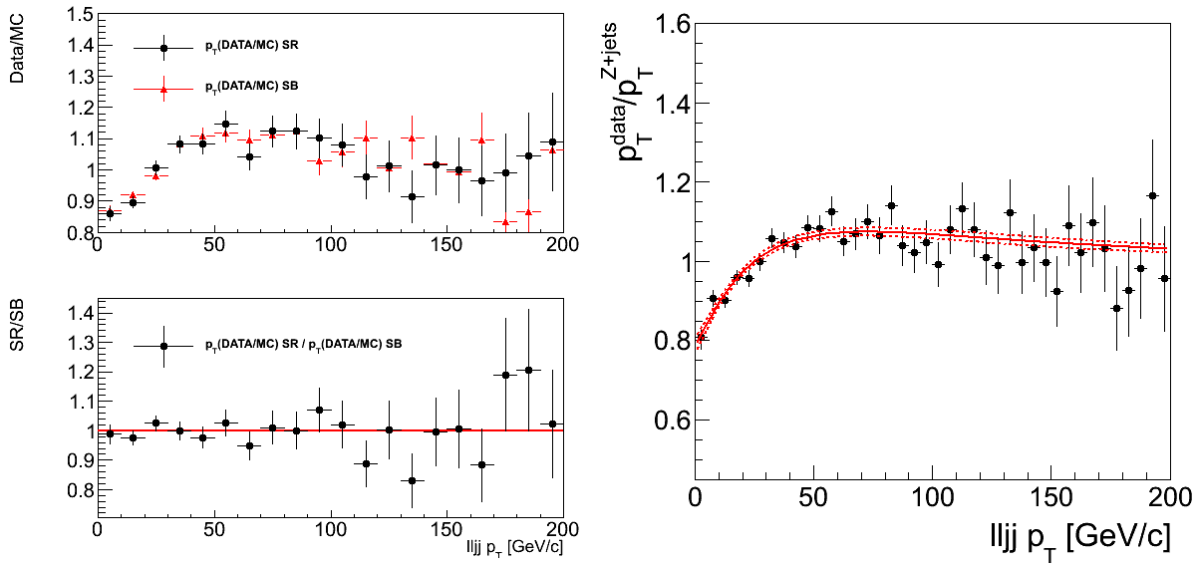


Figure 6: Left-top: data over Z+jets ratio of the p_{T} distributions of the $\ell^+\ell^- jj$ system in the m_{jj} sideband (red triangles) and signal (black dots) regions. In the signal region, unblinded ICHEP data are used. In the sideband, the full dataset is utilized. Left-bottom: ratio of the ratios above. Right: ratio of the data over simulation p_{T} distributions together with the fit to $f(x)$ (red solid line). The red dashed lines are the $\pm 1\sigma$ statistical error bands calculated propagating the full correlation matrix of the fit.

accounts for other small backgrounds (as WW + jets, $Z \to \tau^+\tau^-$ + jets, single top, fakes) where the lepton flavor symmetry can be invoked as well.

In this study we use the Powheg + Pythia $t\bar{t} \to 2\ell 2\nu + X$ Monte Carlo sample. Assuming a cross-section of 23.38 pb for this process, the top MC is normalized to the data for events with $M_{e^\pm\mu^\mp} > 50$ GeV/c^2 using a K-factor of 1.05. Other top MC samples as the Madgraph $t\bar{t}$ inclusive sample produce consistent results, but different K-factors are needed.

Top-pair Monte Carlo studies show that the $e^\pm\mu^\mp$ vs. $e^+e^- + \mu^+\mu^-$ symmetry works very well at the level of the shapes of the distributions of all considered variables. Also, the relative event normalization is consistent with one, within the MC finite statistical errors. For instance, in the case of Powheg + Pythia top MC, the $e^\pm\mu^\mp/(e^+e^- + \mu^+\mu^-)$ relative event normalizations are 1.007 ± 0.007 after selection and kinematical cuts, and 1.01 ± 0.01 after b-tagging.

Figure 8 shows a comparison of the $e^+e^- + \mu^+\mu^-$ and $e^\pm\mu^\mp$ top MC distributions of two relevant variables, for events with at least two leptons and two jets passing selection cuts. More distributions are available in appendix F. The selection step is specified in each plot. In this studt the category "$\geq$ 1 b-tag" includes events with at least one jet tagged using the JPM prescription. The normalization is arbitrary.

The 2012 $e^\pm\mu^\mp$ data yields are compared to the sum of top MC prediction and other small backgrounds in Table 5, while distributions of relevant variables are shown in Figure 9 and appendix F. Events selected contain at least two leptons and two jets passing selection cuts. Only the hardest-$\sum P_T$ dilepton combination and the dijet combination with largest JP discriminator values are considered. Pile-up corrections have been applied. Other extra cuts are detailed where appropriate.

The table and figure above include an estimation of WW, $Z \to \tau^+\tau^-$, and single top contribu-

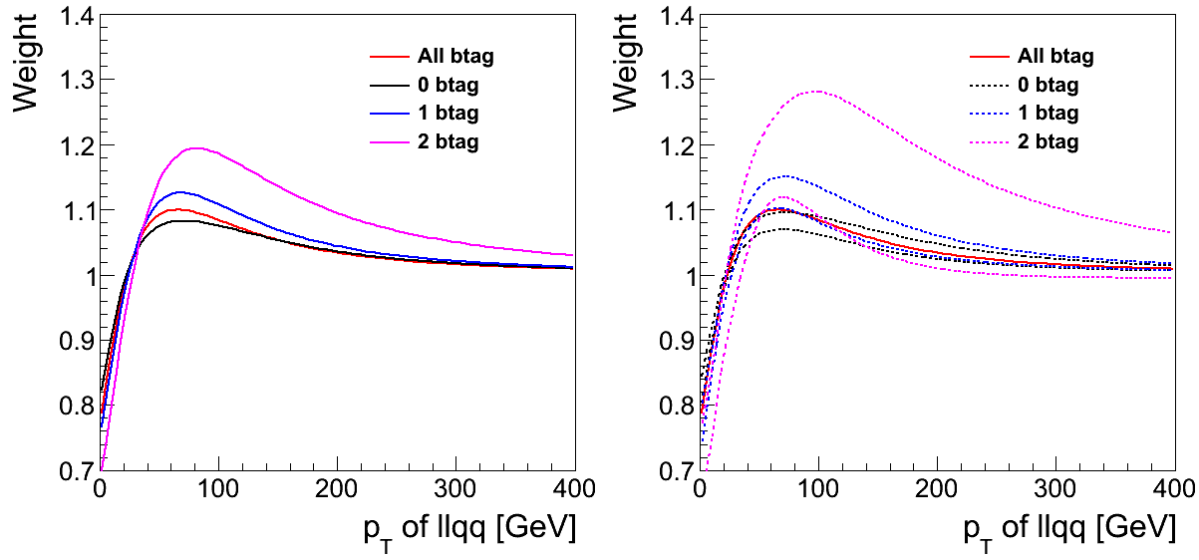


Figure 7: Fits to the ratios of the data over simulation p_{T} distributions of the $\ell^+\ell^- jj$ system in the different b-tag categories and their combination (left) and their corresponding 1σ bands (right).

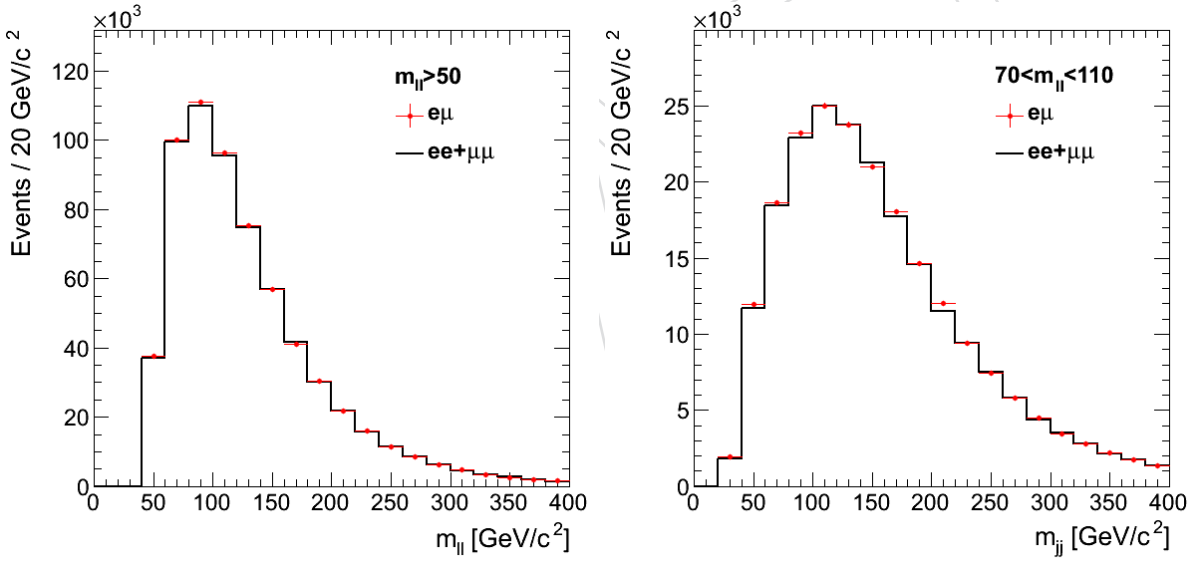


Figure 8: Powheg + Pythia top MC $\mathrm{e}^{\pm}\mu^{\mp}$ to $(\mathrm{e}^+\mathrm{e}^- + \mu^+\mu^-)$ comparison for two variables after different steps of the selection, as specified in the legends: dilepton invariant mass (left) and dijet invariant mass (right).

tions from Monte Carlo. The fake component is estimated from $\mathrm{e}^{\pm}\mu^{\mp}$ data; the yield of events with one or two non-isolated leptons (in the combined relative isolation region 0.25 - 0.85), is extrapolated into the isolated lepton region assuming a flat distribution in the combined relative isolation variable. Changing the size of the non-isolation region changes the fake prediction by at most 10%. The e-μ symmetry holds in reasonable approximation for the non-isolated lepton data.

The sample composition before b-tagging is 86% $\mathrm{t\bar{t}}$, 5% fakes, and 9% other small backgrounds. After requiring 1 JPL b-tag (1 JPM and 1 JPL b-tags) the relative fractions change to 91%(95%)

Cuts	Top MC	Total MC	$e\mu$ data
$M_{ll} > 50\ \mathrm{GeV}/c^2$	26685.0	30238.6	30240
$70\ \mathrm{GeV}/c^2 < M_{ll} < 110\ \mathrm{GeV}/c^2$	9108.0	10403.8	10308
$70\ \mathrm{GeV}/c^2 < M_{jj} < 110\ \mathrm{GeV}/c^2$	1939.2	2261.4	2277
≥ 1 JPL b-tagged jet	1766.5	1941.6	1961
≥ 1 JPM b-tagged jet	1431.1	1545.1	1562
≥ 1 JPL & ≥ 1 JPM	873.7	922.4	950
1 JPL + 1 JPM, MET Sig < 10	357.5	381.9	406

Table 5: Comparison of 2012 $e\mu$ data to Powheg + Pythia top MC event yields, corresponding to an integrated luminosity of 19.6 fb^{-1}. "Total MC" contains the top, WW, $Z \to \tau^+\tau^-$, single top, and fakes contributions. Every cut in a line assumes all cuts in lines above.

$t\bar{t}$, 4%(2%) fakes, and 5%(3%) other small backgrounds.

Now, we test the $e^{\pm}\mu^{\mp}$ vs. $e^+e^- + \mu^+\mu^-$ symmetry using a top-enriched subsample of the data. Figure 10 (left) shows the MET significance distribution after requiring 1 JPM b-tag For values sufficiently large of MET significance the number of events of the $e^{\pm}\mu^{\mp}$ and $e^+e^- + \mu^+\mu^-$ samples are equal within statistical errors. In order to test the agreeement in shape, Figure 10 (right) contains the "Higgs" invariant mass distributions after requiring 2 (1 JPM + 1 JPL) btags, MET significance > 8, and $|M_{\ell^+\ell^-} - M_Z| > 20\ \mathrm{GeV}/c^2$. One can observe agreement on both the normalization and shape of the $e^{\pm}\mu^{\mp}$ and $e^+e^- + \mu^+\mu^-$ distributions. Appendix F has additional supporting figures for two btags (1 JPM and 1 JPL).

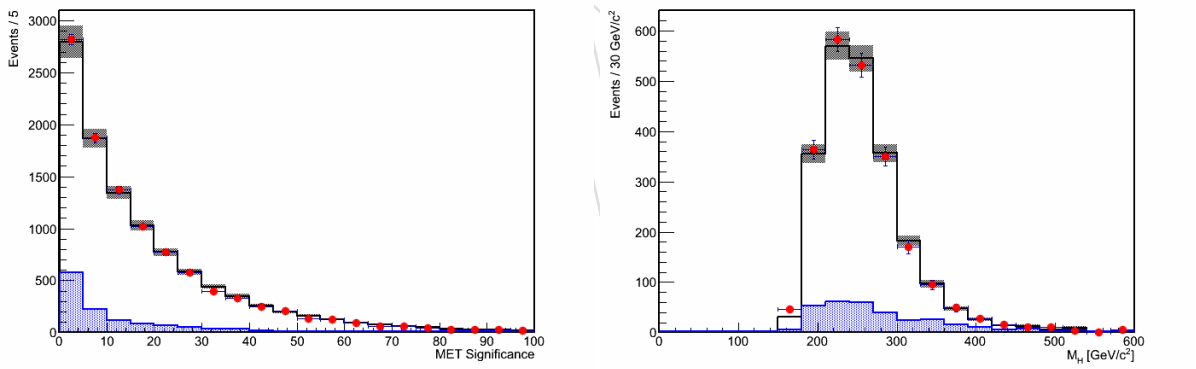


Figure 9: Comparison of 2012 $e^{\pm}\mu^{\mp}$ data to Powheg + Pythia top MC, corresponding to an integrated luminosity of 19.6 fb^{-1}. Red dots are $e^{\pm}\mu^{\mp}$ data; white histogram top Monte Carlo; blue histogram other small backgrounds. MET significance (left) and "Higgs" invariant mass (right).

6 Systematics

In this section the systematic uncertainties affecting the analysis, the method used to estimate them and their estimated values are described.

Luminosity uncertainty

The latest recommendation for the 2012 datasamples is the uncertainty on LHC luminosity of 4.4% [31].

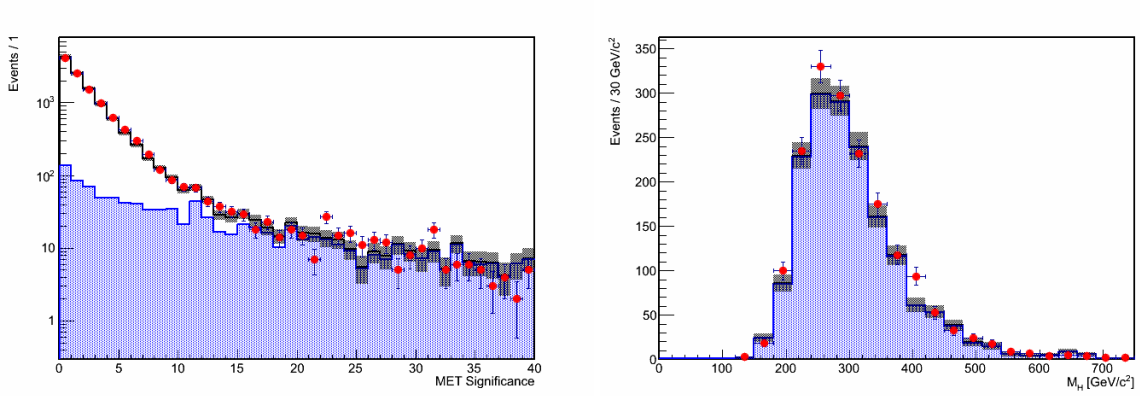


Figure 10: (Left) MET significance distribution for dilepton data compared to the sum of Drell-Yan Monte Carlo plus $e^{\pm}\mu^{\mp}$ data for events with 1 JPM b-tag. (Right) "Higgs" invariant mass (right) for $e^+e^- + \mu^+\mu^-$ and $e^{\pm}\mu^{\mp}$ data for events outside the leptonic Z mass window, with two (1 JPM + 1 JPL) b-tags, and MET significance > 8. Other cuts are detailed in the text. Red dots are $e^+e^- + \mu^+\mu^-$ data; white histogram Drell Yan Monte Carlo; blue histogram $e^{\pm}\mu^{\mp}$ data (plus other small backgrounds).

Higgs cross-section and branching fractions

The Higgs production cross-section uncertainty depends on production mechanism, either gluon fusion or weak boson fusion (WBF). However, since the gluon fusion mechanism dominates, it drives the total uncertainty. We use gg and WBF errors separately and for each mass point according to Yellow Report prescription. The total weighted error is in the range 13.4–18.0%. We note that this uncertainty is relevant only for the measurement of the ratio to SM expectation R, while it does not affect the absolute cross-section measurement.

Uncertainties in the signal expectation

The main systematic uncertainties on signal normalization are summarized in Table 6, and are discussed in more detail in the subsection below.

Table 6: Summary of systematic uncertainties on signal normalization. Most sources are multiplicative errors on the cross-section measurement, except for expected Higgs cross-section (which is relevant for the measurement of the ratio to SM expectation R).

source	0 b-tag	1 b-tag	2 b-tag	comment
muon trigger & reco	1.1%			tag-&-probe study
electron trigger & reco	3.6%			tag-&-probe study
jet reco	1%–5%			JES & JER uncertainty negligible, correlated between categories
b-tagging	1-4%	1-5%	5-8%	anti-correlated between categegories
MET	$< 1\%$			loose requirement
pile-up	1-2%			correlated between categories
production mechanism (PDF)	1-4%			PDF4LHC, acceptance only
production mechanism (lineshape)	0-3%			only for $M_H > 400$ GeV/c^2
luminosity	4.4%			same for all analyses
Higgs cross-section (for R)	13–18%			detailed table from YR available

Lepton energy scale, resolution, selection, and trigger

Lepton trigger and selection is common among several $H \to ZZ$ analyses and we benefit from common study based on tag-and-probe techniques. In particular, recent studies within the framework of Ref. [18] indicate systematics of 1.0% due trigger, 0.5(3.3)% due to muon (electron) identification, 0.2(0.8)% due to muon (electron) isolation mostly independent of the mass hypothesis, 0.1(0.5)% due to muon (electron) momentum/energy scale.

Jet Energy Scale and Resolution

The main uncertainty in jet reconstruction comes from jet energy scale (JES) uncertainty, while the uncertainty on the resolution contributes a much negligible effect to the total uncertainty. Our estimates show that JES variation by $\pm 1\sigma$ changes reconstruction efficiency of a 400 GeV Higgs by about 1.2%. The effect on the jets transverse momentum and dijet invariant mass is sizable and it drives the bias on the acceptance. The change in selection efficiency as function of Higgs mass is summarized in table 6

Table 7: Summary of signal efficiency changes due to systematic uncertainties on the jet energy scale.

M_H GeV	JES $+1\sigma$	JES -1σ
230	4.26%	-4.17%
300	1.33%	-1.30%
400	1.23%	-1.20%
600	-0.79%	0.86%
800	-1.08%	1.10%
1000	-1.16%	1.21%

Pileup

The number of true interactions per bunch crossing in the simulated samples was re-weighed to match the distributions in data. The main source of systematics may come from the uncertainty on the measurement of the amount of pileups in data. This uncertainty is studied by re-estimating the number of true interactions in data with different values of minimum-bias cross section as input, using 65.84 mb and 72.77 mb as recommended by the CMS pileup group [32], which is $\pm 5\%$ difference with respect to the central value 69.30 mb. The re-estimated distributions and the central value are compared in Figure 11.

The number of true interactions are weighted in the MC to match the shifted distributions in data in Figure 11 and re-compute the signal efficiency. This leads to a change in the signal efficiency $< 1\%$ for the 0- and 1-btag categories, and about 2% for the 2-btag category, for $M_{\mathrm{H}} < 650$ GeV/c^2, approximately independent of lepton channel, as detailed in appendix G .

Heavy quark flavor tagging uncertainty

A data-to-Monte Carlo scale factor (SF_b) has been measured for events containing b-jets as a function of p_T and η for the jets. This SF_b corrects for the more efficient identification of b-jets in Monte Carlo compared to data. Likewise, a mistag rate scale factor (SF_{mistag}) for light quarks misreconstructed as b-jets has been measured over a range of p_T and η for the jets. To study the systematic effects of b-tagging, both the SF_b and SF_{mistag} were simultaneously varied up and down by the uncertainty related to each SF.

The study was performed separately for the muon and electron channels, calculating the effect for signal MC.

Tables in appendix G give the b-tagging systematic uncertainty for the signal, for muons and electrons. The systematic effect is computed in final regions moving the SF by plus and minus its uncertainty to the number of tagged jets with the nominal SF. The uncertainty is reported (in %) for the cases where both jets are tagged, at least 1 jet is tagged, and no jets are tagged. In the analysis, the exact systematic uncertainty as a function of the Higgs mass is applied.

MET uncertainty

The dominant effects are from the knowledge of the rest of the event, such as jet energy reconstruction and pileup. Therefore, both of the above subsections cover MET uncertainty to a large extent. Additionally we investigated how much the MET rescaling procedure described in section 4 affects the signal selection efficiency, by counting the number of signal events migrating over the MET threshold due to the scaling procedure. The requirement on the MET significance translates thus into about $\sim 0.5\%$ uncertainty on the final efficiency.

Production mechanism

The expected kinematics of the Higgs production is subject to uncertainties due to limited knowledge of the underlying parton distribution functions (PDFs) as well as the shortcomings in the theoretical prediction (missing higher orders in the perturbation series). These uncertainties are propagated to an uncertainty on the selection acceptance and efficiency. Their additional effect on the Higgs production cross section is discussed in a separate section below.

The PDF uncertainties is evaluated according to the PDF4LHC recommendations, by evaluating the selection efficiency for the PDF sets cteq66 [33], MSTW2008NLO [34] and NNPDF2.1 [35] and their error sets. The envelope of the various PDF sets is used as the total uncertainty, as recommended and amounts to 1-4%. The uncertainty noticeably increases for very high Higgs masses. A summary of systematic uncertainties on the signal acceptance following PDF4LHC recommendations can be found in table **??**.

Additional uncertainties arise due to uncertainties on the Higgs signal shape that is theoretically calculated. The shape uncertainty is evaluated in the recommended way for the Higgs decaying into a pair of Z boson, correctly accounting for the correct lineshape (i.e. reweighting

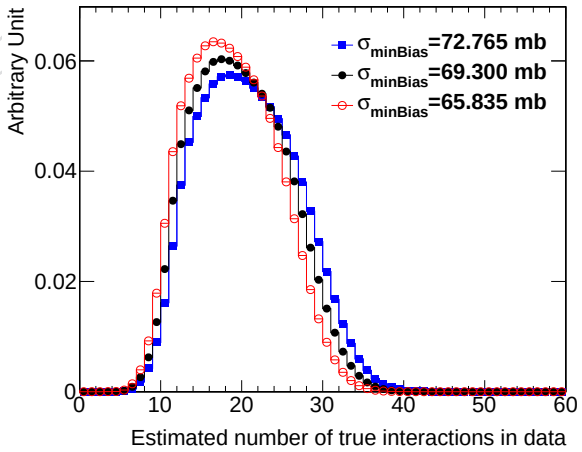


Figure 11: Estimated number of true interactions in 2012 data, assuming different values of minimum-bias cross section. The central value is 69.3 mb (solid circles).

of the given shape in POWHEG) which also accounts for interference effects. The full description of the reweighting and uncertainty method is given in [36] and provides an uncertainty that contributes in two ways: Due to the mass-dependence of the selection efficiency, the total signal effiency is affected by the line shape. The uncertainty is negligible below 400 GeV and rises to $\sim$ 3% at 600 GeV, with only small dependence on btag category.

Additionally the line-shape used in the CLs procedure is re-extracted with the alternative line-shape models (Figure 12 (left)). The tail caused by mismatched jets is not affected at all as it is a random mixture of events, averaging out any shifts from the uncertainty. The core of the signal distribution is only weakly affected byt the uncertainty. In the worst case (the highest mass we consider), the peak-position shifts by ~ 2 GeV/c^2 (compared to a sigma of 60 GeV/c^2) and the sigma changes by ~ 1 GeV/c^2. Due the minuscule effect of this uncertainty, it is not propagated further.

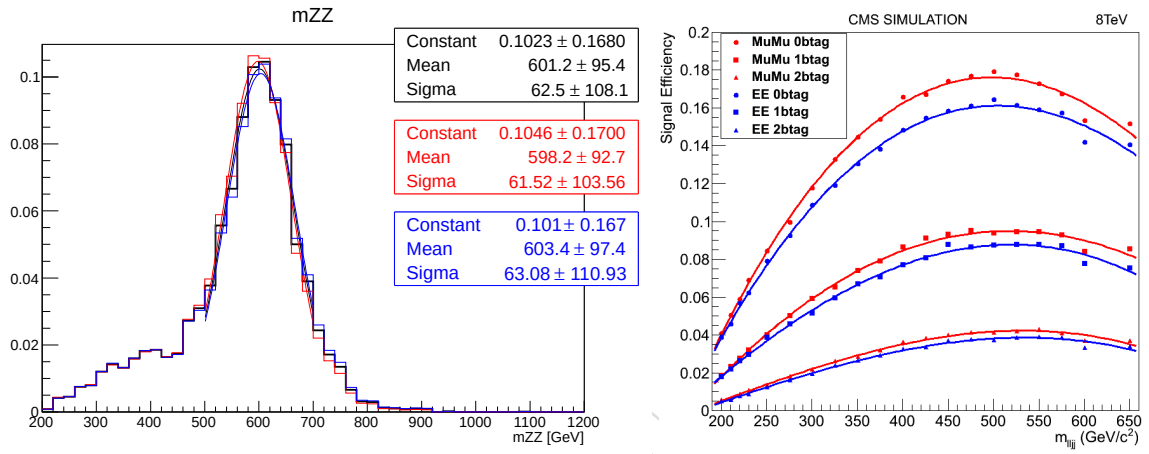


Figure 12: Left: reconstructed $M_{\mathrm{H}} = 600\,\mathrm{GeV}/c^2$ Higgs signal (area normalized) with the nominal lineshape (black) and systematic variations (blue/red). Gaussian fits to the core of the distribution are overlaid. Right: parameterization of the signal efficiencies, as function of the Higgs mass hypothesis, in the three btag categories, in the muon and electron channels.

Background systematic uncertainties

The effect of the main systematic uncertainties affecting the background normalization and shape are summarized in Table 9. The effects are similar to those affecting the signal.

The impact on the shape of the $M_{\ell\ell jj}$ distribution for background events coming from the mass-

Table 8: Summary of systematic uncertainties on the signal acceptance following PDF4LHC recommendations.

PDF	$M_H = 200$ GeV	$M_H = 400$ GeV	$M_H = 600$ GeV	$M_H = 800$ GeV	$M_H = 1000$ GeV
cteq66	$^{+0.6}_{-0.7}$	$^{+0.8}_{-1.0}$	$^{+0.8}_{-1.1}$	$^{+1.5}_{-2.0}$	$^{+2.6}_{-3.2}$
MSTW2008NLO	$^{-0.2}_{-0.5}$	$^{+0.6}_{+0.2}$	$^{+0.8}_{+0.4}$	$^{+1.5}_{+0.7}$	$^{+2.5}_{+1.2}$
NNPDF2.1	$^{+0.8}_{+0.2}$	$^{+1.4}_{+0.75}$	$^{+1.5}_{+0.9}$	$^{+2.7}_{+1.4}$	$^{+4.3}_{2.4}$
Total	$^{+0.8}_{-0.7}$	$^{+1.4}_{-0.8}$	$^{+1.5}_{-1.1}$	$^{+2.7}_{-2.0}$	$^{+4.3}_{-3.2}$

Table 9: Summary of systematic uncertainties on the background determination.

source	0 b-tag	1 b-tag	2 b-tag
muon trigger & reco	2%		
muon momentum scale	0-1%		
electron trigger & reco	2%		
electron energy scale	0-2%		
jet energy scale	0-5%		
b-tagging	0-1%	0-3%	0-5%
MET	0.5%		
pile-up	0-1%		
p_T^{llqq} weighting	0-3%		
Diboson cross section	15%		
luminosity	4.4%		

dependent systematic uncertainties (electron energy scale, muon momentum scale, jet energy scale, pile-up and $p_{\mathrm{T}}^{\ell\ell jj}$-based weigthing) is depicted in Figure 13. The relative variation ranges from 1% at $M_{\ell\ell jj} \simeq 200\,\mathrm{GeV}/c^2$ to about 3% at around 600 GeV/c^2, as displayed in Figure 14.

7 Statistical analysis and results

Signal determination

The signal efficiency of the selection described in section 4 is evaluated as the ratio between the number of selected events in each of the six channels (electron and muon channels, 0-, 1-, and 2-btag categories) under study and the total number of generated events in the Monte-Carlo samples. The signal efficiency as a function of the Higgs mass is fitted to a polinomium in order to be estimatated for those Higgs mass hypothesis where no Monte-Carlo sample is available, as shown in Figure 12 (right).

The narrow width approximation used in the 2011 analysis breaks down at high Higgs mass (typically $>$ 400 GeV/c^2) due to the very large Higgs width ($>$ 70 GeV/c^2). The problem has been discussed in details in Ref. [37] and a more correct approach to describe the Higgs invariant-mass distribution has been proposed, known as Complex Pole Scheme (CPS). The total Higgs production cross-section has been recomputed by the Higgs Cross-Section Working Group to include corrections due to CPS at high Higgs mass [38]. In the 2011 [4] and 2012 [39] published analysis, CPS effects were included in the cross section calculation, but neglected for the signal shape (covered by an appropriate uncertainty). In this analysis we properly reweight the simulated signal samples to follow the CPS.

At high Higgs mass the interference between the Higgs signal and the $gg \to ZZ$ background becomes large, as recently discussed in Ref. [40]. The effect of interference has been shown to be constructive below the Higgs mass peak and destructive above. It has therefore a negligible effect on the total cross-section (1-2%) but it biases the ZZ invariant-mass distribution. Moreover the interference has been computed only at LO while the signal is known at NNLO. In this analysis we follow the approach proposed in Ref. [40] to estimate the uncertainty due to missing higher perturbative order on the interference and the simulated line shape is reqeighted accordingly.

After this is done, the obtained distribution after all the selection criteria should reflect the expectation for the Higgs events, both in yield and shape. This reweighting has been included

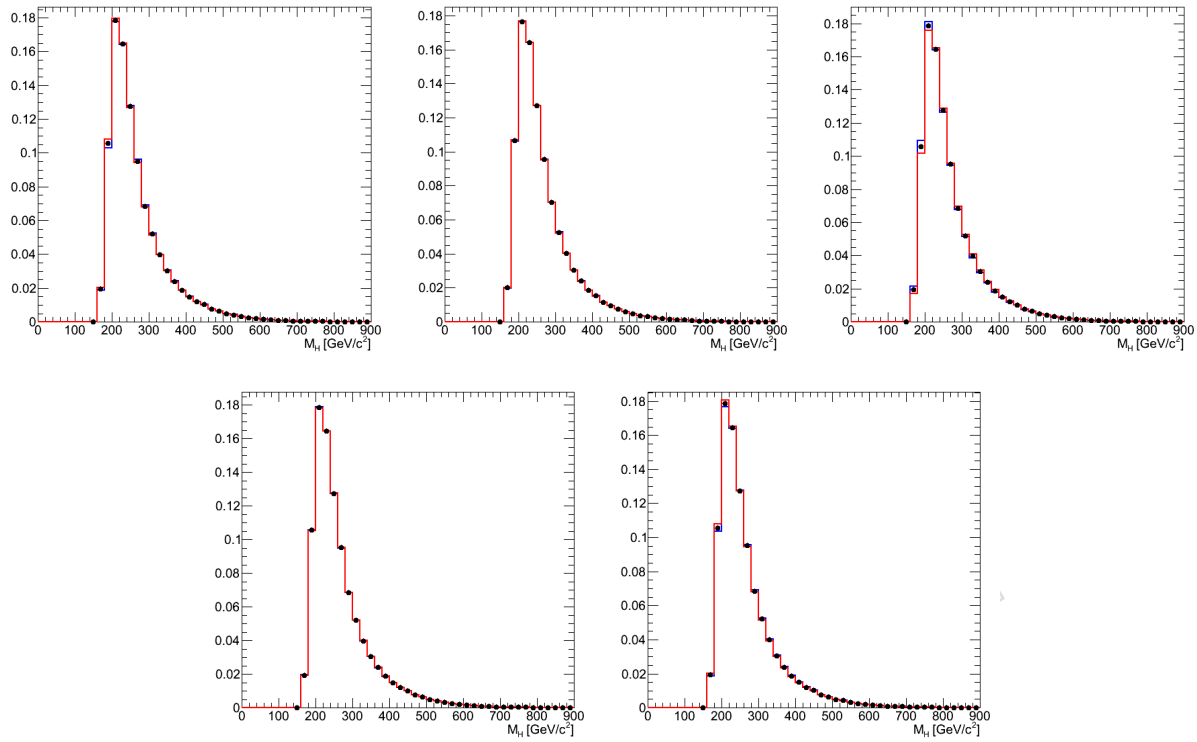


Figure 13: Shape variation of the $M_{\ell\ell jj}$ distribution for Z+jets simulated events when varying the systematic uncertainties (from top-left to bottom-right): electron energy scale, muon momentum scale, jet energy scale, pile-up and $p_{\mathrm{T}}^{\ell\ell jj}$-based weigthing.

in the analysis in order to account for the correct distributions.

Results

Expected upper limits to the Standard Model Higgs production cross section are determined as a function of the Higgs boson mass. The official tool developed by the CMS Higgs combination is used [41] The tool uses the `RooStats` [42] engine from ROOT as internal implementation. and the "asymptotic CLs" [43] method is used.

The limit is expressed as ratio r of the upper limit to the cross section times branching fraction to $\ell\ell q\bar{q}$ divided by its standard model expectation. A value of the Higgs boson mass M_{H} is excluded if, for that mass hypothesis, r is less than one.

Since the decay products of the Higgs boson can be fully reconstructed, it is possible to use the reconstructed higgs boson mass, $M_{\ell\ell jj}$, distribution to discriminate Higgs signal events against background events. The reconstructed mass will peak around the true Higgs boson mass, M_{H}, for the signal, while for background processes it will have a broader distribution. Consequently, a shape-based treatment of the expected and observed distribution of the invariant mass of the Higgs candidate (i.e. the 4-object mass) increases the sensitivity of the analysis, compared to a simpler counting experiment. Since the expected shape of the background and the signal cannot be obtained from first principles, an analytical funcion cannot be used. Therefore, a binned histogram-based calculation has been used for the shape analysis.

The normalization of the simulated background (Z+jets and diboson) in the signal region is

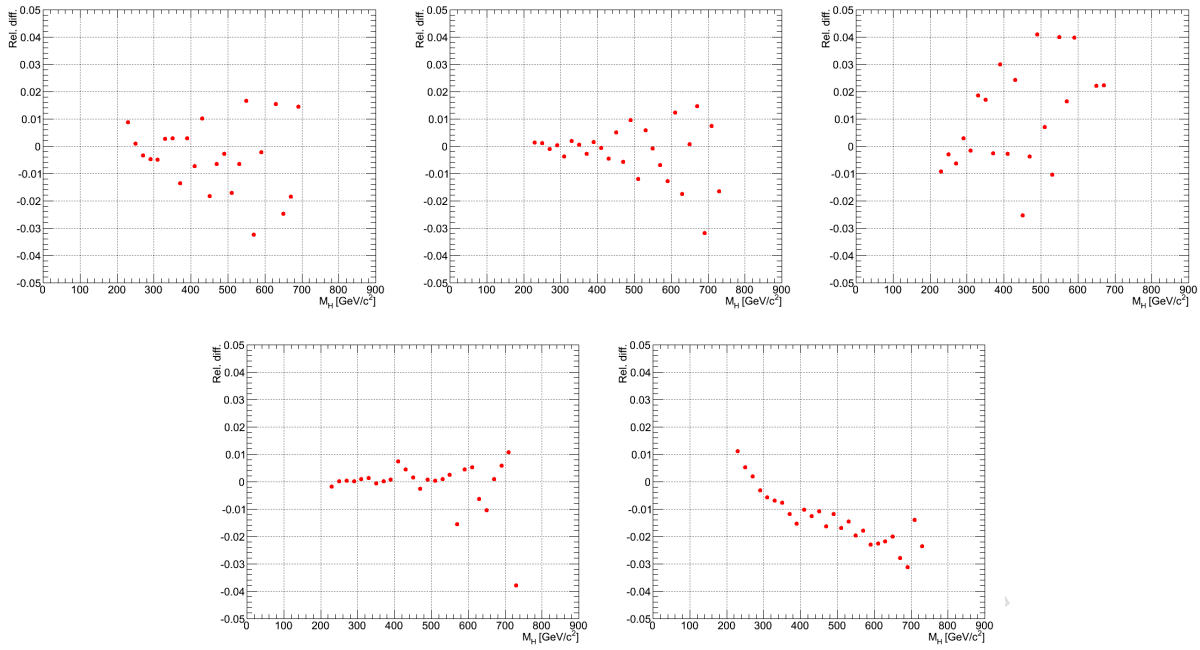


Figure 14: Relative difference on the $M_{\ell\ell jj}$ distribution for Z+jets simulated events when varying the systematic uncertainties (from top-left to bottom-right): electron energy scale, muon momentum scale, jet energy scale, pile-up and $p_{\mathrm{T}}^{\ell\ell jj}$-based weigthing.

allowed to vary, the actual contraint coming from the number of events is the m_{jj} sideband. The observed and expected numbers of events in the sideband region are given as input to the limit calculation tool. The calculation to discriminate signal from background events is performed independetly for the six individual analysis channels (electrons and muons in the 0, 1, and 2-btag categories), and then combined taking into account the correlations among the systematic uncertainties. These uncertainties may affect either the shape of the distributions or their normalization, and are properly taken into account in the statistical analysis.

The mass distributions of the $\ell^+\ell^- jj$ system are depicted in Figure 15 for ICHEP data in the signal region. They include the systematic background shape uncertainty and the uncertainty on the background normalization to the data on the sideband. The full dataset (19.6 fb^{-1}) is used to assess the good modelling of the background in the m_{jj} sideband region (Figure 16).

As a cross check, we have studied the $M_{\ell\ell jj}$ distributions for the electron and muon channels separatelly, both in the m_{jj} sideband and signal regions (appendix H). They show an excellent agreement.

The possible sources of systematic uncertainties on both the expected signal and background yields and the shape of signal and background distributions are considered in the limit extraction procedure as nuisance parameters.

Based on the background and signal expectations and their correspondent uncertainties, the observed limit on the ratio of the SM Higgs boson production cross section to the SM Higgs expectation is determined between 230 GeV/c^2 and 650 GeV/c^2. The $M_{\ell\ell jj}$ region below 200 GeV/c^2 presents a very sharp rising edge difficult to have under control and is excluded from the analysis.

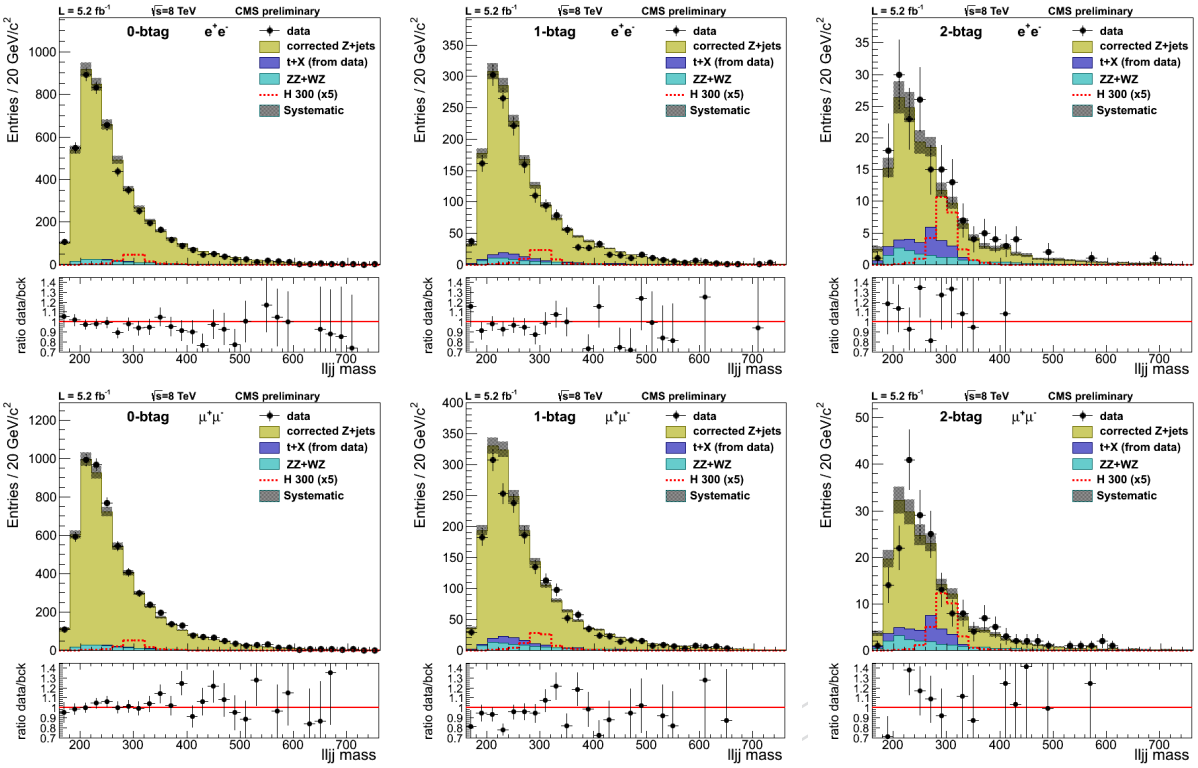


Figure 15: Mass distributions of the $\ell^+\ell^- jj$ system for events in the signal region in the electron (top) and muon (channels). From left to right, plots correspond to the 0-, 1-, and 2-btag categories. The dots are unblinded ICHEP data, brownish histogram corrected Z+jets simulation, light blue simulated diboson background and dark blue $t\bar{t}$ events from data (which include single top, WW, $Z \to \tau^+\tau^-$+jets). The grey band indicates the systematic background shape uncertainty and the uncertainty on the background normalization to the data on the sideband.

Figure 17 show the expected and observed limit using the ICHEP dataset in order to get and estimation on how the results look like for the unblided set of the data. Results were cross-checked with an independent statistical method (see appendix I).

Plot in Figure 18 shows the expected limit for the full dataset recorded during 2012 at 8TeV, corresponding to 19.6 fb^{-1}. The limit is shown as a ratio to the SM predicted cross section as a function of the Higgs mass. With the increased luminosity in in the full 2012 dataset, the 2-btag category becomes the most powerful contribution to the combination of the six channels (see Figure 19).

8 Conclusions

We have performed a search for a SM-like Higgs boson with a mass between 230 GeV/c^2 and 650 GeV/c^2 in the decay $H \to ZZ \to \ell^+\ell^- q\bar{q}$, with the Z boson subsequently decaying to $Z \to \ell^+\ell^-$ and $Z \to$, using of data collected in summer 2012 from LHC proton-proton collisions at a centre-of-mass energy of 8 TeV. The analysis, using the first 5fb^{-1} of data recorded between March and June, excludes the existence of a hypothetical standard model Higgs boson in the mass range between 269 GeV/c^2 to 304 GeV/c^2 and between 360 GeV/c^2 to 555 GeV/c^2.

This analysis aims at analyzing the full 2012 data coprresponding to an integrated luminosity

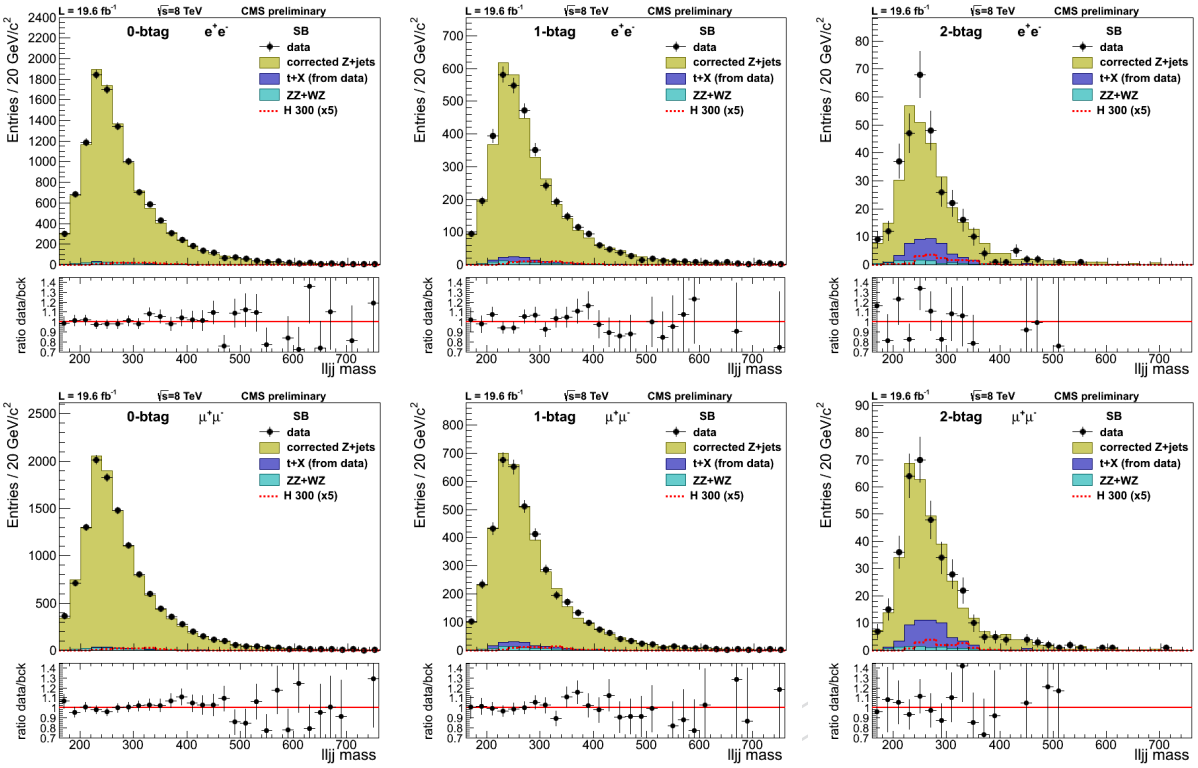


Figure 16: Mass distributions of the $\ell^+\ell^- jj$ system for events in the m_{jj} sideband region in the electron (top) and muon (channels). From left to right, plots correspond to the 0-, 1-, and 2-btag categories. The dots are data (19.6 fb^{-1}), brownish histogram corrected Z+jets simulation, light blue simulated diboson background and dark blue $\bar{\text{t}}\text{t}$ events from data (which include single top, WW, $\text{Z} \to \tau^+\tau^-$+jets).

of 19.6 fb^{-1}, after unblinding is granted by the Higgs PAG. The expected exclusion range is set using the full dataset between 230 GeV/c^2 and 650 GeV/c^2.

References

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

[2] ATLAS Collaboration, "Observation of a new particle in the search for the Standard Model Higgs boson with the ATLAS detector at the LHC", *Phys.Lett.* **B716** (2012) 1–29, doi:10.1016/j.physletb.2012.08.020, arXiv:1207.7214.

[3] T. Aaltonen et al, The CDF, D0 Collaborations, the TEVNPHWG Working Group, "Combined CDF and D0 Upper Limits on Standard Model Higgs Boson Production with up to 8.2 fb$^-$1 of Data", *FERMILAB-CONF-11-044-E* (2011) arXiv:1103.3233.

[4] CMS Collaboration, "Search for a Higgs boson in the decay channel $H \to ZZ^{(\star)} \to q\bar{q}l^-l^+$", *JHEP* **04** (2012) 036, doi:10.1007/JHEP04(2012)036, arXiv:1202.1416.

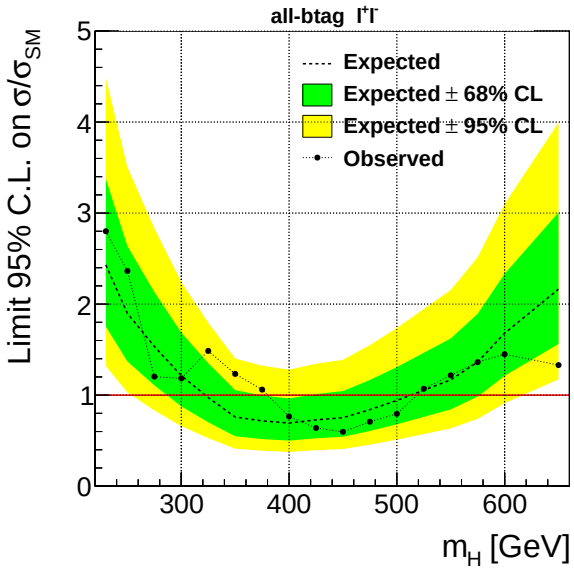


Figure 17: Observed (solid) and expected (dashed) 95% CL upper limit on the ratio of the production cross section to the SM expectation for the Higgs boson obtained using the CL_s technique. The 68% and 95% ranges of expectation for the background-only model are also shown with green and yellow bands, respectively. The solid line at 1 indicates the expectation for a SM-Higgs-like boson. These results correspond to a check performed on the unblinded part of the dataset (ICHEP data) to verify everything is under control.

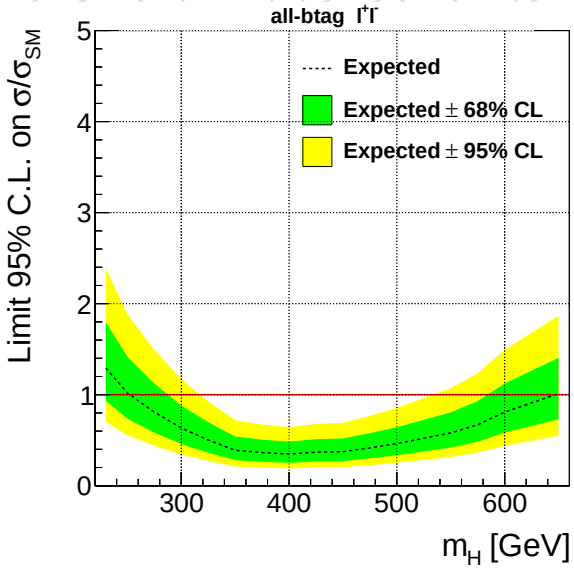


Figure 18: Observed (solid) and expected (dashed) 95% CL upper limit on the ratio of the production cross section to the SM expectation for the Higgs boson obtained using the CL_s technique. The 68% and 95% ranges of expectation for the background-only model are also shown with green and yellow bands, respectively. The solid line at 1 indicates the expectation for a SM-Higgs-like boson. [PLOT TO BE CHANGED AFTER UNBLINDING]

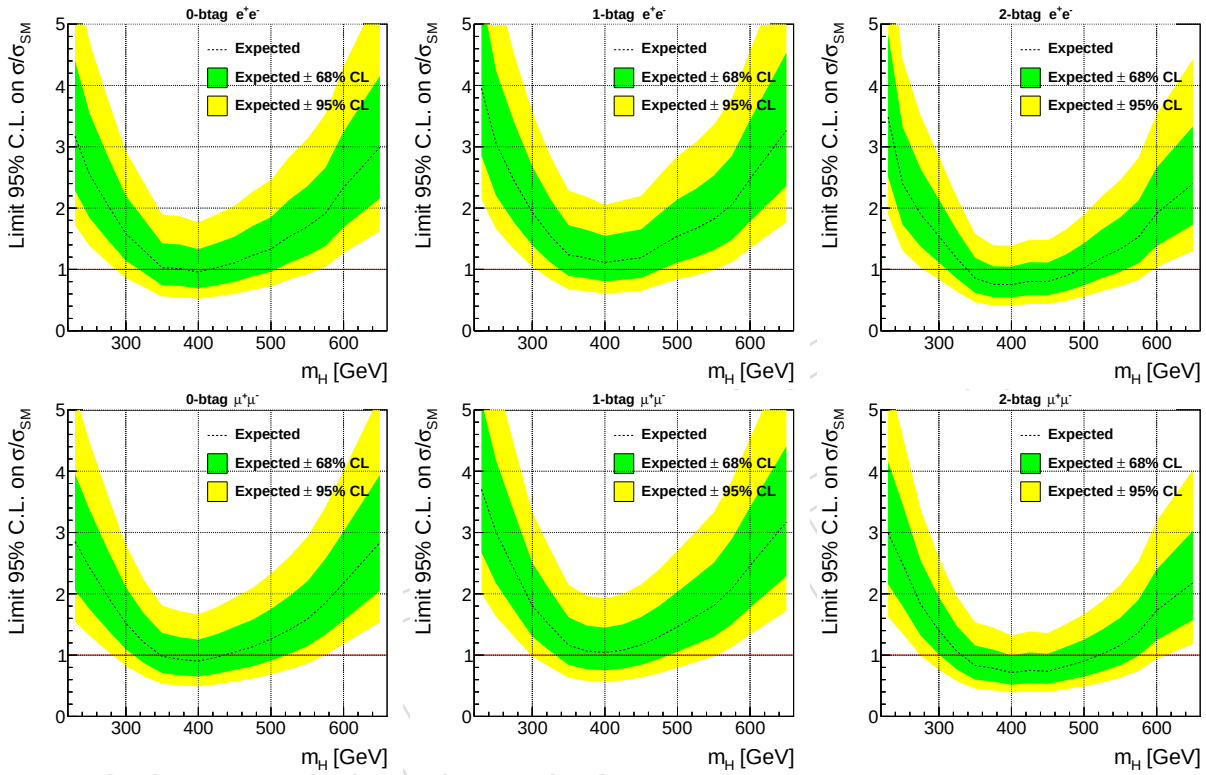


Figure 19: Observed (solid) and expected (dashed) 95% CL upper limit on the ratio of the production cross section to the SM expectation for the Higgs boson obtained using the CL_s technique, separate for the different lepton flavor and b-tag categories. The 68% and 95% ranges of expectation for the background-only model are also shown with green and yellow bands, respectively. The solid line at 1 indicates the expectation for a SM-Higgs-like boson. [PLOT TO BE CHANGED AFTER UNBLINDING]

[5] LHC Higgs Cross Section Working Group et al., "Handbook of LHC Higgs Cross Sections: 1. Inclusive Observables", *CERN-2011-002* (CERN, Geneva, 2011) arXiv:1101.0593.

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

[7] Particle Data Group Collaboration, "Review of particle physics", *J. Phys.* **G37** (2010) 075021, doi:10.1088/0954-3899/37/7A/075021.

[8] S. Frixione, P. Nason, and C. Oleari, "Matching NLO QCD computations with Parton Shower simulations: the POWHEG method", *JHEP* **11** (2007) 070, doi:10.1088/1126-6708/2007/11/070, arXiv:0709.2092.

[9] The CMS generator group, "https://twiki.cern.ch/twiki/bin/viewauth/CMS/StandardModelCrossSectionsat8TeV", twiki, CERN, (2012).

[10] J. Alwall et al., "MadGraph/MadEvent v4: the new web generation", *JHEP* **09** (2007) 028, doi:10.1088/1126-6708/2007/09/028, arXiv:0706.2334.

[11] T. Sjöstrand, S. Mrennan and P. Z. Skands, "PYTHIA 6.4 Physics and Manual", *JHEP* **05** (2006) 026, doi:10.1088/1126-6708/2006/05/026, arXiv:hep-ph/0603175.

[12] CMS, "https://twiki.cern.ch/twiki/bin/view/CMS/HiggsZZ2l2q2012Skims", twiki, CERN, (2012).

[13] CMS Collaboration Collaboration, "Performance of muon reconstruction and identification in pp collisions at $\sqrt{s}$=7 TeV", *CMS PAS* **CMS-MUO-10-004under approval** (2010).

[14] CMS Collaboration, "Performance of CMS muon identification in pp collisions at $\sqrt{s} = 7$ TeV", *CMS PAS* **MUO-2010-002** (2010).

[15] CMS Collaboration, "Jet Performance in pp Collisions at $\sqrt{s}$=7 TeV", CMS Physics Analysis Summary CMS-PAS-JME-10-003, CERN, (2010).

[16] CMS Collaboration, "Commissioning of the Particle-Flow Reconstruction in Minimum-Bias and Jet Events from pp Collisions at 7 TeV", CMS Physics Analysis Summary CMS-PAS-PFT-10-002, CERN, (2010).

[17] CMS Collaboration, "Trigger strategies for Higgs searches in 2011", CMS Analysis Note CMS-AN-2011-065, CERN, (2011).

[18] CMS Collaboration, "Search for the Standard Model Higgs Boson in the Decay Channel $H \to ZZ \to 2l2b$", CMS Analysis Note CMS-AN-2011-399, CERN, (2011).

[19] W. Adam et al., "Reconstruction of Electrons with the Gaussian-Sum Filter in the CMS Tracker at the LHC", CMS Note 2005/001, CERN, (2005).

[20] CMS, "https://twiki.cern.ch/twiki/bin/view/CMS/EgammaCutBasedIdentification", twiki, CERN, (2012).

[21] CMS, "https://twiki.cern.ch/twiki/bin/view/CMSPublic/SWGuideMuonId", twiki, CERN, (2012).

[22] M. Cacciari, G. P. Salam, and G. Soyez, "The Anti-k(t) jet clustering algorithm", *JHEP* **0804** (2008) 063, doi:10.1088/1126-6708/2008/04/063, arXiv:0802.1189.

[23] CMS Collaboration, "Jet Energy Corrections determination at 7 TeV", CMS Physics Analysis Summary CMS-PAS-JME-10-010, CERN, (2010).

[24] M. Cacciari, G. P. Salam, and G. Soyez, "FastJet user manual (for version 3.0.2)", *Eur. Phys. J. C* **72** (2012) 1896, doi:10.1140/epjc/s10052-012-1896-2, arXiv:1111.6097.

[25] P. Van Hove and D. Bloch, "Mistag rate on b-tagging in 2012 data", CMS Note CMS-AN-2012/195, CERN, (2012).

[26] CMS Collaboration, "Identification of b-quark jets with the CMS experiment", arXiv:1211.4462v1.

[27] Y. Gao et al., "Spin determination of single-produced resonances at hadron colliders", *Phys.Rev.* **D81** (2010) 075022, doi:10.1103/PhysRevD.81.075022, arXiv:1001.3396.

[28] A. De Rujula et al., "Higgs look-alikes at the LHC", *Phys.Rev.* **D82** (2010) 013003, doi:10.1103/PhysRevD.82.013003, arXiv:1001.5300.

[29] CMS Collaboration, "Angular Analysis of Resonances $pp \to X \to ZZ$", CMS Analysis Note CMS-AN-2010-351, CERN, (2010).

[30] CMS Collaboration, "Missing transverse energy performance of the CMS detector", *JINST* **6** (2011) 09001, doi:10.1088/1748-0221/6/09/P09001.

[31] CMS Collaboration, "CMS Luminosity Based on Pixel Cluster Counting - Summer 2012 Update", CMS Physics Analysis Summary CMS-PAS-LUM-12-001, CERN, (2012).

[32] Michael Hildreth, "Estimating Systematic Errors Due to Pileup Modeling", twiki, CERN, (2012).

[33] P. M. Nadolsky et al., "Implications of CTEQ global analysis for collider observables", *Phys.Rev.* **D78** (2008) 013004, doi:10.1103/PhysRevD.78.013004, arXiv:0802.0007.

[34] A. Martin et al., "Parton distributions for the LHC", *Eur.Phys.J.* **C63** (2009) 189–285, doi:10.1140/epjc/s10052-009-1072-5, arXiv:0901.0002.

[35] R. D. Ball et al., "Impact of Heavy Quark Masses on Parton Distributions and LHC Phenomenology", *Nucl.Phys.* **B849** (2011) 296–363, doi:10.1016/j.nuclphysb.2011.03.021, arXiv:1101.1300.

[36] CMS Collaboration, "Updated results on the new boson discovered in the search for the standard model Higgs boson in the H→ZZ→ 4ℓ channel in pp collisions at $\sqrt{s} = 7$ and 8 TeV", CMS Analysis Note CMS-AN-2012-367, CERN, (2012).

[37] S. Goria, G. Passarino, and D. Rosco, "The Higgs Boson Lineshape", *Nucl.Phys.* **B864** (2012) 530–579, doi:10.1016/j.nuclphysb.2012.07.006, arXiv:1112.5517.

[38] CMS, "LHC Higgs Cross Section Working Group web page", https://twiki.cern.ch/twiki/bin/view/lhcphysics/, CERN, (2012).

[39] CMS Collaboration Collaboration, "Search for a standard-model-like Higgs boson with a mass of up to 1 TeV at the LHC", *Submitted to Eur. Phys. J. C* (2012) arXiv:1304.0213v1.

[40] G. Passarino, "Higgs Interference Effects in gg $\to$ ZZ and their Uncertainty", *JHEP* **1208** (2012) 146, doi:10.1007/JHEP08(2012)146, arXiv:1206.3824.

[41] C. H. C. Group, "Documentation of the RooStats-based statistics tools for Higgs PAG", *CMS TWiki* SWGuideHiggsAnalysisCombinedLimit (2011).

[42] L. Moneta et al., "The RooStats Project", in *13th International Workshop on Advanced Computing and Analysis Techniques in Physics Research (ACAT2010)*. SISSA, 2010. arXiv:1009.1003. PoS(ACAT2010)057.

[43] G. Cowan, K. Cranmer, E. Gross, O. Vitells, "Asymptotic formulae for likelihood based tests of new physics", *Eur. Phys. J.* **C71** (2011) 1554, arXiv:physics/1007.1727v2.

A Data and Monte Carlo samples

Table 10: Data samples used in the analysis.

Channel	Dataset	Luminosity [pb^{-1}]
$2\mu 2q$	/DoubleMu/Run2012A-13Jul2012-v1/AOD	808
	/DoubleMu/Run2012A-recover-06Aug2012-v1/AOD	82
	/DoubleMu/Run2012B-13Jul2012-v4/AOD	4429
	/DoubleMu/Run2012C-24Aug2012-v1/AOD	495
	/DoubleMu/Run2012C-EcalRecover_11Dec2012-v1/AOD	134
	/DoubleMu/Run2012C-PromptReco-v2/AOD	6394
	/DoubleMu/Run2012D-PromptReco-v1/AOD	7274
$2e2q$	/DoubleElectron/Run2012A-13Jul2012-v1/AOD	808
	/DoubleElectron/Run2012A-recover-06Aug2012-v1/AOD	82
	/DoubleElectron/Run2012B-13Jul2012-v4/AOD	4429
	/DoubleElectron/Run2012C-24Aug2012-v1/AOD	495
	/DoubleElectron/Run2012C-EcalRecover_11Dec2012-v1/AOD	134
	/DoubleElectron/Run2012C-PromptReco-v2/AOD	6394
	/DoubleElectron/Run2012D-PromptReco-v1/AOD	7274
$e\mu qq$	/MuEG/Run2012A-13Jul2012-v1/AOD	808
	/MuEG/Run2012A-recover-06Aug2012-v1/AOD	82
	/MuEG/Run2012B-13Jul2012-v4/AOD	4429
	/MuEG/Run2012C-24Aug2012-v1/AOD	495
	/MuEG/Run2012C-EcalRecover_11Dec2012-v1/AOD	134
	/MuEG/Run2012C-PromptReco-v2/AOD	6394
	/MuEG/Run2012D-PromptReco-v1/AOD	7274

Table 11: The signal samples, $H \to ZZ \to \ell^+\ell^- q\bar{q}$ ($\ell = e, \mu, \tau$), simulated with POWHEG are */GluGluToHToZZTo2L2Q_M-xyz_8TeV-powheg-pythia6/Summer12_DR53X-PU_S10_START53_V7A-v1/AODSIM* , where *xyz* is the Higgs boson mass hypothesis, M_H. The cross section times branching fraction for each M_H value is listed in pb.

M_H (GeV/c^2)	$\sigma \times$ Br($H \to ZZ \to \ell^+\ell^- q\bar{q}$) [pb]	M_H (GeV/c^2)	$\sigma \times$ Br($H \to ZZ \to \ell^+\ell^- q\bar{q}$) [pb]
200	0.2566	400	0.1111
210	0.2538	425	0.0914
220	0.2416	450	0.7311
230	0.2278	475	0.6000
250	0.2022	500	0.4719
275	0.1751	525	0.0380
300	0.1563	550	0.0305
325	0.1478	575	0.0250
350	0.1482	600	0.0201
375	0.1360		

Table 12: Background simulated samples of the Summer12 production used in the analysis. The equivalent luminosity of the processed events for each sample is computed using the (N)NLO cross section in the 3rd column.

Process	dataset	σ [pb]	luminosity [fb^{-1}]
Z+jets (inclusive)	/DYJetsToLL_M-50_TuneZ2Star_8TeV-madgraph-tarball/ Summer12_DR53X-PU_S10_START53_V7A-v1/AODSIM	3503.71	8.7
Z+1 jet (exclusive)	/DY1JetsToLL_M-50_TuneZ2Star_8TeV-madgraph/ Summer12_DR53X-PU_S10_START53_V7A-v1/AODSIM	660.6	36.4
Z+2 jet (exclusive)	/DY2JetsToLL_M-50_TuneZ2Star_8TeV-madgraph/ Summer12_DR53X-PU_S10_START53_V7A-v1/AODSIM	215.1	101.6
Z+3 jet (exclusive)	/DY3JetsToLL_M-50_TuneZ2Star_8TeV-madgraph/ Summer12_DR53X-PU_S10_START53_V7A-v1/AODSIM	65.79	167.4
Z+4 jet (exclusive)	/DY4JetsToLL_M-50_TuneZ2Star_8TeV-madgraph/ Summer12_DR53X-PU_S10_START53_V7A-v1/AODSIM	27.59	232.1
$t\bar{t}$	/TTTo2L2Nu2B_8TeV-powheg-pythia6/ Summer12_DR53X-PU_S10_START53_V7A-v1/AODSIM	23.38	461
ZZ	/ZZ_TuneZ2star_8TeV_pythia6_tauola/ Summer12_DR53X-PU_S10_START53_V7A-v1/AODSIM	17.654	549
WZ	/WZ_TuneZ2star_8TeV_pythia6_tauola/ Summer12_DR53X-PU_S10_START53_V7A-v1/AODSIM	22.88	424
WW	/WW_TuneZ2star_8TeV_pythia6_tauola/ Summer12_DR53X-PU_S10_START53_V7A-v1/AODSIM	57.1097	168

B Trigger efficiencies and Monte Carlo correction factors

Table 13: Working point loose to the HLT Ele8 leg tag-and-probe efficiencies and scale factors.

η coverage	p_T range (GeV/c)	efficiency (data)	efficiency (MC)	data/MC ratio
$0.0 < \|\eta\| < 0.8$	$20 < p_T < 40$	0.986 ± 0.001	0.988 ± 0.001	0.997 ± 0.001
$0.8 < \|\eta\| < 1.4$		0.936 ± 0.001	0.946 ± 0.001	0.990 ± 0.002
$1.6 < \|\eta\| < 2.0$		0.901 ± 0.002	0.905 ± 0.003	0.995 ± 0.004
$2.0 < \|\eta\| < 2.5$		0.944 ± 0.002	0.944 ± 0.002	1.000 ± 0.003
$0.0 < \|\eta\| < 0.8$	$40 < p_T < 200$	0.991 ± 0.001	0.994 ± 0.001	0.997 ± 0.000
$0.8 < \|\eta\| < 1.4$		0.976 ± 0.001	0.978 ± 0.001	0.998 ± 0.001
$1.6 < \|\eta\| < 2.0$		0.945 ± 0.002	0.946 ± 0.002	0.999 ± 0.002
$2.0 < \|\eta\| < 2.5$		0.962 ± 0.002	0.962 ± 0.002	1.000 ± 0.002

Table 14: Working point loose to the HLT Ele17 leg tag-and-probe efficiencies and scale factors.

η coverage	p_T range (GeV/c)	efficiency (data)	efficiency (MC)	data/MC ratio
$0.0 < \|\eta\| < 0.8$	$20 < p_T < 40$	0.983 ± 0.001	0.984 ± 0.001	0.999 ± 0.001
$0.8 < \|\eta\| < 1.4$		0.932 ± 0.001	0.940 ± 0.001	0.991 ± 0.002
$1.6 < \|\eta\| < 2.0$		0.895 ± 0.002	0.898 ± 0.003	0.997 ± 0.004
$2.0 < \|\eta\| < 2.5$		0.933 ± 0.002	0.935 ± 0.003	0.998 ± 0.003
$0.0 < \|\eta\| < 0.8$	$40 < p_T < 200$	0.989 ± 0.001	0.991 ± 0.001	0.998 ± 0.001
$0.8 < \|\eta\| < 1.4$		0.972 ± 0.001	0.973 ± 0.001	0.999 ± 0.001
$1.6 < \|\eta\| < 2.0$		0.938 ± 0.002	0.939 ± 0.002	0.999 ± 0.002
$2.0 < \|\eta\| < 2.5$		0.951 ± 0.002	0.955 ± 0.002	0.996 ± 0.003

Table 15: Dimuon trigger efficiencies, calculated using the Muon POG official numbers, for two tight muons, both with $p_T > 20$ GeV/c, in four bins of pseudorapidity for each of the two muons.

Muon η	$0.0 < \|\eta\| < 0.9$	$0.9 < \|\eta\| < 1.2$	$1.2 < \|\eta\| < 2.1$	$2.1 < \|\eta\| < 2.4$
$0.0 < \|\eta\| < 0.9$	0.938 ± 0.011	0.880 ± 0.014	0.864 ± 0.012	0.880 ± 0.021
$0.9 < \|\eta\| < 1.2$	0.880 ± 0.014	0.836 ± 0.021	0.824 ± 0.017	0.819 ± 0.047
$1.2 < \|\eta\| < 2.1$	0.864 ± 0.012	0.824 ± 0.017	0.813 ± 0.010	0.804 ± 0.021
$2.1 < \|\eta\| < 2.4$	0.880 ± 0.021	0.819 ± 0.047	0.804 ± 0.021	0.784 ± 0.063

As a cross check, we have compared the trigger efficiencies calculated from data with the trigger efficiencies provided by the HLT trigger simulation. Table 16 shows the trigger simulation efficiencies for the DY Z+jets and Higss signal (300 GeV/c^2) simulated samples. Table 17 lists the ratio between the trigger efficiencies calculated from data and the trigger simulation. The overall ratio, averaged over η, is 0.94, both for the DY Z+jets and Higss signal (300 GeV/c^2) simulated samples. This discrepancy is under study and can be explained by the missing cut on the longitudinal distance between the two muons (Δz) in the muon trigger simulation. In any case, this small discrepancy would only be relevant for the signal, given the background normalization is contrained to the data in the m_{jj} sideband region.

We have performed the same study for electrons. The overall trigger effiencies calculated from data and from the trigger simulation, averaged over p_{T} and η, agree within 1%.

Table 16: Muon trigger efficiencies from trigger simulation for the DY Z+jets and Higss signal (300 GeV/c^2) simulated samples.

	DY Z+jets				Higgs signal, 300 GeV/c^2			
muon η	$0-0.9$	$0.9-1.2$	$1.2-2.1$	$2.1-2.4$	$0-0.9$	$0.9-1.2$	$1.2-2.1$	$2.1-2.4$
$0-0.9$	0.97	0.94	0.93	0.92	0.97	0.94	0.94	0.93
$0.9-1.2$	0.94	0.91	0.91	0.89	0.94	0.91	0.90	0.91
$1.2-2.1$	0.93	0.91	0.90	0.88	0.93	0.91	0.91	0.88
$2.1-2.4$	0.92	0.88	0.88	0.85	0.90	0.89	0.88	0.97

Table 17: Ratio between the trigger efficiencies calculated from data (Table 15) and from the trigger simulation (Table 16) for the DY Z+jets and Higss signal (300 GeV/c^2) simulated samples.

	DY Z+jets				Higgs signal, 300 GeV/c^2			
muon η	$0-0.9$	$0.9-1.2$	$1.2-2.1$	$2.1-2.4$	$0-0.9$	$0.9-1.2$	$1.2-2.1$	$2.1-2.4$
$0-0.9$	0.97	0.94	0.93	0.96	0.97	0.94	0.92	0.95
$0.9-1.2$	0.94	0.91	0.90	0.92	0.94	0.92	0.91	0.90
$1.2-2.1$	0.93	0.91	0.90	0.92	0.93	0.91	0.90	0.91
$2.1-2.4$	0.95	0.93	0.91	0.92	0.97	0.92	0.91	0.81

C Lepton identification requirements and efficiencies

The standard tag-&-probe method used to evaluate the lepton identification efficiencies from data requires the reconstruction of the dilepton system with invariant mass in the range [60-120] GeV/c^2. One of the leptons, called tag, is required to pass full selection criteria and to match the tighter leg of the trigger. The other lepton candidate, called probem is selected with criteria that depend on the efficiency being measured. The sample is divided into two exclusive subsamples depending on whether the probe lepton passes or fails the selection criteria currently under investigation. Due to the presence of background events, the signal yields are obtained with a fit to the invariant mass distribution of the dilepton system. The measured efficiency is calculated as a function of p_T and η of the probe lepton from the relative yields of the signal in subsamples with passing or failing probes. Finally the data to MC scale factors are deduced by dividing the efficiencies in data to the ones obtained from MC using exactly same procedure. Scale factors are used instead of raw efficiencies in order to benefit from partial cancellations of systematic uncertainites associated with the procedure. The total efficiency measurement is factorized into five sequential relative efficiency measurements: tracking, reconstruction, identification, isolation and the total trigger efficiency, given by the product:

$$\epsilon_{lepton} = \epsilon_{tracking} \times \epsilon_{RECO/Tracking} \times \epsilon_{ID/RECO} \times \epsilon_{ISO/ID} \times \epsilon_{Trigger/ISO}$$

Table 18: Electron ID requirements for the Loose ID working point.

Variable	Barrel cut	Endcap cut
$\Delta\eta_{trk,supercluster}$	< 0.007	< 0.009
$\Delta\phi_{trk,supercluster}$	< 0.15	< 0.1
$\sigma_{i\eta,i\eta}$	< 0.01	< 0.03
H/E	< 0.12	< 0.10
d_0 (wrt primary vertex)	$< 0.2\,\text{mm}$	$< 0.2\,\text{mm}$
d_z (wrt primary vertex)	$< 2\,\text{mm}$	$< 2\,\text{mm}$
$\lvert 1/E - 1/p \rvert$	< 0.05	< 0.05
$I_{PF,\,corr}/p_\text{T}$	< 0.15	< 0.15
Missing hits	≤ 1	≤ 1
Conversion vertex fit prob.	$< 10^{-6}$	$< 10^{-6}$

Table 19: Muon ID requirements for the Tight ID working point.

Variable	Cut
isGlobalMuon	True
isPFMuon	True
isTrackerMuon	True
$\chi^2/ndof$ (global fit)	< 10
Muon chamber hits in global fit	> 0
Muon stations with muon segments	> 1
d_{xy} (from tracker, wrt primary vertex)	$< 2\,\text{mm}$
d_z (from tracker, wrt primary vertex)	$< 5\,\text{mm}$
Valid pixel hits (tracker track)	> 0
Tracker layers with hits	> 5
$I_{PF,\,corr}/p_\text{T}$	< 0.12

Table 20: Data to simulation scale factors for electron (upper) and muon (lower) identification requirements in various η ranges.

electron p_T	$0.0 < \lvert\eta\rvert < 0.8$	$0.8 < \lvert\eta\rvert < 1.442$	$1.556 < \lvert\eta\rvert < 2.0$	$2.0 < \lvert\eta\rvert < 2.5$
20 - 30	1.005 ± 0.003	0.981 ± 0.003	0.980 ± 0.005	1.017 ± 0.006
30 - 40	1.004 ± 0.001	0.991 ± 0.001	0.992 ± 0.002	1.019 ± 0.003
40 - 50	1.008 ± 0.001	0.994 ± 0.001	1.004 ± 0.002	1.005 ± 0.001
50 - 200	1.008 ± 0.001	0.999 ± 0.001	1.006 ± 0.003	1.009 ± 0.002

muon p_T	$0.0 < \lvert\eta\rvert < 0.8$	$0.8 < \lvert\eta\rvert < 2.1$	$2.1 < \lvert\eta\rvert < 2.4$
20 - 40	1.0043 ± 0.0004	1.0074 ± 0.0005	1.022 ± 0.001
40 - 100	1.0012 ± 0.0004	1.0043 ± 0.0004	1.014 ± 0.001

D Extra plots

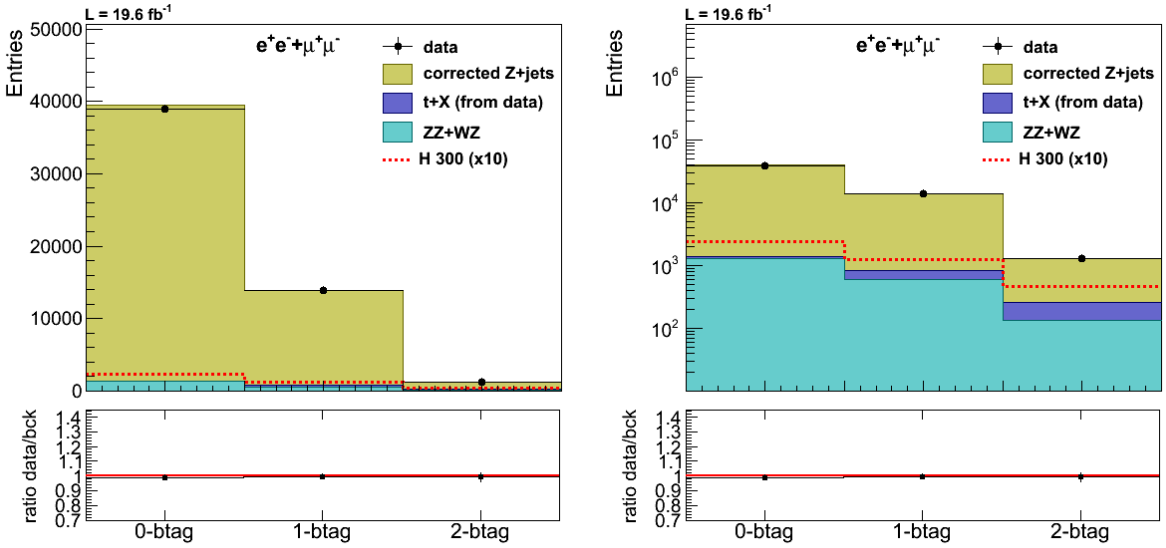


Figure 20: Number of number of events in each btag category after full selection, for the electron and muon channel combined, in linear (left) and logarithmic scales (right). Dots indicate data, brownish histogram corrected Z+jets simulation, light blue simulated diboson background and dark blue $t\bar{t}$ events from data (which

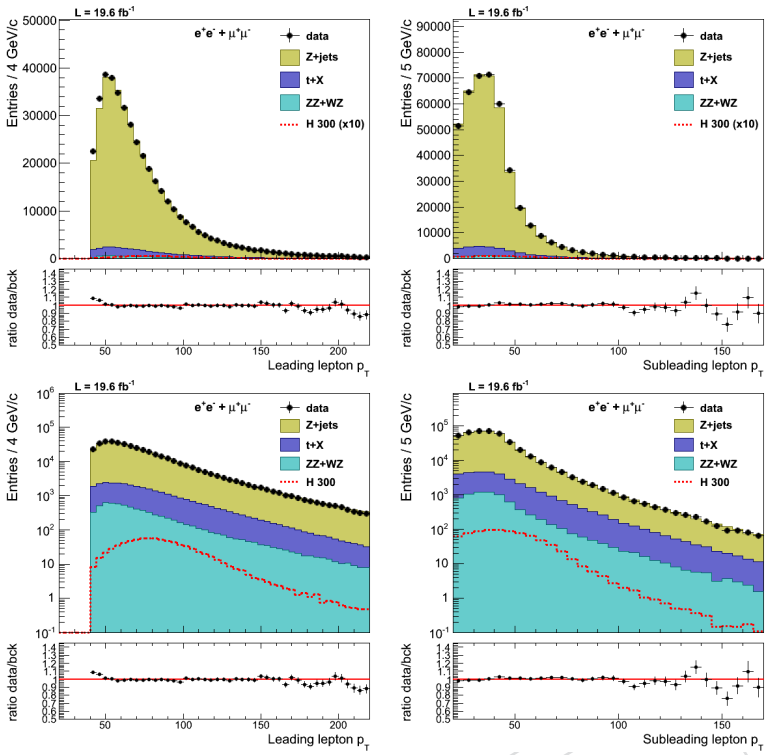


Figure 21: Distributions of the p_{T} –in linear (upper) and logarithmic scales (lower)– of the leading (left) and subleading lepton (right) after preselection cuts. Dots indicate data, brownish histogram corrected Z+jets simulation, light blue simulated diboson background and dark blue $\mathrm{t\bar{t}}$ events from data (which include single top, WW, $Z \to \tau^{+}\tau^{-}$+jets).

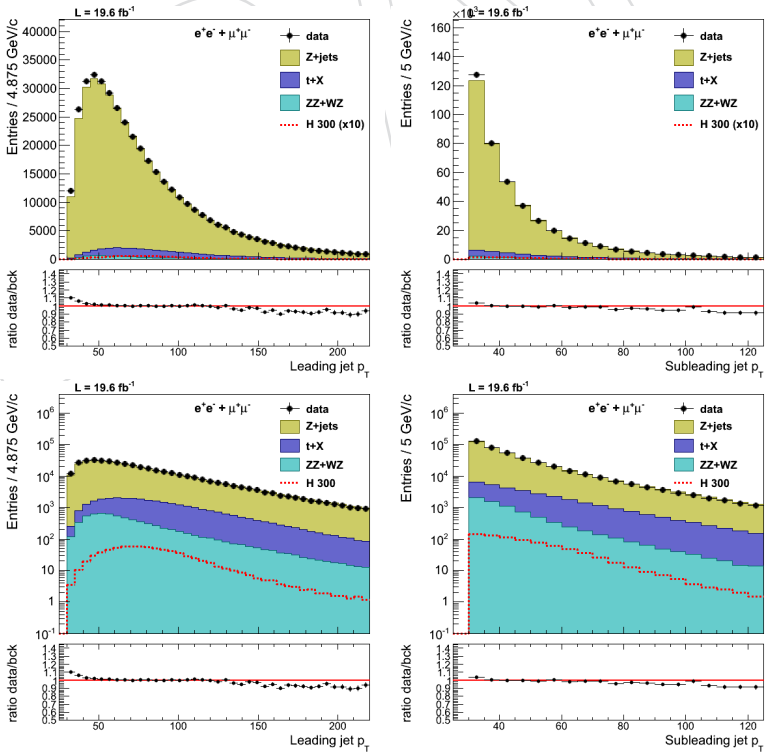


Figure 22: Similar distributions for the leading (left) and subleading jet (right).

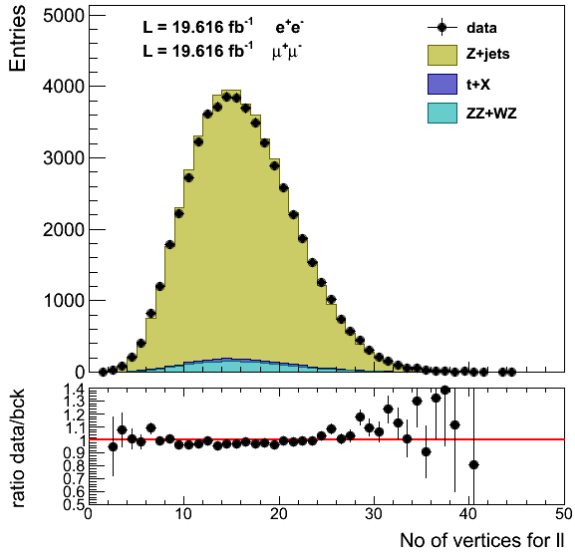


Figure 23: Reconstructed number of good vertices in data (points with error bars) and in reweighed simulation (solid histograms), from the electron and muon channels combined. The number of true interactions in MC has been reweighed to match the estimated distributions in data.

E Expected signal and background events per btag category

Numbers of signal and background events expected for 19.6 fb^{-1} in the mass range $[M_{\mathrm{H}} - 6\%, M_{\mathrm{H}} + 10\%]$ in the three btag categories, separately for the $\mu^+\mu^- jj$ and $\mathrm{e}^+\mathrm{e}^- jj$ channels (denoted $\mu\mu$ and ee, respectively).

Table 21: Expected yields in the 0-btag category.

M_{H} (GeV/c^2)	signal		Z+jets		eμ data		diboson		total background	
	$\mu\mu$	ee	$\mu\mu$	ee	$\mu\mu$	ee	$\mu\mu$	ee	$\mu\mu$	ee
250	90.0	80.4	6160.5	5312.2	24	19	201.7	175.6	6386.3	5506.7
300	90.0	80.6	3334.3	2800.3	11	9	121.1	106.6	3466.6	2915.7
400	82.6	73.5	1254.4	1065.9	1	0	55.9	49.3	1310.9	1115.6
500	33.4	30.2	429.0	381.9	1	1	23.9	20.4	454.0	403.2
600	11.9	10.5	179.3	152.3	0	0	11.3	10.8	190.6	163.2

Table 22: Expected yields in the 1-btag category.

M_{H} (GeV/c^2)	signal		Z+jets		eμ data		diboson		total background	
	$\mu\mu$	ee	$\mu\mu$	ee	$\mu\mu$	ee	$\mu\mu$	ee	$\mu\mu$	ee
250	44.7	40.1	2124.7	1779.0	87	68	89.7	74.2	2301.2	1921.4
300	46.8	40.8	1174.5	987.0	33.0	25	60.8	49.1	1268.3	1062.1
400	45.0	40.8	459.8	406.7	6	5	26.9	24.1	492.9	435.7
500	18.9	17.0	181.1	155.8	1	0	13.9	10.1	195.5	166.3
600	7.0	6.4	80.9	61.4	0	0	6.5	5.5	87.4	66.9

Table 23: Expected yields in the 2-btag category.

M_{H} (GeV/c^2)	signal		Z+jets		eμ data		diboson		total background	
	$\mu\mu$	ee	$\mu\mu$	ee	$\mu\mu$	ee	$\mu\mu$	ee	$\mu\mu$	ee
250	15.6	14.8	180.9	138.6	42	33	18.7	16.4	241.7	188.0
300	18.5	16.0	91.2	70.9	28	22	12.2	11.7	131.4	104.6
400	19.8	16.8	32.2	24.6	1	0	7.0	5.3	39.7	30.4
500	8.2	7.5	9.6	8.9	0	0	2.7	2.5	12.3	11.4
600	3.0	2.8	5.6	4.5	0	0	1.7	1.5	7.3	6.0

F Determination of $t\bar{t}$ background from data

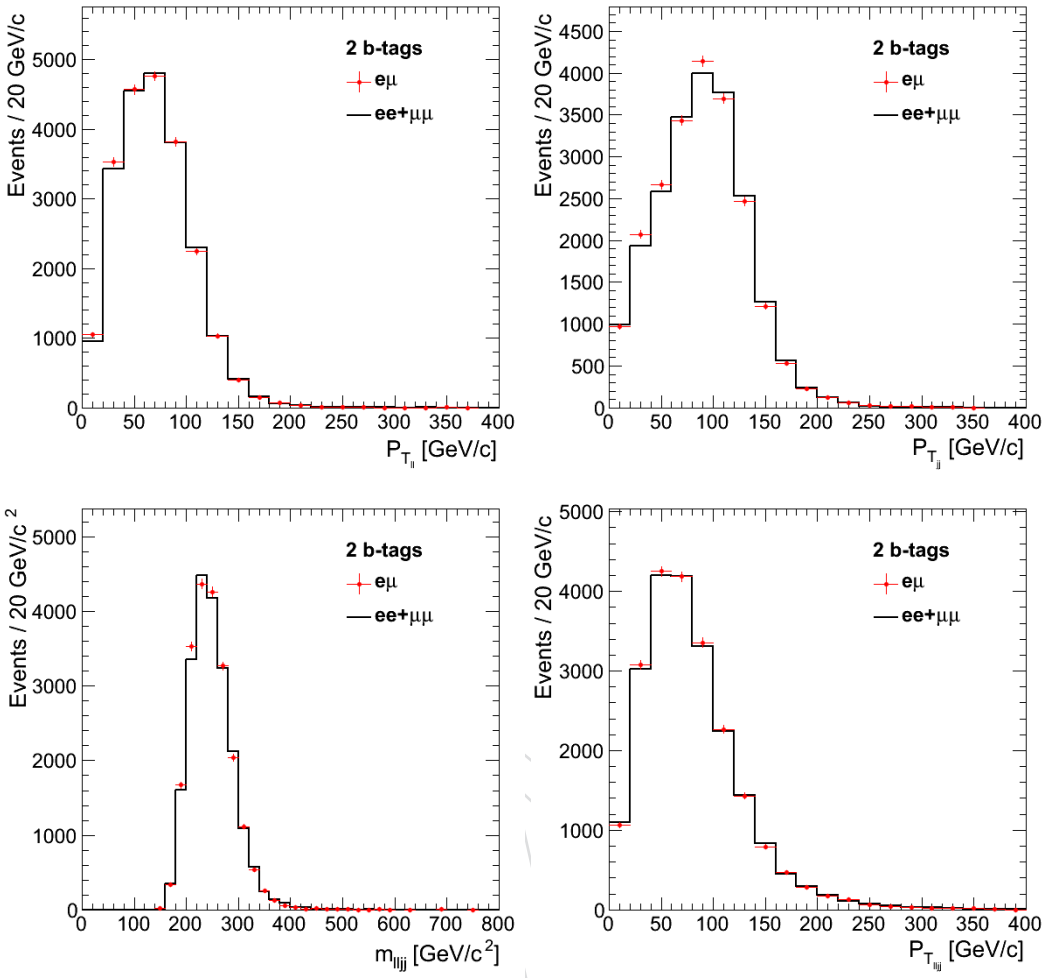


Figure 24: Powheg + Pythia top MC $e^{\pm}\mu^{\mp}$ to $(e^+e^- + \mu^+\mu^-)$ comparison for several variables after different steps of the selection, as specified in the legends. Top: dilepton (left) and di-jet transverse momentum (right). Bottom: dilepton + dijet "Higgs" invariant mass (left) and transverse momentum (right).

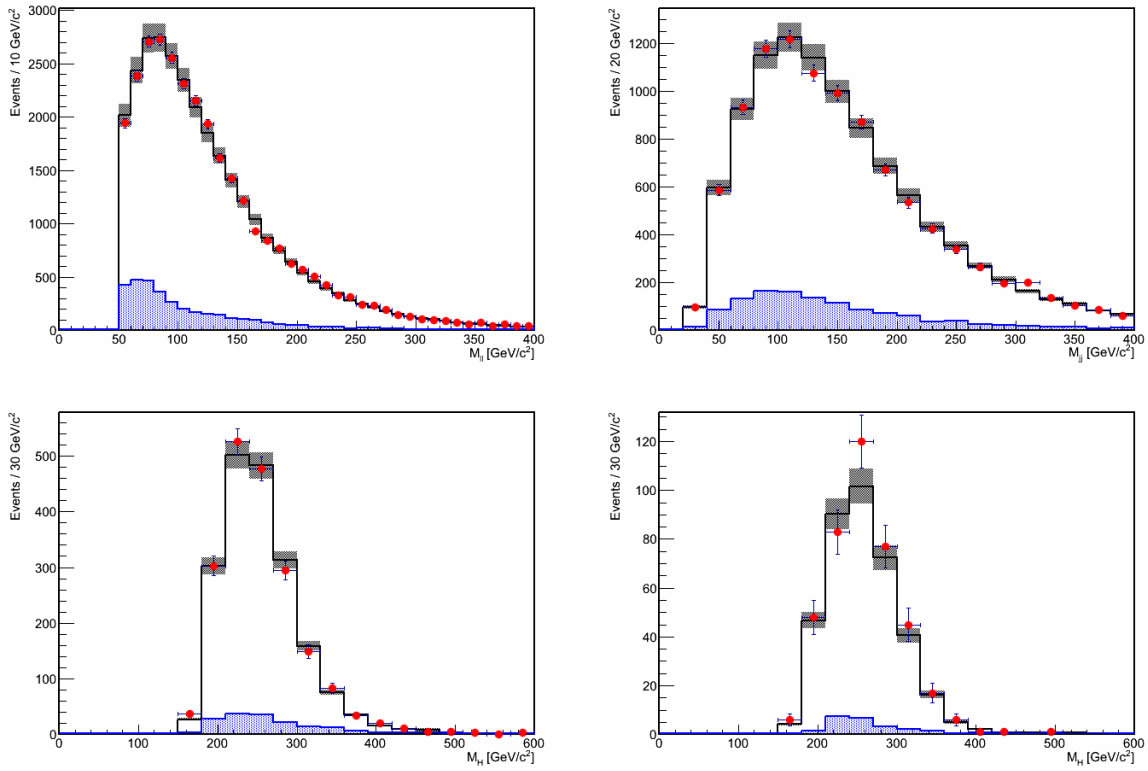


Figure 25: Comparison of 2012 $e^{\pm}\mu^{\mp}$ data to Powheg + Pythia top MC, corresponding to an integrated luminosity of 19.6 fb^{-1}. Red dots are $e^{\pm}\mu^{\mp}$ data; white histogram top Monte Carlo; blue histogram other small backgrounds. Top: dilepton invariant mass (left) and dijet invariant mass (right). Bottom: "Higgs" invariant mass for events with 1 JPL b-tag (left), and two (1 JPM + 1 JPL) b-tags and MET significance < 10 (right).

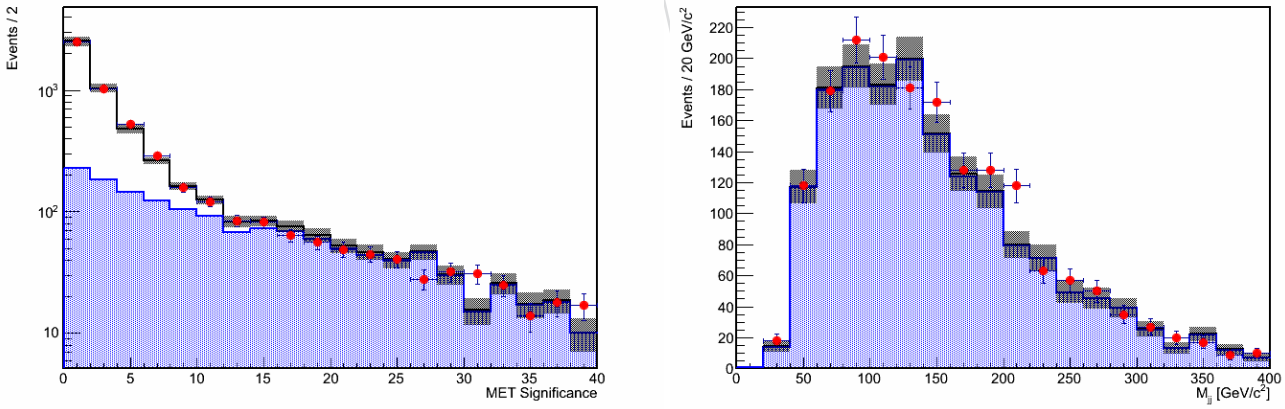


Figure 26: MET significance distribution for dilepton data compared to the sum of Drell-Yan Monte Carlo plus $e^{\pm}\mu^{\mp}$ data for events with and two (1 JPM + 1 JPL) btags (left). Dijet invariant mass (right) for $e^{+}e^{-} + \mu^{+}\mu^{-}$ and $e^{\pm}\mu^{\mp}$ data for events outside the leptonic Z mass window, with two (1 JPM + 1 JPL) b-tags, and MET significance > 8. Other cuts are detailed in the text. Red dots are $e^{+}e^{-} + \mu^{+}\mu^{-}$ data; white histogram Drell Yan Monte Carlo; blue histogram $e^{\pm}\mu^{\mp}$ data (plus other small backgrounds).

G Systematic uncertainties on the signal

Table 24: Systematic uncertainty on the signal in the muon channel.

M_H (GeV/c^2)	0-btag		1-btag		2-btag	
	$1\sigma_{UP}$ (%)	$1\sigma_{DOWN}$ (%)	$1\sigma_{UP}$ (%)	$1\sigma_{DOWN}$ (%)	$1\sigma_{UP}$ (%)	$1\sigma_{DOWN}$ (%)
200	-3.2	1.5	5.8	-1.3	4.0	-6.7
210	-3.4	1.2	4.8	-1.9	8.5	-2.5
220	-3.1	1.7	4.5	-1.7	6.8	-6.1
230	-3.4	1.8	4.8	-2.1	7.4	-5.3
250	-2.9	2.1	3.8	-2.4	6.3	-5.5
275	-3.4	1.5	4.5	-1.0	6.1	-5.5
300	-3.6	1.4	4.6	-0.69	6.8	-5.7
350	-3.8	1.7	4.7	-0.84	6.7	-6.1
375	-3.7	1.7	5.0	-0.31	5.2	-7.5
400	-3.9	1.6	4.5	-0.74	6.8	-5.7
425	-4.1	1.6	5.1	-0.59	5.9	-5.8
450	-4.1	1.8	4.0	-1.1	8.1	-5.0
475	-3.9	1.8	4.1	-0.6	7.0	-6.5
500	-3.6	2.1	3.6	-0.97	7.1	-6.7
525	-4.1	2.0	4.8	-0.65	6.4	-7.1
550	-3.6	1.9	3.5	-0.87	6.9	-5.6
575	-4.4	1.8	4.9	-0.78	6.7	-5.7
600	-4.3	1.9	4.4	-1.3	7.8	-4.9

Table 25: Systematic uncertainty on the signal in the electron channel.

M_H (GeV/c^2)	0-btag		1-btag		2-btag	
	$1\sigma_{UP}$ (%)	$1\sigma_{DOWN}$ (%)	$1\sigma_{UP}$ (%)	$1\sigma_{DOWN}$ (%)	$1\sigma_{UP}$ (%)	$1\sigma_{DOWN}$ (%)
200	-3.4	1.8	3.9	-3.7	12	-0.18
210	-3.5	1.2	5.4	-1.1	6.9	-5.5
220	-2.9	1.9	4.1	-2.0	7.0	-7.2
230	-3.6	1.4	4.9	-1.9	8.7	-3.5
250	-2.9	2.2	4.0	-2.4	5.1	-6.2
275	-3.5	1.5	4.7	-1.3	6.3	-4.6
300	-3.9	1.5	5.9	-0.26	5.9	-7.7
350	-3.9	1.5	4.7	-0.41	7.1	-6.5
375	-3.9	1.8	4.5	-1.1	7.5	-5.7
400	-3.9	1.9	4.0	-1.1	8.2	-6.0
425	-4.0	1.8	4.8	-0.67	6.9	-6.6
450	-3.9	1.5	3.8	-0.43	7.8	-5.3
475	-4.3	1.8	5.2	-0.54	6.1	-6.2
500	-4.1	2.0	5.2	-0.83	5.8	-6.7
525	-4.0	2.0	3.8	-1.3	8.1	-5.5
550	-4.0	1.8	4.5	-1.1	6.2	-5.1
575	-4.0	1.9	4.1	-0.67	7.1	-6.2
600	-4.4	1.8	5.2	-0.68	6.7	-6.1

H $M_{\ell\ell jj}$ distributions for the electron and muon channels

The $M_{\ell\ell jj}$ distributions for the electron and muon channels display an excellent agreement both in the m_{jj} sideband, where the full data sample is used for the comparisons, and signal regions, made with unblinded ICHEP data.

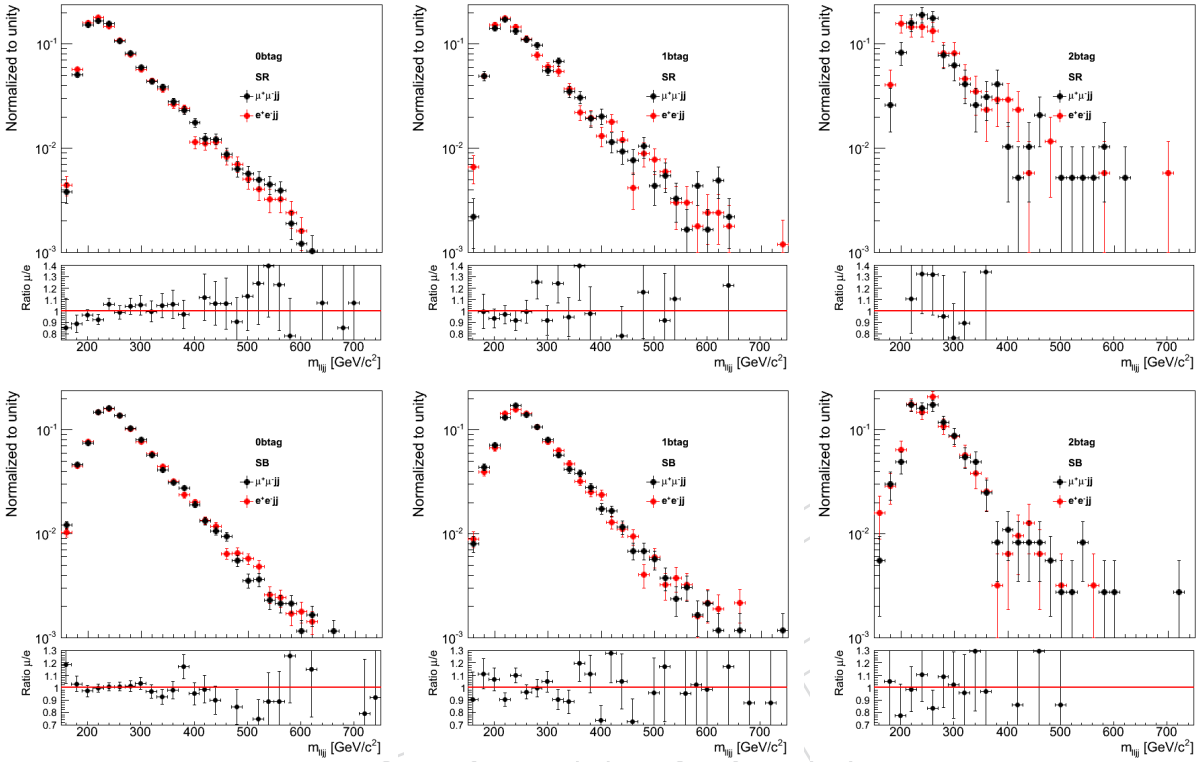


Figure 27: Mass distributions of the $\ell^+\ell^- jj$ system for events in the electron and muon channels: (top) unblinded ICHEP data in the m_{jj} signal region and (bottom) full data sample in the sideband region. From left to right, plots correspond to the 0-, 1-, and 2-btag categories.

I Limit Cross Checks

Another approach, referred to as *cut-and-count* analysis, uses only the number of events selected with a reconstructed Higgs mass in the window $[M_{\mathrm{H}} - 6\%, M_{\mathrm{H}} + 10\%]$. The Higgs cross section limit is determined from the expected number of signal and background events passing the selections s and b respectively. Results (shown in Figure 28) are compatible with the full results.

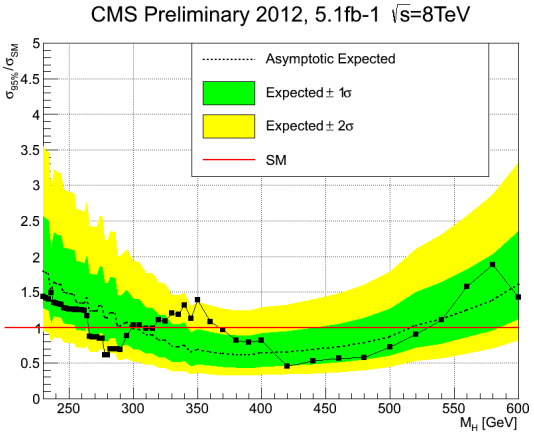


Figure 28: Observed (solid) and expected (dashed) 95% CL upper limit on the ratio of the production cross section to the SM expectation for the Higgs boson obtained using the *cut-and-count* technique. The 68% and 95% ranges of expectation for the background-only model are also shown with green and yellow bands, respectively. The solid line at 1 indicates the expectation for a SM-Higgs-like boson. These results correspond to a check performed on the unblinded part of the dataset (ICHEP data) to verify everything is under control.