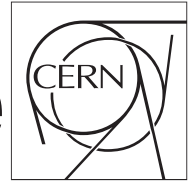


The Compact Muon Solenoid Experiment

# CMS Draft Note

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



2012/11/26

Head Id: 155899

Archive Id: 159169

Archive Date: 2012/11/01

Archive Tag: trunk

## Search for the standard model Higgs Boson in the decay channel $H \rightarrow ZZ \rightarrow 2\ell 2q$ at CMS at $\sqrt{s}=8$ TeV

The CMS Collaboration<sup>3</sup>

<sup>1</sup> Northeastern University

<sup>2</sup> Fermilab

<sup>3</sup> CERN

### Abstract

A search for the standard model Higgs boson decaying to two  $Z$  bosons with subsequent decay to a final state with two leptons and two quark-jets,  $H \rightarrow ZZ^{(*)} \rightarrow (q\bar{q})(\ell^-\ell^+)$ , is presented. Data corresponding to an integrated luminosity of about  $5.1 \text{ fb}^{-1}$  of LHC proton-proton collisions at  $\sqrt{s}=8$  TeV were collected and analyzed by the CMS experiment. The selection to discriminate between signal and background events is based on kinematic and topological quantities, which include the angular spin correlations of the decay products. The events are classified according to probability of the jets to originate from quarks of light or heavy flavor. No evidence for a Higgs boson is found and upper limits on the Higgs boson production cross section are set in the range of masses between 200 and 600 GeV.

This box is only visible in draft mode. Please make sure the values below make sense.

PDFAuthor: Oscar Gonzalez, Matthias Mozer, et al  
PDFTitle: Search for the standard model Higgs Boson in the decay channel  $H \rightarrow ZZ \rightarrow 2\ell 2q$  at CMS  
PDFSubject: CMS  
PDFKeywords: CMS, physics, software, computing

Please also verify that the abstract does not use any user defined symbols



# Contents

1	1	Introduction . . . . .	2
2	2	Data and Monte Carlo samples . . . . .	2
3	3	Reconstruction and Base Selection . . . . .	3
4	3.1	Reweighting the pileup distributions in MC . . . . .	3
5	3.2	Trigger . . . . .	3
6	3.3	Lepton Selection . . . . .	4
7	3.4	Jet Selection . . . . .	8
8	3.5	Higgs candidates reconstruction . . . . .	9
9	3.6	Kinematic fit . . . . .	10
10	3.7	B-tagging of jets . . . . .	11
11	4	Performance of the B-tagging selection . . . . .	11
12	5	Final Selection . . . . .	12
13	5.1	Classification of events . . . . .	12
14	5.2	Helicity Discriminant . . . . .	13
15	5.3	MET . . . . .	16
16	5.4	Optimization . . . . .	16
17	5.5	Blinding policy/Unblinding strategy . . . . .	17
18	5.6	Unblinded results . . . . .	22
19	6	Background Determination from Data . . . . .	22
20	6.1	Background estimation from the $m_{JJ}$ sidebands . . . . .	25
21	6.2	$t\bar{t}$ Background Determination from Data . . . . .	29
22	6.3	Diboson background . . . . .	36
23	7	Systematics . . . . .	36
24	7.1	Luminosity uncertainty . . . . .	36
25	7.2	Higgs cross-section and branching fractions . . . . .	36
26	7.3	Uncertainties in the background prediction . . . . .	36
27	7.4	Uncertainties in the signal expectation . . . . .	36
28	8	Signal Shape Parameterization . . . . .	44
29	9	Statistical Analysis and Results . . . . .	45
30	10	Updating the 7-TeV Results . . . . .	51
31	11	Conclusion . . . . .	52
32	A	Neural Network . . . . .	58
33	A.1	Signal Optimization Based on Helicity Neural Network . . . . .	58
34			

## 1 Introduction

The Higgs boson 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 candidate has recently been found with a mass of 125 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 in a several decay modes. This comprehensive effort aims at gaining sensitivity over a large range of Higgs masses by combining many different analyses. We expect that this effort will finally explore the region with  $M_H \geq 2m_Z$  which is not probed at the Tevatron. In this note we report a study of the search for the Higgs boson in  $H \rightarrow ZZ$  when one of  $Z$  decays as  $Z \rightarrow \ell^- \ell^+$  and the other as  $Z \rightarrow q\bar{q}$ . The dominant background is  $Z$ +jets production. Other minor backgrounds are  $t\bar{t}$  and diboson production. The data and Monte Carlo (MC) samples used in this analysis are described in Section 2. The analysis strategy including specific channel optimization and the enhanced signal-to-background selection are discussed from Sections 3 to 5. Contamination is estimated from background-enriched control samples. Methods to provide a data-driven derivation of the background is presented in Section 6. A detailed discussion of the systematic errors, and other possible source of uncertainties is included in Section 7. The strategy was validated using  $4.9 \text{ fb}^{-1}$  data collected by the CMS experiment in 2011 as described in [3, 4]. The analysis mostly follows the analysis on 2011 data, with the following modifications and improvements: Lepton Id and b-tagging are reworked according to POG recommendations. The Quark-Gluon discrimination was removed as uncertainties could not be obtained on short notice. It is planned to extended the analysis to higher higgs masses and treat Higgs production in VBF separately from the gluon fusion process, though these features are not complete yet. An effort has been started to observe the newly discovered boson also in the  $2l2q$  channel, but further work will be necessary to reach sensitivity to that particle, which lies below the kinematic thresholds for the 2011 low mass analysis.

## 2 Data and Monte Carlo samples

We have analyzed  $5.1 \text{ fb}^{-1}$  of proton-proton collisions data at  $\sqrt{s} = 8 \text{ TeV}$  collected in 2012. The data sample was prompt-reco'd with CMSSW\_5.2.X release and comprises runs in the range 190459-193621 (Run2012A) and 193834-195947 (Run2012B) ("ICHEP dataset"). Several primary datasets have been considered for the analysis, according to the channel under investigation, or for dedicated background studies, which are listed in Table 1. We use only lumi-sections that have been declared good for analysis by the central certification team, and provided in the JSON file `Cert.190456-195947.8TeV.PromptReco.Collisions12.JSON.v2.txt`

Table 1: Data samples used for the analysis.

Channel	Dataset
$2\mu 2q$	/DoubleMu/Run2012A-PromptReco-v1/AOD /DoubleMu/Run2012B-PromptReco-v1/AOD
$2e 2q$	/DoubleElectron/Run2012A-PromptReco-v1/AOD /DoubleElectron/Run2012B-PromptReco-v1/AOD
$t\bar{t}$ bkg control sample	/MuEG/Run2012A-PromptReco-v1/AOD /MuEG/Run2012B-PromptReco-v1/AOD

We use Monte Carlo (MC) samples from the Summer12 official CMS production in order to study properties of the SM Higgs boson signal (for a multiplicity of Higgs mass hypotheses  $M_H$ ) and of the relevant background processes. The MC samples are reconstructed with release CMSSW\_5\_2\_X (CMSSW\_5\_0\_X in case of di-boson samples).

The  $H \rightarrow ZZ \rightarrow 2\ell 2q$  signal MC samples ( $\ell = e, \mu, \tau$ ) generated for different  $M_H$  values are listed in Table 2, along with the cross-sections times branching fraction of the process. The SM Higgs cross-sections and the  $H \rightarrow ZZ$  branching fraction are provided by the LHC Higgs XS WG [5, 6], while  $\text{Br}(Z \rightarrow \ell\ell)$ ,  $\ell = e, \mu, \tau$  and  $\text{Br}(Z \rightarrow qq)$  are taken from the PDG [7]. The samples are generated using the POWHEG generator.

The background MC samples are listed in Table 3, along with the SM (N)NLO cross sections taken from the CMS generator group twiki [8], and the equivalent luminosity of the processed events. The dominant Drell-Yan background is studied with an inclusive, high di-lepton mass ( $M_{\ell\ell} > 50 \text{ GeV}$ ),  $Z$ +jets sample generated with the MadGraph generator. Background from top events (mainly  $t\bar{t}$  events) are studied with an inclusive  $t\bar{t}$  sample generated with MadGraph. The background from SM di-boson events is studied with inclusive  $ZZ$ ,  $WZ$ , and  $WW$  samples generated with the Pythia generator.

All the data and Monte Carlo samples used in the analysis have been pre-selected and processed centrally for PAT-tification by employing the common 2l2q skim [9].

### 3 Reconstruction and Base Selection

Input objects to the analysis are reconstructed electrons [10], muons [11], and Particle Flow jets [12, 13] which satisfy kinematic and identification cuts that are described in the following sections.

#### 3.1 Reweighing the pileup distributions in MC

The presence of additional interactions with respect to the primary one, known as pile-up (PU), is expected to deposit additional energy in the trackers and calorimeters, reducing the isolation efficiency of the signal leptons. In order to have an accurate estimate of the signal efficiency, we examine the distributions of the number of true interactions in data and MC. A minimum-bias cross section of 69.4 mb is used as an input to estimate the number of true interactions per bunch crossing in data, following the instruction in Refs. [14, 15]. The distributions of the number of true interactions in data and the un-weighted Summer2012 MC are shown in Fig. 1. In the following this distribution in MC has been re-weighted to match the data. In Fig. 2 we show consistency in the number of reconstructed vertexes between data and MC after MC samples have been re-weighted. The final systematics on the signal efficiency due to pileup reweighing is discussed in Section 7.4.3.

#### 3.2 Trigger

We analyze events in the DoubleMu and DoubleElectron datasets. In each of these datasets there is at least one un-prescaled trigger with looser requirements than our offline selections. Only events which satisfy the lowest-threshold un-prescaled trigger for the dataset are considered for the analysis. The trigger requirements are summarized in Table 4. More details on the trigger strategies for Higgs searches are available in [16]. Since the level of precision of the trigger emulation in simulation is not well known, no trigger is applied on MC samples.

Table 2: POWHEG signal MC samples  $H \rightarrow ZZ \rightarrow 2l2q$ ,  $l = e, \mu, \tau$ .

$M_H$ (GeV)	Name	$\sigma(H \rightarrow 2l2q)$ [pb]
200	/GluGluToHToZZTo2L2Q_M-200.8TeV-powheg-pythia6/Summer12-PU_S7_START52.V9-v1/AODSIM	0.2566
210	/GluGluToHToZZTo2L2Q_M-210.8TeV-powheg-pythia6/Summer12-PU_S7_START52.V9-v1/AODSIM	0.2538
220	/GluGluToHToZZTo2L2Q_M-220.8TeV-powheg-pythia6/Summer12-PU_S7_START52.V9-v1/AODSIM	0.2416
230	/GluGluToHToZZTo2L2Q_M-230.8TeV-powheg-pythia6/Summer12-PU_S7_START52.V9-v1/AODSIM	0.2278
250	/GluGluToHToZZTo2L2Q_M-250.8TeV-powheg-pythia6/Summer12-PU_S7_START52.V9-v1/AODSIM	0.2022
275	/GluGluToHToZZTo2L2Q_M-275.8TeV-powheg-pythia6/Summer12-PU_S7_START52.V9-v1/AODSIM	0.1751
300	/GluGluToHToZZTo2L2Q_M-300.8TeV-powheg-pythia6/Summer12-PU_S7_START52.V9-v1/AODSIM	0.1563
325	/GluGluToHToZZTo2L2Q_M-325.8TeV-powheg-pythia6/Summer12-PU_S7_START52.V9-v1/AODSIM	0.1478
350	/GluGluToHToZZTo2L2Q_M-350.8TeV-powheg-pythia6/Summer12-PU_S7_START52.V9-v1/AODSIM	0.1482
375	/GluGluToHToZZTo2L2Q_M-375.8TeV-powheg-pythia6/Summer12-PU_S7_START52.V9-v1/AODSIM	0.1360
400	/GluGluToHToZZTo2L2Q_M-400.8TeV-powheg-pythia6/Summer12-PU_S7_START52.V9-v1/AODSIM	0.1111
425	/GluGluToHToZZTo2L2Q_M-425.8TeV-powheg-pythia6/Summer12-PU_S7_START52.V9-v1/AODSIM	0.0914
450	/GluGluToHToZZTo2L2Q_M-450.8TeV-powheg-pythia6/Summer12-PU_S7_START52.V9-v1/AODSIM	0.7311
475	/GluGluToHToZZTo2L2Q_M-475.8TeV-powheg-pythia6/Summer12-PU_S7_START52.V9-v1/AODSIM	0.6
500	/GluGluToHToZZTo2L2Q_M-500.8TeV-powheg-pythia6/Summer12-PU_S7_START52.V9-v1/AODSIM	0.4719
525	/GluGluToHToZZTo2L2Q_M-525.8TeV-powheg-pythia6/Summer12-PU_S7_START52.V9-v1/AODSIM	0.0380
550	/GluGluToHToZZTo2L2Q_M-550.8TeV-powheg-pythia6/Summer12-PU_S7_START52.V9-v1/AODSIM	0.0305
575	/GluGluToHToZZTo2L2Q_M-575.8TeV-powheg-pythia6/Summer12-PU_S7_START52.V9-v1/AODSIM	0.025
600	/GluGluToHToZZTo2L2Q_M-600.8TeV-powheg-pythia6/Summer12-PU_S7_START52.V9-v1/AODSIM	0.0201

Instead, proper event weights are assigned to MC events according to the probabilities of lepton candidates to pass the trigger. The trigger efficiency tables for leptons satisfying the same identification criteria as in the analysis are computed in bins of  $(p_T, \eta)$  from data using tag & probe techniques.

### 3.3 Lepton Selection

We reconstruct  $Z \rightarrow ee$  and  $Z \rightarrow \mu\mu$  candidates from pairs of same-flavour, opposite-charge lepton candidates, which satisfy kinematic and identification criteria.

Electron candidates are reconstructed with the GSF algorithm and, in order to assure good

Table 3: Background Monte Carlo samples from 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	Name	$\sigma$ [pb]	lumi [ $\text{fb}^{-1}$ ]
Z+jets (inclusive)	/DYJetsToLL_M-50_TuneZ2Star_8TeV-madgraph-tarball/ Summer12-PU_S7_START52_V9-v2/AODSIM	3503.71	8.7
Z+1 jet (exclusive)	/DY1JetsToLL_M-50_TuneZ2Star_8TeV-madgraph/ Summer12-PU_S7_START52_V9-v1/AODSIM	660.6	36.4
Z+2 jet (exclusive)	/DY2JetsToLL_M-50_TuneZ2Star_8TeV-madgraph/ Summer12-PU_S7_START52_V9-v1/AODSIM	215.1	10.9
Z+3 jet (exclusive)	/DY3JetsToLL_M-50_TuneZ2Star_8TeV-madgraph/ Summer12-PU_S7_START52_V9-v1/AODSIM	65.79	167.4
Z+4 jet (exclusive)	/DY4JetsToLL_M-50_TuneZ2Star_8TeV-madgraph/ Summer12-PU_S7_START52_V9-v1/AODSIM	27.59	232.0
tt	/TTJets_TuneZ2star_8TeV-madgraph-tauola/ Summer12-PU_S7_START52_V5-v1/AODSIM	225.197	29.9
ZZ	/ZZ_TuneZ2star_8TeV_pythia6_tauola/ Summer12-PU_S7_START50_V15-v1/AODSIM	17.654	555
WZ	/WZ_TuneZ2star_8TeV_pythia6_tauola/ Summer12-PU_S7_START50_V15-v1/AODSIM	22.88	437
WW	/WW_TuneZ2star_8TeV_pythia6_tauola/ Summer12-PU_S7_START50_V15-v1/AODSIM	57.1097	175

electron reconstruction, a fiducial cut is applied: the  $\eta$  of the electron supercluster must be inside the ECAL acceptance volume ( $|\eta| < 2.5$ ) but outside the ECAL barrel-endcap overlap region ( $1.4442 < |\eta| < 1.566$ ). Muon candidates must have been reconstructed by both the GlobalMuon and the PF muon reconstruction algorithms and must satisfy the acceptance cut  $|\eta| < 2.4$ .

Electron candidates must satisfy the standard “Loose” working point of the cut-based electron ID for 2012 analyses [17]. The cuts are listed in Table 5 and comprise proper electron identification requirements, an isolation cut, and conversions rejection criteria. Muon candidates must satisfy the standard “Tight” working point of the cut-based muon ID for 2012 analyses [18]. The cuts are listed in Table 6 and comprise proper muon identification requirements plus an

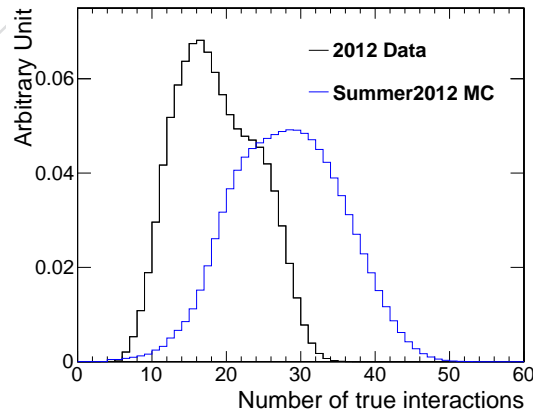


Figure 1: Number of true interactions in 2012 data and Summer2012 MC. In the following the MC sample has been reweighed to match the data.



Table 4: Event trigger requirements for data.

Dataset	trigger requirement
DoubleMu	HLT_Mu17_Mu8    HLT_Mu17_TkMu8
DoubleElectron	HLT_Ele17_CaloIdT_TrkIdVL_CaloIsoVL_TrkIsoVL_ Ele8_CaloIdT_TrkIdVL_CaloIsoVL_TrkIsoVL

129 isolation cut.

Given a lepton candidate, the PF isolation is defined as the sum of “isodeposits” (i.e.  $p_T$  or  $E_T$ ) of charged hadrons ( $I_{ch}$ ), neutral hadrons ( $I_{nh}$ ), and photons ( $I_{ph}$ ), computed in a  $\Delta R$  cone around the lepton direction. In order to assure independence of the isolation from the number of PU interactions, we use a corrected PF isolation definition:

$$I_{PF, corr} = I_{ch}(PFnoPU) + \max(I_{nh} + I_{ph} - \rho \cdot A_{eff}, 0) \quad (1)$$

130 In the above corrected definition we consider only isodeposits from charged hadrons not com-  
 131 ing from PU vertices ( $I_{ch}(PFnoPU)$ ), and subtract an overall PU energy contribution, estimated  
 132 as the average energy density in the event ( $\rho$ ) multiplied by an effective area  $A_{eff}$ . We strictly  
 133 follow the recommendations of egamma and muon POGs. The isolation cone for electrons is  
 134 defined as  $\Delta R < 0.3$ , while for muons is defined as  $\Delta R < 0.4$ . PF muon isolation employs for  
 135 PU correction the energy density  $\rho'$  computed in the central region of the detector after charged  
 136 PU subtraction (“RhoCentralNeutral”, available in AOD). The proper  $A_{eff}$  values are provided  
 137 by the POGs in bins of lepton  $\eta$ . We apply a cut on the relative PF isolation ( $I_{PF, corr} / p_T$ ) as re-  
 138 ported in Tables 5 and 6.

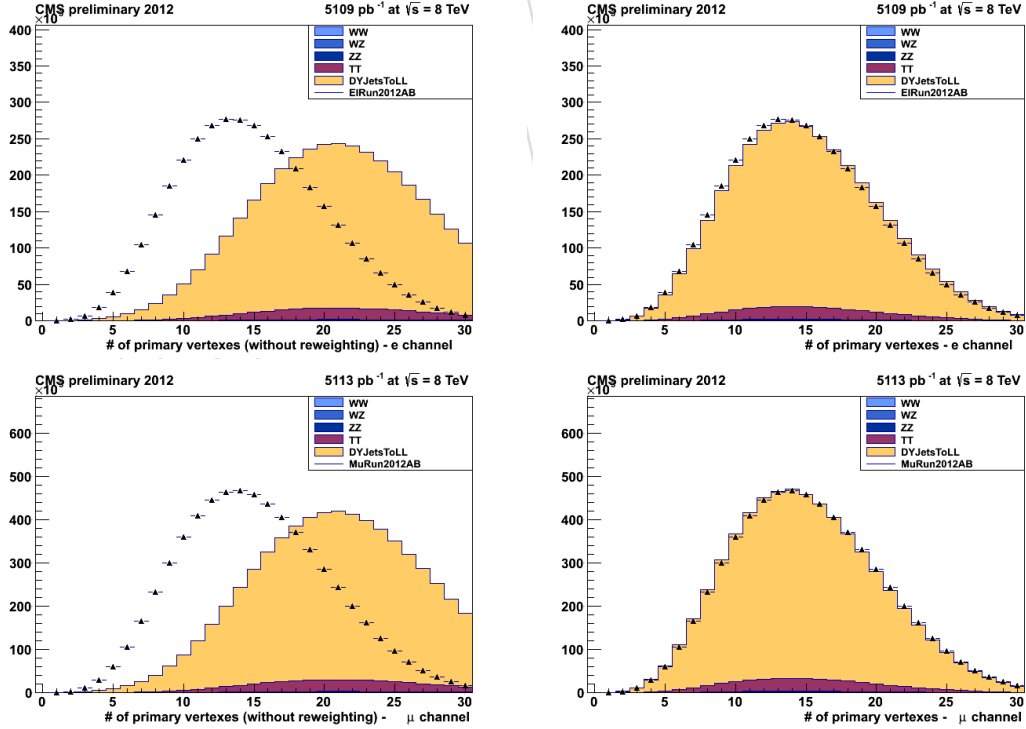


Figure 2: Number of reconstructed vertexes before (left) and after (right) re-weighting the MC sample. Points with error bars show data after loose pre-selection, histograms show contribution of dominant background channels.



Table 5: 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)	$< 2 \text{ mm}$	$< 2 \text{ mm}$
$ 1/E - 1/p $	$< 0.05$	$< 0.05$
$I_{PF,corr}/p_T$	$< 0.15$	$< 0.15$
Missing hits	$\leq 1$	$\leq 1$
Conversion vertex fit prob.	$< 10^{-6}$	$< 10^{-6}$

Table 6: Muon ID requirements for the Tight ID working point.

Variable	Cut
isGlobalMuon	True
isPFMuon	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	$> 5$
$I_{PF,corr}/p_T$	$< 0.12$

The efficiencies for lepton identification (mentioned in table 5 and 6) are evaluated from data using Tag and Probe method which require the reconstruction of di-lepton system with invariant mass to fall in the range [60-120] GeV. One of the lepton is required to pass full selection criteria with a match with the tighter leg of the trigger and is called a tag. The other lepton candidate is selected with criteria which depend on efficiency being measured and this is called a probe. The sample is then 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 events from backgrounds, the signal yields have to be obtained via a fit to invariant mass distribution of the di-lepton system. The measured efficiency is then deduced from relative yields of signal in subsamples with passing or failing probes. This procedure is performed separately in different bins of probe pseudorapidity( $\eta$ ) and transverse momenta ( $p_T$ ) to obtain results as function of probe ( $p_T, \eta$ ). Finally the data to MC scale factors are deduced by dividing efficiencies in data to the ones obtained from MC using exactly same procedure. We use scale factors instead of raw efficiencies, in order to benefit from possible cancellation of systematic uncertainties associated with the procedure. Table 7 and 8 reports the scale factors for electron and muon selection requirements, respectively.

The total efficiency measurement can be factorized into five sequential relative measurements. The five steps are: tracking efficiency, reconstruction efficiency, identification efficiency, isolation efficiency and the online or total trigger efficiency. The total is thus given by the product:

$$\epsilon_{\text{lepton}} = \epsilon_{\text{tracking}} \times \epsilon_{\text{RECO/Tracking}} \times \epsilon_{\text{ID/RECO}} \times \epsilon_{\text{ISO/ID}} \times \epsilon_{\text{Trigger/ISO}} \quad (2)$$

Table 7: Data to MC Scale Factors for Electron Identification requirements

$p_T$	$0.0 <  \eta  < 0.8$	$0.8 <  \eta  < 1.442$	$1.556 <  \eta  < 2.0$	$2.0 <  \eta  < 2.5$
20 - 30	$1.017 \pm 0.004$	$0.998 \pm 0.005$	$1.009 \pm 0.008$	$1.102 \pm 0.010$
30 - 40	$1.019 \pm 0.002$	$1.008 \pm 0.002$	$1.010 \pm 0.004$	$1.066 \pm 0.005$
40 - 50	$1.016 \pm 0.001$	$1.002 \pm 0.002$	$1.009 \pm 0.003$	$1.040 \pm 0.004$
50-200	$1.005 \pm 0.002$	$0.993 \pm 0.003$	$1.003 \pm 0.005$	$1.019 \pm 0.007$

Table 8: Data to MC Scale Factors for Muon Identification requirements

$p_T$	$0.0 <  \eta  < 0.8$	$0.8 <  \eta  < 2.1$	$2.1 <  \eta  < 2.4$
20.00 - 40.00	$1.00425 \pm 0.00042$	$1.00740 \pm 0.00046$	$1.0216 \pm 0.0014$
40.00 - 100.00	$1.00119 \pm 0.00039$	$1.00425 \pm 0.00039$	$1.0140 \pm 0.0014$

Table 9: Working Point Loose to the HLT Ele8 Leg Scale Factors

$\eta$ coverage	$p_T$ range (GeV)	Efficiency (data)	Efficiency (MC)	Data/MC ratio
$0.0 <  \eta  < 0.8$	$10 < p_T < 20$	$47.35\% \pm 0.89\%$	$59.14\% \pm 1.18\%$	$0.801 \pm 0.022$
$0.8 <  \eta  < 1.4$	$10 < p_T < 20$	$34.26\% \pm 0.82\%$	$47.72\% \pm 1.08\%$	$0.718 \pm 0.024$
$1.6 <  \eta  < 2.0$	$10 < p_T < 20$	$44.39\% \pm 1.42\%$	$48.86\% \pm 1.80\%$	$0.909 \pm 0.044$
$2.0 <  \eta  < 2.5$	$10 < p_T < 20$	$45.19\% \pm 1.48\%$	$54.05\% \pm 2.35\%$	$0.836 \pm 0.046$
$0.0 <  \eta  < 0.8$	$20 < p_T < 40$	$98.56\% \pm 0.04\%$	$98.83\% \pm 0.05\%$	$0.997 \pm 0.001$
$0.8 <  \eta  < 1.4$	$20 < p_T < 40$	$93.60\% \pm 0.11\%$	$94.56\% \pm 0.12\%$	$0.990 \pm 0.002$
$1.6 <  \eta  < 2.0$	$20 < p_T < 40$	$90.06\% \pm 0.20\%$	$90.54\% \pm 0.25\%$	$0.995 \pm 0.004$
$2.0 <  \eta  < 2.5$	$20 < p_T < 40$	$94.44\% \pm 0.18\%$	$94.41\% \pm 0.24\%$	$1.000 \pm 0.003$
$0.0 <  \eta  < 0.8$	$40 < p_T < 200$	$99.13\% \pm 0.03\%$	$99.43\% \pm 0.03\%$	$0.997 \pm 0.000$
$0.8 <  \eta  < 1.4$	$40 < p_T < 200$	$97.63\% \pm 0.06\%$	$97.83\% \pm 0.07\%$	$0.998 \pm 0.001$
$1.6 <  \eta  < 2.0$	$40 < p_T < 200$	$94.47\% \pm 0.14\%$	$94.59\% \pm 0.16\%$	$0.999 \pm 0.002$
$2.0 <  \eta  < 2.5$	$40 < p_T < 200$	$96.24\% \pm 0.14\%$	$96.24\% \pm 0.16\%$	$1.000 \pm 0.002$

We reconstruct  $Z \rightarrow ee$  and  $Z \rightarrow \mu\mu$  candidates from pairs of opposite-charge leptons. The leading lepton of the pair must have  $p_T > 40$  GeV, while the next-to-leading lepton must have  $p_T > 20$  GeV. The invariant mass of the pair must be  $70 < m_{\ell\ell} < 110$  GeV. The di-lepton invariant mass for the selected  $Z \rightarrow \ell\ell$  candidates is shown in Fig. 3.

### 3.4 Jet Selection

The PF jets are reconstructed with the anti- $k_T$  algorithm [19] with radius parameter set to  $R = 0.5$ . Jets are required to be inside the tracker acceptance ( $|\eta| < 2.4$ ) thus allowing high reconstruction efficiency and precise energy measurements using PF techniques. Jet-energy corrections are applied to data and MC as explained in [20]. Correction for PU energy is applied at the L1-correction level, by using Fastjet algorithm. In order to remove jets which originate from

Table 10: Working Point Loose to the HLT Ele17 Leg Scale Factors

$\eta$ coverage	$p_T$ range (GeV)	Efficiency (data)	Efficiency (MC)	Data/MC ratio
$0.0 <  \eta  < 0.8$	$10 < p_T < 20$	$95.45\% \pm 0.38\%$	$96.94\% \pm 0.43\%$	$0.985 \pm 0.006$
$0.8 <  \eta  < 1.4$	$10 < p_T < 20$	$85.21\% \pm 0.62\%$	$86.72\% \pm 0.74\%$	$0.983 \pm 0.011$
$1.6 <  \eta  < 2.0$	$10 < p_T < 20$	$83.87\% \pm 1.06\%$	$84.12\% \pm 1.33\%$	$0.997 \pm 0.020$
$2.0 <  \eta  < 2.5$	$10 < p_T < 20$	$86.77\% \pm 1.02\%$	$88.87\% \pm 1.52\%$	$0.976 \pm 0.020$
$0.0 <  \eta  < 0.8$	$20 < p_T < 40$	$98.30\% \pm 0.05\%$	$98.41\% \pm 0.06\%$	$0.999 \pm 0.001$
$0.8 <  \eta  < 1.4$	$20 < p_T < 40$	$93.16\% \pm 0.11\%$	$94.00\% \pm 0.13\%$	$0.991 \pm 0.002$
$1.6 <  \eta  < 2.0$	$20 < p_T < 40$	$89.48\% \pm 0.21\%$	$89.78\% \pm 0.26\%$	$0.997 \pm 0.004$
$2.0 <  \eta  < 2.5$	$20 < p_T < 40$	$93.31\% \pm 0.19\%$	$93.50\% \pm 0.26\%$	$0.998 \pm 0.003$
$0.0 <  \eta  < 0.8$	$40 < p_T < 200$	$98.89\% \pm 0.04\%$	$99.11\% \pm 0.04\%$	$0.998 \pm 0.001$
$0.8 <  \eta  < 1.4$	$40 < p_T < 200$	$97.15\% \pm 0.07\%$	$97.27\% \pm 0.08\%$	$0.999 \pm 0.001$
$1.6 <  \eta  < 2.0$	$40 < p_T < 200$	$93.80\% \pm 0.14\%$	$93.87\% \pm 0.17\%$	$0.999 \pm 0.002$
$2.0 <  \eta  < 2.5$	$40 < p_T < 200$	$95.08\% \pm 0.16\%$	$95.46\% \pm 0.18\%$	$0.996 \pm 0.003$

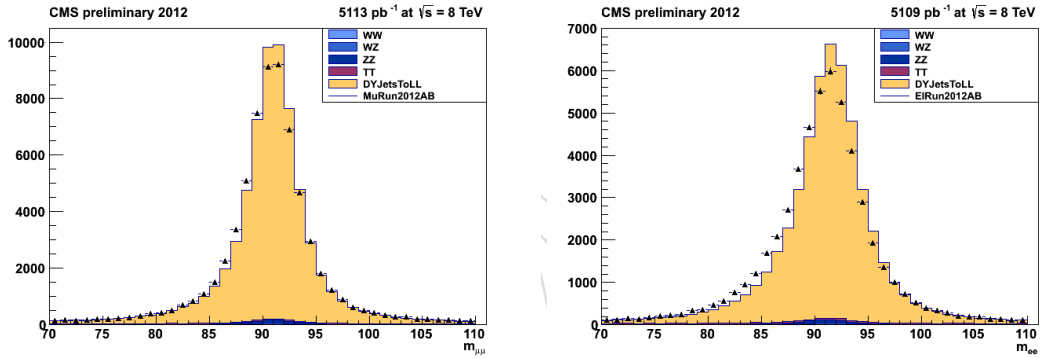


Figure 3: Dilepton invariant mass in data and MC of  $Z \rightarrow \ell\ell$  candidates after lepton selection. Left: Muon channel. Right: Electron channel. The selection described in section 3 is applied but the cut on  $m_{jj}$ .

PU interactions, we select only jets with  $\beta \geq 0.2$ , where  $\beta$  is defined as the sum of transverse momenta of all charged particles in the jet coming from the primary vertex, normalized to the total sum of transverse momenta of all charged particles in the jet (see Subsection 3.1).

We reconstruct  $Z \rightarrow q\bar{q}$  candidates from jet-jet pairs. In order to reject fake candidates made by low- $p_T$  jets from QCD background, we require that both jets of the pair must have  $p_T > 30$  GeV.

### 3.5 Higgs candidates reconstruction

All the di-lepton and di-jet pairs satisfying the above requirements are combined to form  $\ell\ell jj$  candidates. In addition, a  $\Delta R > 0.5$  cut is applied between each lepton and jet within a candidate in order to avoid double counting of the same object reconstructed in different collections (for instance leptons inside a jet). In the following, we refer to the entire selection procedure described above as “pre-selection” of  $\ell\ell jj$  candidates.

The distribution of the di-jet invariant mass  $m_{jj}$  for  $\ell\ell jj$  candidates reconstructed in the electron and muon channels is shown in Fig. 4. In order to suppress the dominant  $Z$ +jets background,

we select only candidates in a signal region of  $m_{jj}$ , defined as  $75 < m_{jj} < 105$  GeV.

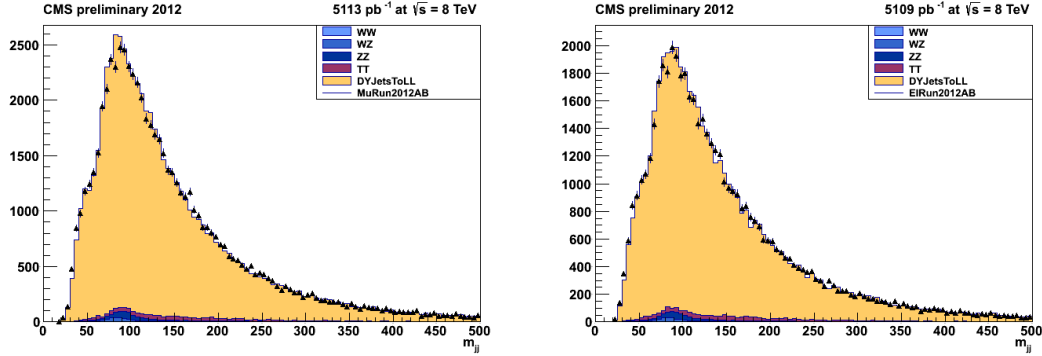


Figure 4: Dijet invariant mass in data and MC for  $\ell\ell jj$  candidates which have passed the pre-selection. Left: Muon channel. Right: Electron channel. The selection described in section 3.

### 3.6 Kinematic fit

Finite resolution of the jet energy is the dominant source of uncertainty in both the di-jet invariant mass  $m_{jj}$  and di-boson invariant mass  $m_{ZZ}$  for Higgs candidates. Therefore, the two variables become highly correlated, as can be seen in results from 2011 shown in Fig. 5 (left). In order to optimally scale the di-jet quadrimomentum to the Z boson mass, we use a kinematic fit to the two jets, as discussed in detail in the analysis notes of previous iterations of this analysis [3, 21]. The fit is provided with parametrizations of jet transverse momentum and angular resolutions as functions of transverse momentum and pseudorapidity, and therefore constrains the mass of the di-jet system to the value of the Z boson mass by modifying the jet fourmomenta in accordance to their expected resolutions. This brings a further improvement in the resolution on the signal invariant mass, as was shown in the 2011 analysis (see [3, 21] for details). The kinematic fit to the di-jet system also removes the correlation between the di-jet and di-boson invariant mass in signal as can be seen in Fig. 5 (right). This allows a straightforward definition of signal and sideband regions, through simple addition of one-variable interval cuts.

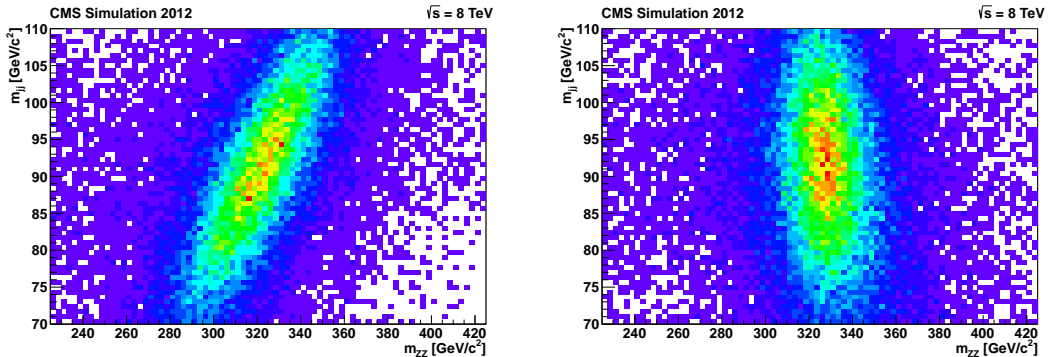


Figure 5: Kinematic fit results. Di-jet invariant mass vs. di-boson invariant mass for Higgs candidates (signal MC,  $m_H = 325$  GeV) after loose selection requirements. Left: before kinematic fit; right: after kinematic fit.

### 3.7 B-tagging of jets

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, we use a b-tagging algorithm in order to identify jets originating from heavy-flavour quarks. However, no selection of candidates based on b-probabilities is employed in the analysis. Instead, the b-tagging information is used to classify  $\ell\ell jj$  candidates into three categories, each one characterized by a certain signal-to-background ratio, and an optimized selection is applied on candidates belonging to different categories. In this way we exploit as much information as possible from our data, thus increasing the analysis sensitivity. Details of the b-tagging algorithm are described in section 4 and candidates categories are defined in section 5.1

## 4 Performance of the B-tagging selection

Due to the relatively large branching fraction of the Z boson decaying into a pair of bottom-antibottom quarks, it is advantageous to exploit the possibility to identify jets originating from heavy-flavour quarks in order to enhance the sensitivity of the analysis.

For this purpose, the reconstructed jets in the event are checked for the possible heavy-flavour content and check whether it can be “tagged” as originating from a bottom quark. In a previous version of the analysis, this “b-tagging” was performed by using the so-called Track-Counting Discriminant [22], but since that tagger is no longer recommended for physics analysis, in this generation of the analysis we have moved to the Jet-Probability (JP) Discriminant.

The JP tagger is based on track impact parameters: it entails computing the compatibility of a set of tracks associated to the jet to have come from the primary vertex. If the probability to come from the primary vertex is low, the jet is likely to be a b-jet. The jet probability estimation is performed through several steps: for each track, the probability to come from the primary vertex is computed and these probabilities are combined to provide the jet probability. The track probability distribution is calibrated by means of the distribution of track impact parameters with negative signs. The negative part of the impact parameter distribution is used for this purpose because it is mainly made up of primary vertex tracks due to detector resolution effects. The advantage of this method with respect to track counting is the fact that a single discriminator is used and that information from all tracks is used at the same time.

The JP tagger has been validated in our selection using a control region that is completely dominated by background. This control region is defined by selecting 2 leptons and 2 jets (using the analysis object definition), requiring  $70 \text{ GeV} < m_{\ell\ell} < 110 \text{ GeV}$ , and one jet passing the tagger. Since we are mostly interested in the “medium” and “loose” version of the tagger, we have validated those versions and confirm that the performance is similar to that of the Track-Counting tagger used previously.

Tables 11 and 12 show the number of events selected with the “loose” working point in the control region using TCHE and JP taggers, and separated in flavor categories. The JP tagger provides similar performance tagging b-jets while reducing the amount of light jets in our major background, Z+jets production.

The addition of the tagging requirement is a fundamental part used to classify the final events in order to use them in a more optimized way to enhance the sensitivity to signal. This is described in Section 5.1.

Muon Events -  $4.6 \text{ fb}^{-1}$ 

	TCHEL			JPL		
	b-jets	c-jets	udsg	b-jets	c-jets	udsg
$t\bar{t}$	668.7	6.3	37.2	686.9	5.0	23.1
Z+jets	1647.8	1314.3	5482.42	1657.6	1317.6	4107.4
Total	2346.5	1359.3	5570.2	2375.2	1364.	4170.0
Higgs (400)	70.6	33.9	52.0	71.0	31.7	32.5

Table 11: Number muon events normalized to  $4.6 \text{ fb}^{-1}$  requiring one jet passing TCHE or JP “loose” working point. Events are break down in flavor categories and background samples, where total includes diboson events.

Electron Events-  $4.6 \text{ fb}^{-1}$ 

	TCHEL			JPL		
	b-jets	c-jets	udsg	b-jets	c-jets	udsg
$t\bar{t}$	598.1	8.2	39.6	607.2	6.6	22.1
Z+jets	1432.4	1131.2	4618.9	1455.7	1177.3	3408.4
Total	2056.9	1170.9	4701.5	2089.6	1217.1	3466.3
Higgs (400)	62.5	29.5	46.7	63.1	27.9	29.0

Table 12: Number electron events normalized to  $4.6 \text{ fb}^{-1}$  requiring one jet passing TCHE or JP “loose” working point. Events are break down in flavor categories and background samples, where total includes diboson events.

## 5 Final Selection

After the basic selection and the specific selections aiming to enhance the sensitivity for events with b jets and for VBF production, a final selection is performed to optimize the discrimination power for a Higgs signal.

### 5.1 Classification of events

Once the events are preselected as described in the previous sections, they are classified accordingly to the tagging-content of the jets associated to the Higgs candidate. Events are classified as “two tags” if the two jets are b-tagged with at least one medium-tag (see Section 4 for details) and one loose-tag. Events failing that requirement are classified as “one tag” if at least one of the two jets satisfy the loose-tag condition. Events failing this requirement are finally classified as “no-tag”.

As these three categories present very different signal-to-background ratio and display some differences in the object description, they are treated separately in the final optimization described in this section. In addition, some differences in the cuts are also needed due to the different background composition, especially in the case of the “two tags” sample, in which the  $t\bar{t}$  background starts to be noticeable.

On the other hand, since the optimization is based mostly on the intrinsic properties of the final state under investigation, there are also parts that are shared by the three categories. Also the methods used to estimate the backgrounds and obtain the final signal-background discriminant are common to all the categories.



Table 13: List of expected background and signal yields in all the channels with  $5.1 \text{ fb}^{-1}$  of data after all selection and within the  $ZZ$  invariant mass range  $[183, 1000]$

	0 b-tag yields	1 b-tag yields	2 b-tag yields
observed yield	$4479.00 \pm 66.93$	$0.00 \pm 66.93$	$0.00 \pm 66.93$
expected background (data)	$4303.20 \pm 60.89$	$3277.98 \pm 56.69$	$251.46 \pm 16.22$
expected background (MC)	$\pm$	$\pm$	$\pm$
200 GeV	$146.80 \pm 4.75 \pm 1.08$	$215.27 \pm 6.23 \pm 0.57$	$17.37 \pm 0.86 \pm 0.37$
250 GeV	$799.00 \pm 16.52 \pm 1.08$	$410.71 \pm 9.89 \pm 0.57$	$39.82 \pm 1.81 \pm 0.37$
300 GeV	$1182.40 \pm 19.48 \pm 1.08$	$470.70 \pm 10.01 \pm 0.57$	$62.58 \pm 2.18 \pm 0.37$
350 GeV	$1598.04 \pm 23.17 \pm 1.08$	$583.28 \pm 11.19 \pm 0.57$	$99.39 \pm 3.06 \pm 0.37$
400 GeV	$1216.67 \pm 17.20 \pm 1.08$	$395.32 \pm 7.36 \pm 0.57$	$86.78 \pm 2.49 \pm 0.37$
500 GeV	$257.05 \pm 3.70 \pm 1.08$	$77.09 \pm 1.49 \pm 0.57$	$21.61 \pm 0.61 \pm 0.37$
600 GeV	$44.23 \pm 0.72 \pm 1.08$	$14.13 \pm 0.32 \pm 0.57$	$5.55 \pm 0.17 \pm 0.37$

## 5.2 Helicity Discriminant

There are several features in the signal  $H \rightarrow ZZ \rightarrow 2\ell 2j$  decay kinematics which discriminate it against background. We can exploit these kinematic differences to optimize selection and maximize signal significance or exclusion power. In the nominal approach we fully explore kinematics in the decay with five angles which characterize the decay.

It has been shown in Refs. [23, 24] that five angular observables fully describe kinematics in the decay  $2 \rightarrow 1 \rightarrow 2 \rightarrow 4$  as in  $ab \rightarrow X \rightarrow ZZ \rightarrow 2\ell 2j$ , and they are orthogonal observables to the three invariant masses of the  $X$  and the two  $Z$ . We should note that 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.

In Fig. 6 we illustrate the angular distribution in the production and decay chain  $ab \rightarrow X \rightarrow P_1 P_2 \rightarrow p_{11} p_{12} p_{21} p_{22}$  with an example of the  $ab \rightarrow X \rightarrow ZZ \rightarrow 4\ell$  or  $2\ell 2q$  (where quarks  $q$  hadronize to jets, which we refer to as  $2\ell 2j$  channel later) chain with two partons  $a$  and  $b$ , such as  $gg$  or  $q\bar{q}$ . The angular distribution can be expressed as a function of three helicity angles  $\theta_1$ ,  $\theta_2$ , and  $\Phi$ , and two production angles  $\theta^*$  and  $\Phi_1$ , as shown in Fig. 6. More details can be found in Refs. [23, 25], where parameterization of both signal and background distributions have been derived and implemented.

Here  $\theta_i$  is the angle between the direction of the  $\ell^-$  or  $q$  from the  $Z \rightarrow \ell^+ \ell^-$  or  $q\bar{q}$  (where the quark-antiquark pair produces two jets) and the direction opposite the  $X$  in the  $Z$  rest frame, and  $\Phi$  is the angle between the decay planes of the two  $Z$  systems. The two  $Z$ 's are distinguished by their decay type or, in case their daughters are the same type of particles, by an arbitrary convention. The production angle  $\theta^*$  is defined as the angle between the parton collision axis  $z$  and the  $X$  decay axis in the  $X$  rest frame. The fifth angle can be defined as  $\Phi_1$ , an angles between the production plane and the first  $Z$  decay plane.

A comparison of angular distribution in data and MC for electron and muon events can be found in Figs. 7 and 8, where we observe good agreement, in samples dominated by background.

Previous work in Ref. [25] has concentrated on using angular information with a likelihood method for extracting signal and background yields from fits. Here we have chosen to adapt



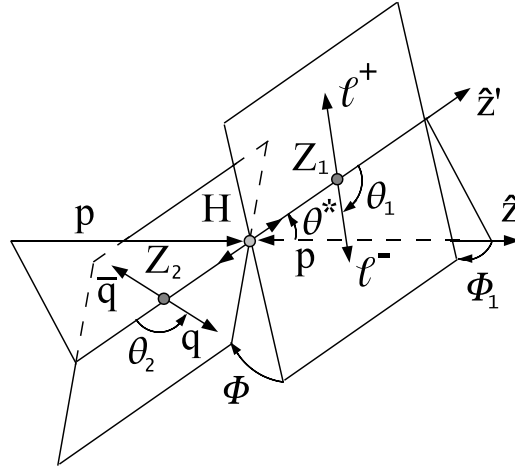


Figure 6: Diagram depicting the decay  $X \rightarrow ZZ \rightarrow 2l2q$  and the 5 angles which describe such a decay.

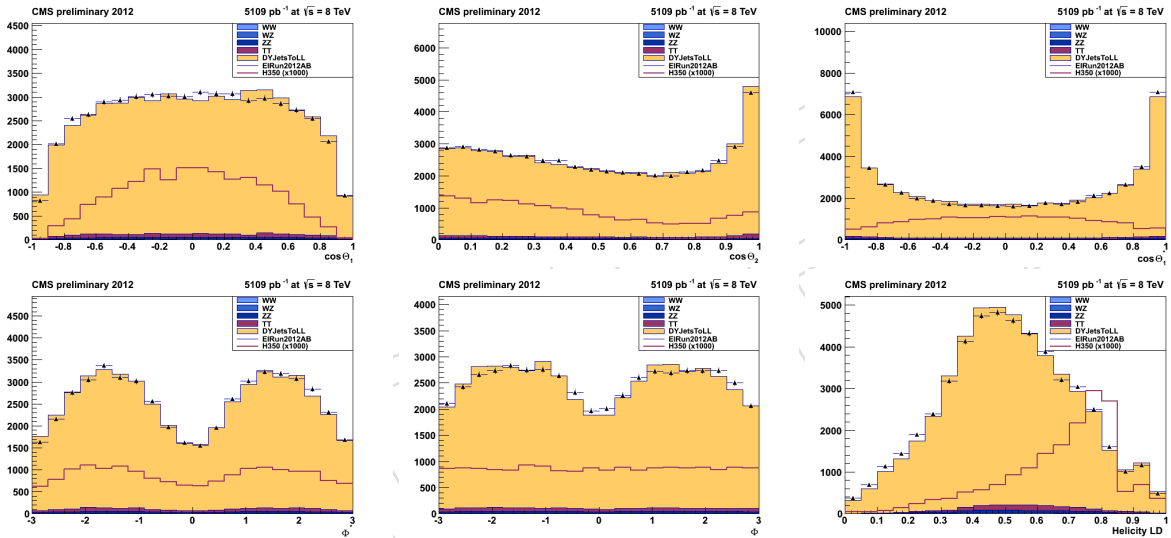


Figure 7: Five angular distributions of  $\cos \theta_1$ ,  $\cos \theta_2$ ,  $\cos \theta^*$ ,  $\Phi$ ,  $\Phi_1$  and the helicity likelihood discriminant for 2012 electron data (points) and Summer 12 Monte Carlo samples (histogram). Open histograms indicate the expected distribution for a Higgs boson with mass 350 GeV, multiplied by a factor of 1000 for illustration. The selection described in section 3 is applied but the cut on  $m_{jj}$ .

this information into a cut-and-count approach by, instead, building a likelihood discriminant from the angular distributions. While some statistical power is lost in reducing the MVA likelihood fit to a 1D discriminant, we gain in simplicity. For example, even if some parameterization of either signal or background effect is not perfect in the likelihood parameterization, the analysis is still not biased, it is only slightly less optimal than with the perfect description.

Assuming the probability distributions of the five helicity angles for both signal and background are known,  $P_{sig}$  and  $P_{bkg}$  respectively, the likelihood discriminant is given by the probability ratio

$$LD = \frac{P_{sig}}{P_{sig} + P_{bkg}}.$$

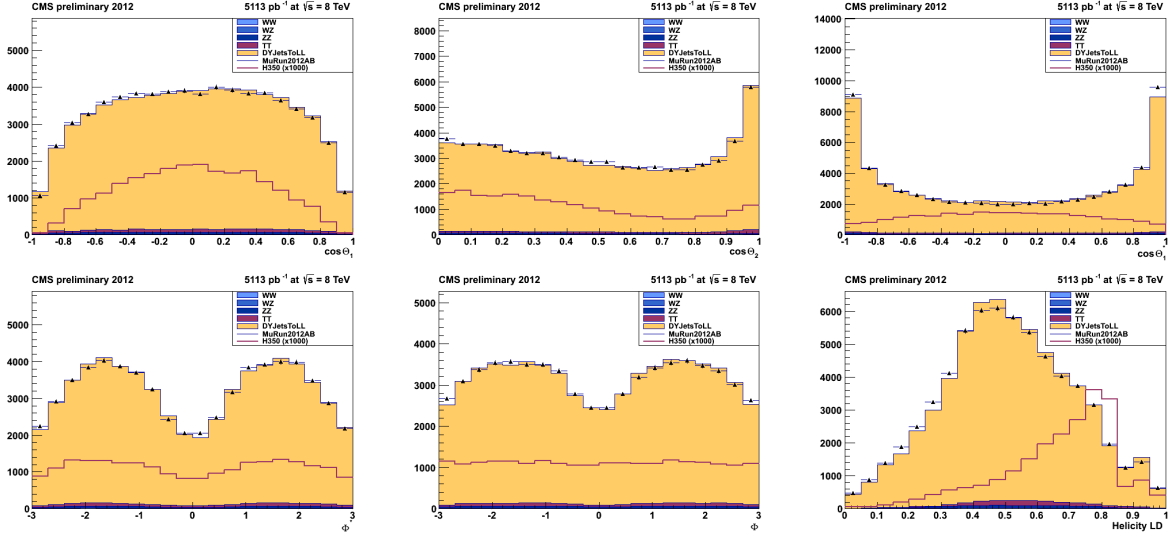


Figure 8: Five angular distributions of  $\cos\theta_1, \cos\theta_2, \cos\theta^*, \Phi, \Phi_1$  and the helicity likelihood discriminant for 2012 muon data (points) and Summer 12 Monte Carlo samples (histogram). Open histograms indicate the expected distribution for a Higgs boson with mass 350 GeV, multiplied by a factor of 1000 for illustration. The selection described in section 3 is applied but the cut on  $m_{jj}$ .

This function has the feature that the signal is most likely to have values close to one and the background is most likely to have values close to zero. Events are then selected by requiring LD to be above certain threshold. This method of selection has been shown to provide at least similar results to those obtained using a set of optimized kinematic cuts. Furthermore, since the helicity angles are largely decoupled from the mass variables, by making selections based on the helicity likelihood discriminant one can better preserve the shape of the background's ZZ invariant mass distribution than with tight kinematic cuts. The method for obtaining such a likelihood discriminant is described below.

The probability distribution function for signal was taken to be a product of the ideal, fully correlated, distribution which is derived in Ref. [23] and a set of four one-dimensional acceptance functions.

$$\mathcal{P}_{\text{sig}} = \mathcal{P}_{\text{ideal}}(\theta^*, \theta_1, \theta_2, \Phi, \Phi_1; M_{ZZ}) \cdot \mathcal{G}_{\theta^*}(\theta^*; M_{ZZ}) \cdot \mathcal{G}_{\theta_1}(\theta_1; M_{ZZ}) \cdot \mathcal{G}_{\theta_2}(\theta_2; M_{ZZ}) \cdot \mathcal{G}_{\Phi_1}(\Phi_1; M_{ZZ})$$

The four acceptance functions,  $\mathcal{G}_{\theta^*}, \mathcal{G}_{\theta_1}, \mathcal{G}_{\theta_2}$ , and  $\mathcal{G}_{\Phi_1}$ , have been obtained empirically from fits to Monte Carlo Fall11 [3]. We do not expect significant change in MC Summer 12 for the analyzed range of mass. However, to extend the analysis to higher masses, the acceptance functions will have to be reevaluated in the new kinematic regime. Projections of  $\mathcal{P}_{\text{sig}}$  can be seen in Fig. 9.

Where as the ideal function,  $\mathcal{P}_{\text{ideal}}$ , is naturally parameters with the ZZ invariant mass, the parameters of the four acceptance functions have all been re-parameterized in terms of  $m_{ZZ}$  only. This was done by fitting eight different Monte Carlo samples each corresponding to a different Higgs mass and then fitting the resulting parameters with either a linear or quadratic function of  $m_{ZZ}$ .

The probability distribution function for the background was approximated with a product of five one-dimensional functions.

$$\mathcal{P}_{\text{bkg}}(\theta^*, \theta_1, \theta_2, \Phi, \Phi_1; m_{ZZ}) = \mathcal{P}_{\theta^*}(\theta^*; m_{ZZ}) \cdot \mathcal{P}_{\theta_1}(\theta_1; m_{ZZ}) \cdot \mathcal{P}_{\theta_2}(\theta_2; m_{ZZ}) \cdot \mathcal{P}_{\Phi}(\Phi; m_{ZZ}) \cdot \mathcal{P}_{\Phi_1}(\Phi_1; m_{ZZ})$$

All functions were obtained empirically from fits to Monte Carlo [3] in 2011. The background shapes are not expected to change from 2011 to 2012. Projections of  $\mathcal{P}_{\text{bkg}}$  can be found in Fig. 10.

Similar to the case of  $\mathcal{P}_{\text{sig}}$ , the background Monte Carlo was divided into bins of  $m_{ZZ}$  and each bin was fit with  $\mathcal{P}_{\text{bkg}}$ . The parameters from each fit were then fit using either linear or quadratic functions of  $m_{ZZ}$ .

Combining  $\mathcal{P}_{\text{sig}}$  and  $\mathcal{P}_{\text{bkg}}$  into LD, we end up with the discriminant that is a function of the five helicity angles and parameterized by a given event's ZZ invariant mass. The helicity likelihood discriminant is plotted in Fig. 7 and 8 for both background and signal Monte Carlo after the basic kinematic selections (i.e. before the classification by b-tag category).

There is a good agreement in the likelihood discriminant (LD) distribution between data and background MC, as it is expected based on agreement of variables entering the LD calculation.

### 5.3 MET

Without further selection, in the 2  $b$ -tag category the dominant background originates from  $t\bar{t}$  decay chain which contains two true  $b$  quark jets. We further reduce this background with a particle flow MET significance requirement of less than 10.

The distribution of PF MET significance for data and MC background are shown in Fig. 11. The requirement of this variable is placed in the tail of the distribution, which makes the disagreement between data and MC background prediction introduce a negligible systematic uncertainty compared to other sources, reason why it is not considered in the analysis.

This rather loose requirement is expected to be robust even in the presence of PileUp.

### 5.4 Optimization

The requirement on the likelihood discriminant has been optimized for each  $b$ -tag category as function of the reconstructed diboson invariant mass,  $m_{ZZ}$ . The optimization procedure has

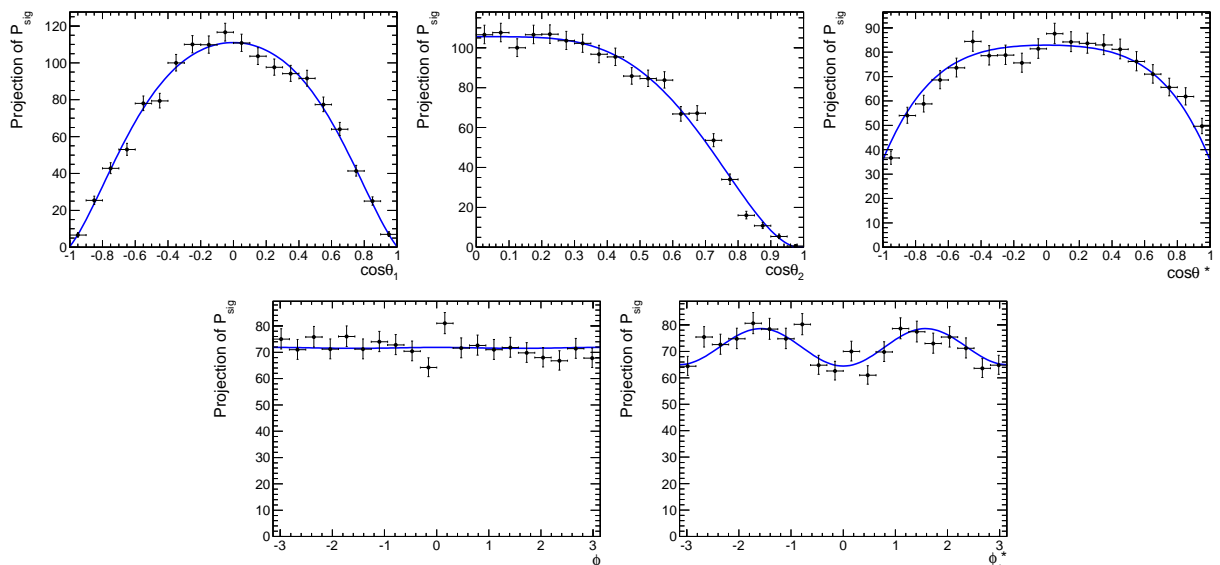


Figure 9: Distributions of  $\cos \theta_1$ ,  $\cos \theta_2$ ,  $\cos \theta^*$ ,  $\Phi$ , and  $\Phi_1$  for a 500 GeV Higgs boson.

Figure 10: Distributions of  $\cos \theta_1, \cos \theta_2, \cos \theta^*, \Phi$ , and  $\Phi_1$  for background around 500 GeV.Figure 11: PF MET significance comparison between data and MC simulation for electron (left) and muon (right) channels. A requirement of PF MET significance less than 10 is demanded in the 2  $b$ -tag category.

been performed for the analysis of data collected in 2011 [3]. The mass dependence of the LD cut was optimized up to  $\sqrt{s} = 600$  GeV. To extend the analysis beyond this mass range will require an additional optimization in the added kinematic range. The same optimized selection is used for the analysis presented in this note. In Table 14 the final optimized selection requirements are listed for the three  $b$ -tag categories.

Table 14: Optimized selection requirements in the three  $b$ -tag categories.

	preselection		
	0 $b$ -tag	1 $b$ -tag	2 $b$ -tag
$b$ -tag	none	JPL	JPM & JPL
Helicity LD	$> (0.55 + 0.00025 \times m_{ZZ})$	$> (0.302 + 0.000656 \times m_{ZZ})$	$> 0.5$
pf MET significance	–	–	$< 10$
$m_{jj}$	[75,105] GeV		
$m_{\ell\ell}$	[70,110] GeV		
$p_T(I^\pm)$	$> 20/40$ GeV		
$p_T(\text{jets})$	$> 30$ GeV		
$ \eta (\ell^\pm)$	$(e^\pm) < 2.5, (\mu^\pm) < 2.4$		
$ \eta (\text{jets})$	$< 2.4$		
lepton quality	see Section 3.3		
jet quality	see Section 3.4		

The expected yield (corresponding to  $1 \text{ fb}^{-1}$  of data) of signal and background events in the full  $ZZ$  invariant mass range  $[0,1000]$  GeV is listed in Table 15 for the electron channel and in Table 16 for the muon channel. Further discrimination of the signal and background will be achieved with the likelihood fit which uses  $m_{ZZ}$  invariant mass distributions.

For further illustration of results, the various background contributions are shown in several invariant mass ranges corresponding to the mass window  $(-6\%, +10\%)$  around the signal mass hypothesis. The expected yields for each  $b$ -tag category and each leptonic channel are shown in Tables from 17 to 22

## 5.5 Blinding policy/Unblinding strategy

In accordance with the Higgs Group proposal in 2012 about keeping the analyses “blind” to sensitive selections, we have implemented a blinding policy and set a well-defined unblinding strategy after the preapproval.

Basically the only distribution that could bias our approach to the analysis is the full invariant mass of the two  $Z$ , which is directly related to the Higgs mass. Therefore we have decided that the distribution itself is to be kept blind as long as possible.

By general motivation, we consider to solely use the events in the hadronic  $Z$  window (see Section 5.4) for any study. However, due to the nature of our background estimation, the dis-

Table 15: List of expected background and signal yields in the electron channel with  $1 \text{ fb}^{-1}$  of data after all selection and within the ZZ invariant mass range [0,1000]

	0 b-tag yields	1 b-tag yields	2 b-tag yields
Background	$1951.83 \pm 34.43$	$1386.68 \pm 30.51$	$94.14 \pm 8.67$
200 GeV	$6.64 \pm 0.32$	$7.47 \pm 0.37$	$1.55 \pm 0.16$
210 GeV	$8.77 \pm 0.40$	$7.96 \pm 0.40$	$1.62 \pm 0.16$
220 GeV	$12.10 \pm 0.46$	$9.22 \pm 0.39$	$2.78 \pm 0.23$
230 GeV	$13.13 \pm 0.46$	$9.95 \pm 0.41$	$2.69 \pm 0.21$
250 GeV	$15.65 \pm 0.49$	$10.24 \pm 0.38$	$2.83 \pm 0.19$
275 GeV	$17.98 \pm 0.49$	$11.61 \pm 0.40$	$3.61 \pm 0.22$
275 GeV	$17.98 \pm 0.49$	$11.61 \pm 0.40$	$3.61 \pm 0.22$
300 GeV	$19.25 \pm 0.47$	$11.45 \pm 0.36$	$3.88 \pm 0.20$
325 GeV	$20.03 \pm 0.46$	$12.14 \pm 0.36$	$4.90 \pm 0.24$
350 GeV	$22.53 \pm 0.49$	$13.30 \pm 0.38$	$5.44 \pm 0.25$
375 GeV	$21.56 \pm 0.47$	$13.32 \pm 0.37$	$5.57 \pm 0.24$
400 GeV	$19.53 \pm 0.41$	$11.05 \pm 0.31$	$5.10 \pm 0.22$
450 GeV	$13.38 \pm 0.30$	$8.10 \pm 0.24$	$3.68 \pm 0.16$
500 GeV	$8.48 \pm 0.18$	$4.52 \pm 0.13$	$2.11 \pm 0.09$
550 GeV	$5.17 \pm 0.13$	$2.98 \pm 0.10$	$1.45 \pm 0.07$
600 GeV	$3.27 \pm 0.08$	$1.77 \pm 0.06$	$0.96 \pm 0.04$

Table 16: List of expected background and signal yields in the muon channel with  $1 \text{ fb}^{-1}$  of data after all selection and within the ZZ invariant mass range [0,1000]

	0 b-tag yields	1 b-tag yields	2 b-tag yields
Background	$2607.14 \pm 44.59$	$1818.05 \pm 38.83$	$125.31 \pm 7.66$
200 GeV	$8.12 \pm 0.39$	$9.69 \pm 0.45$	$2.37 \pm 0.22$
210 GeV	$10.90 \pm 0.46$	$10.42 \pm 0.47$	$2.06 \pm 0.19$
220 GeV	$14.80 \pm 0.52$	$12.04 \pm 0.48$	$2.90 \pm 0.22$
230 GeV	$16.50 \pm 0.53$	$12.63 \pm 0.47$	$3.77 \pm 0.25$
250 GeV	$19.23 \pm 0.53$	$13.54 \pm 0.47$	$4.14 \pm 0.26$
275 GeV	$23.25 \pm 0.58$	$14.20 \pm 0.45$	$4.50 \pm 0.24$
275 GeV	$23.25 \pm 0.58$	$14.20 \pm 0.45$	$4.50 \pm 0.24$
300 GeV	$23.94 \pm 0.53$	$15.45 \pm 0.44$	$5.22 \pm 0.24$
325 GeV	$25.27 \pm 0.53$	$15.58 \pm 0.42$	$5.76 \pm 0.26$
350 GeV	$27.85 \pm 0.54$	$17.13 \pm 0.44$	$6.73 \pm 0.28$
375 GeV	$28.14 \pm 0.54$	$16.30 \pm 0.41$	$6.60 \pm 0.27$
400 GeV	$25.36 \pm 0.48$	$13.82 \pm 0.35$	$6.57 \pm 0.25$
450 GeV	$17.69 \pm 0.36$	$10.01 \pm 0.27$	$4.46 \pm 0.18$
500 GeV	$10.90 \pm 0.21$	$5.49 \pm 0.14$	$2.80 \pm 0.11$
550 GeV	$6.62 \pm 0.15$	$3.69 \pm 0.11$	$1.82 \pm 0.08$
600 GeV	$4.22 \pm 0.09$	$2.12 \pm 0.06$	$1.21 \pm 0.05$

Table 17: Expected yields of signal and background with  $1 \text{ fb}^{-1}$  based on simulation in the 0 b-tag category. The numbers show  $2e2j$  expectations. Tighter  $m_{ZZ}$  mass requirements are applied as (-6%, +%) of the mass hypothesis.

Mass [GeV]	Signal	Z+Jets	$t\bar{t}$	ZZ/WZ/WW	TotalBkg
200	$4.97 \pm 0.28$	$371.23 \pm 17.53$	$2.29 \pm 0.94$	$16.09 \pm 0.90$	$389.61 \pm 17.57$
210	$6.99 \pm 0.36$	$468.71 \pm 19.08$	$2.23 \pm 0.99$	$19.98 \pm 0.98$	$490.92 \pm 19.13$
220	$9.73 \pm 0.43$	$489.28 \pm 18.71$	$4.00 \pm 1.19$	$20.23 \pm 0.96$	$513.50 \pm 18.77$
230	$10.27 \pm 0.41$	$518.79 \pm 19.46$	$3.98 \pm 1.03$	$19.96 \pm 0.93$	$542.73 \pm 19.51$
250	$12.28 \pm 0.44$	$487.10 \pm 18.34$	$3.90 \pm 0.89$	$19.48 \pm 0.90$	$510.49 \pm 18.38$
275	$13.69 \pm 0.44$	$386.88 \pm 13.10$	$6.83 \pm 1.73$	$18.42 \pm 0.87$	$412.12 \pm 13.24$
275	$13.69 \pm 0.44$	$386.88 \pm 13.10$	$6.83 \pm 1.73$	$18.42 \pm 0.87$	$412.12 \pm 13.24$
300	$14.15 \pm 0.40$	$297.59 \pm 12.79$	$5.02 \pm 1.84$	$15.23 \pm 0.81$	$317.85 \pm 12.95$
325	$14.64 \pm 0.40$	$239.89 \pm 12.42$	$2.24 \pm 1.27$	$12.06 \pm 0.74$	$254.19 \pm 12.50$
350	$15.94 \pm 0.41$	$197.54 \pm 11.81$	$1.47 \pm 0.57$	$10.15 \pm 0.68$	$209.16 \pm 11.84$
375	$14.37 \pm 0.39$	$143.53 \pm 7.85$	$1.18 \pm 0.51$	$9.04 \pm 0.68$	$153.75 \pm 7.89$
400	$12.72 \pm 0.33$	$114.00 \pm 6.45$	$0.51 \pm 0.26$	$7.50 \pm 0.63$	$122.01 \pm 6.49$
450	$7.81 \pm 0.23$	$74.37 \pm 5.30$	$0.17 \pm 0.17$	$4.38 \pm 0.43$	$78.92 \pm 5.32$
500	$4.79 \pm 0.14$	$43.93 \pm 2.58$	$0.17 \pm 0.17$	$3.48 \pm 0.39$	$47.58 \pm 2.62$
550	$2.78 \pm 0.10$	$33.59 \pm 3.72$	$0.00 \pm 0.00$	$2.61 \pm 0.33$	$36.21 \pm 3.73$
600	$1.49 \pm 0.05$	$21.98 \pm 3.80$	$0.00 \pm 0.00$	$1.74 \pm 0.23$	$23.72 \pm 3.81$

Table 18: Expected yields of signal and background with  $1 \text{ fb}^{-1}$  based on simulation in the 0 b-tag category. The numbers show  $2\mu 2j$  expectations. Tighter  $m_{ZZ}$  mass requirements are applied as (-6%, +%) of the mass hypothesis.

Mass [GeV]	Signal	Z+Jets	$t\bar{t}$	ZZ/WZ/WW	TotalBkg
200	$6.52 \pm 0.36$	$488.09 \pm 22.98$	$2.71 \pm 1.23$	$20.35 \pm 1.03$	$511.15 \pm 23.03$
210	$8.66 \pm 0.42$	$597.36 \pm 22.91$	$6.72 \pm 2.13$	$26.50 \pm 1.17$	$630.58 \pm 23.03$
220	$11.97 \pm 0.48$	$630.88 \pm 22.71$	$8.58 \pm 2.30$	$26.95 \pm 1.15$	$666.41 \pm 22.86$
230	$13.37 \pm 0.49$	$655.24 \pm 22.62$	$10.40 \pm 2.39$	$26.31 \pm 1.10$	$691.95 \pm 22.77$
250	$15.00 \pm 0.47$	$649.27 \pm 23.15$	$11.43 \pm 2.71$	$24.84 \pm 1.06$	$685.54 \pm 23.33$
275	$17.39 \pm 0.49$	$550.19 \pm 22.27$	$7.52 \pm 2.14$	$22.88 \pm 1.02$	$580.59 \pm 22.39$
275	$17.39 \pm 0.49$	$550.19 \pm 22.27$	$7.52 \pm 2.14$	$22.88 \pm 1.02$	$580.59 \pm 22.39$
300	$17.91 \pm 0.47$	$404.54 \pm 15.93$	$4.93 \pm 2.32$	$18.82 \pm 0.91$	$428.29 \pm 16.12$
325	$18.55 \pm 0.46$	$315.77 \pm 14.01$	$3.90 \pm 2.27$	$16.06 \pm 0.87$	$335.72 \pm 14.22$
350	$19.35 \pm 0.46$	$259.00 \pm 13.71$	$1.45 \pm 0.81$	$14.49 \pm 0.84$	$274.94 \pm 13.76$
375	$18.09 \pm 0.43$	$186.53 \pm 9.92$	$1.37 \pm 0.82$	$12.92 \pm 0.82$	$200.82 \pm 9.98$
400	$16.17 \pm 0.38$	$146.83 \pm 8.63$	$0.38 \pm 0.27$	$9.47 \pm 0.64$	$156.68 \pm 8.66$
450	$10.48 \pm 0.28$	$107.83 \pm 8.28$	$0.21 \pm 0.21$	$5.90 \pm 0.50$	$113.94 \pm 8.29$
500	$6.19 \pm 0.17$	$55.62 \pm 2.95$	$0.05 \pm 0.05$	$5.25 \pm 0.53$	$60.92 \pm 3.00$
550	$3.53 \pm 0.12$	$37.72 \pm 3.14$	$0.00 \pm 0.00$	$3.41 \pm 0.42$	$41.13 \pm 3.17$
600	$1.92 \pm 0.06$	$25.75 \pm 2.72$	$0.00 \pm 0.00$	$2.19 \pm 0.34$	$27.94 \pm 2.74$

Table 19: Expected yields of signal and background with  $1 \text{ fb}^{-1}$  based on simulation in the 1 b-tag category. The numbers show  $2e2j$  expectations. Tighter  $m_{ZZ}$  mass requirements are applied as (-6%, +%) of the mass hypothesis.

Mass [GeV]	Signal	Z+Jets	$tt$	ZZ/WZ/WW	TotalBkg
200	$6.02 \pm 0.34$	$358.86 \pm 16.69$	$44.54 \pm 6.39$	$17.42 \pm 0.86$	$420.81 \pm 17.90$
210	$5.73 \pm 0.33$	$365.28 \pm 17.51$	$46.09 \pm 6.55$	$17.98 \pm 0.87$	$429.35 \pm 18.71$
220	$7.51 \pm 0.36$	$355.57 \pm 17.31$	$53.09 \pm 7.19$	$18.72 \pm 0.93$	$427.38 \pm 18.77$
230	$7.86 \pm 0.37$	$327.05 \pm 15.22$	$57.35 \pm 7.82$	$18.68 \pm 0.92$	$403.08 \pm 17.14$
250	$7.35 \pm 0.31$	$259.00 \pm 11.47$	$55.35 \pm 7.21$	$17.32 \pm 0.91$	$331.67 \pm 13.58$
275	$8.51 \pm 0.35$	$190.16 \pm 10.39$	$38.88 \pm 5.55$	$12.45 \pm 0.73$	$241.50 \pm 11.81$
275	$8.51 \pm 0.35$	$190.16 \pm 10.39$	$38.88 \pm 5.55$	$12.45 \pm 0.73$	$241.50 \pm 11.81$
300	$7.94 \pm 0.29$	$137.25 \pm 9.59$	$20.16 \pm 3.85$	$10.05 \pm 0.69$	$167.46 \pm 10.35$
325	$8.15 \pm 0.29$	$93.32 \pm 6.46$	$12.17 \pm 2.91$	$8.62 \pm 0.66$	$114.11 \pm 7.12$
350	$8.59 \pm 0.31$	$70.66 \pm 5.18$	$6.64 \pm 2.22$	$7.25 \pm 0.61$	$84.55 \pm 5.67$
375	$8.49 \pm 0.30$	$57.57 \pm 4.82$	$3.89 \pm 2.08$	$5.60 \pm 0.53$	$67.05 \pm 5.28$
400	$7.00 \pm 0.24$	$47.79 \pm 6.17$	$1.95 \pm 1.35$	$3.44 \pm 0.33$	$53.17 \pm 6.32$
450	$4.49 \pm 0.18$	$40.33 \pm 6.46$	$2.00 \pm 1.36$	$2.65 \pm 0.33$	$44.98 \pm 6.61$
500	$2.42 \pm 0.10$	$20.74 \pm 2.52$	$0.40 \pm 0.29$	$1.92 \pm 0.28$	$23.06 \pm 2.55$
550	$1.45 \pm 0.07$	$10.79 \pm 1.30$	$0.00 \pm 0.00$	$1.31 \pm 0.23$	$12.10 \pm 1.32$
600	$0.77 \pm 0.04$	$5.34 \pm 1.01$	$0.00 \pm 0.00$	$0.71 \pm 0.15$	$6.05 \pm 1.02$

Table 20: Expected yields of signal and background with  $1 \text{ fb}^{-1}$  based on simulation in the 1 b-tag category. The numbers show  $2\mu 2j$  expectations. Tighter  $m_{ZZ}$  mass requirements are applied as (-6%, +%) of the mass hypothesis.

Mass [GeV]	Signal	Z+Jets	$tt$	ZZ/WZ/WW	TotalBkg
200	$7.92 \pm 0.41$	$511.29 \pm 23.57$	$43.36 \pm 6.22$	$23.78 \pm 1.03$	$578.43 \pm 24.40$
210	$8.35 \pm 0.43$	$494.62 \pm 22.42$	$61.12 \pm 7.52$	$24.45 \pm 1.04$	$580.18 \pm 23.68$
220	$9.05 \pm 0.41$	$501.16 \pm 22.73$	$64.21 \pm 7.53$	$24.46 \pm 1.07$	$589.84 \pm 23.97$
230	$9.90 \pm 0.43$	$446.61 \pm 21.20$	$66.68 \pm 7.07$	$24.03 \pm 1.07$	$537.32 \pm 22.38$
250	$9.95 \pm 0.41$	$340.99 \pm 17.00$	$52.31 \pm 6.19$	$19.69 \pm 0.93$	$412.98 \pm 18.11$
275	$10.58 \pm 0.39$	$246.02 \pm 13.16$	$37.12 \pm 4.71$	$15.17 \pm 0.79$	$298.30 \pm 14.00$
275	$10.58 \pm 0.39$	$246.02 \pm 13.16$	$37.12 \pm 4.71$	$15.17 \pm 0.79$	$298.30 \pm 14.00$
300	$11.07 \pm 0.38$	$174.70 \pm 9.49$	$19.64 \pm 3.25$	$12.41 \pm 0.74$	$206.75 \pm 10.06$
325	$10.86 \pm 0.35$	$133.40 \pm 8.22$	$13.78 \pm 2.76$	$9.11 \pm 0.61$	$156.30 \pm 8.69$
350	$11.48 \pm 0.36$	$106.06 \pm 7.56$	$11.82 \pm 3.28$	$7.77 \pm 0.56$	$125.66 \pm 8.26$
375	$10.40 \pm 0.33$	$76.82 \pm 5.75$	$7.10 \pm 2.87$	$6.49 \pm 0.55$	$90.41 \pm 6.45$
400	$8.45 \pm 0.27$	$64.14 \pm 6.24$	$1.02 \pm 0.61$	$5.66 \pm 0.53$	$70.82 \pm 6.29$
450	$5.70 \pm 0.20$	$41.57 \pm 5.09$	$0.06 \pm 0.05$	$4.17 \pm 0.41$	$45.80 \pm 5.10$
500	$2.94 \pm 0.11$	$27.52 \pm 4.82$	$0.01 \pm 0.01$	$2.02 \pm 0.28$	$29.55 \pm 4.83$
550	$1.82 \pm 0.08$	$13.90 \pm 2.09$	$0.00 \pm 0.00$	$1.16 \pm 0.18$	$15.07 \pm 2.09$
600	$1.02 \pm 0.05$	$7.71 \pm 1.07$	$0.00 \pm 0.00$	$0.91 \pm 0.17$	$8.62 \pm 1.08$



Table 21: Expected yields of signal and background with  $1 \text{ fb}^{-1}$  based on simulation in the 2 b-tag category. The numbers show  $2e2j$  expectations. Tighter  $m_{ZZ}$  mass requirements are applied as (-6%, +%) of the mass hypothesis.

Mass [GeV]	Signal	Z+Jets	$tt$	ZZ/WZ/WW	TotalBkg
200	$1.40 \pm 0.15$	$12.46 \pm 2.36$	$1.52 \pm 0.84$	$2.79 \pm 0.32$	$16.77 \pm 2.53$
210	$1.47 \pm 0.15$	$13.39 \pm 2.36$	$6.66 \pm 2.76$	$3.35 \pm 0.40$	$23.41 \pm 3.65$
220	$2.48 \pm 0.22$	$12.25 \pm 3.15$	$8.35 \pm 2.98$	$3.34 \pm 0.37$	$23.95 \pm 4.36$
230	$2.42 \pm 0.20$	$16.83 \pm 3.46$	$10.08 \pm 3.45$	$3.39 \pm 0.37$	$30.30 \pm 4.90$
250	$2.61 \pm 0.18$	$22.49 \pm 6.99$	$6.19 \pm 2.12$	$2.41 \pm 0.26$	$31.10 \pm 7.31$
275	$3.20 \pm 0.21$	$11.90 \pm 1.77$	$5.25 \pm 1.41$	$1.82 \pm 0.22$	$18.97 \pm 2.27$
275	$3.20 \pm 0.21$	$11.90 \pm 1.77$	$5.25 \pm 1.41$	$1.82 \pm 0.22$	$18.97 \pm 2.27$
300	$3.21 \pm 0.18$	$7.14 \pm 1.19$	$3.17 \pm 1.20$	$1.84 \pm 0.23$	$12.15 \pm 1.70$
325	$4.00 \pm 0.21$	$4.48 \pm 0.87$	$1.01 \pm 0.42$	$1.79 \pm 0.27$	$7.28 \pm 1.00$
350	$4.31 \pm 0.22$	$4.79 \pm 1.01$	$0.52 \pm 0.33$	$1.90 \pm 0.30$	$7.20 \pm 1.10$
375	$4.22 \pm 0.21$	$4.33 \pm 1.07$	$0.00 \pm 0.00$	$1.65 \pm 0.29$	$5.99 \pm 1.10$
400	$3.58 \pm 0.19$	$3.38 \pm 0.82$	$0.00 \pm 0.00$	$1.18 \pm 0.22$	$4.55 \pm 0.85$
450	$2.52 \pm 0.14$	$2.69 \pm 0.61$	$0.00 \pm 0.00$	$0.49 \pm 0.10$	$3.17 \pm 0.62$
500	$1.24 \pm 0.07$	$1.21 \pm 0.31$	$0.00 \pm 0.00$	$0.61 \pm 0.19$	$1.82 \pm 0.36$
550	$0.86 \pm 0.05$	$0.45 \pm 0.13$	$0.00 \pm 0.00$	$0.69 \pm 0.20$	$1.14 \pm 0.23$
600	$0.50 \pm 0.03$	$0.15 \pm 0.06$	$0.00 \pm 0.00$	$0.29 \pm 0.09$	$0.45 \pm 0.11$

Table 22: Expected yields of signal and background with  $1 \text{ fb}^{-1}$  based on simulation in the 2 b-tag category. The numbers show  $2\mu 2j$  expectations. Tighter  $m_{ZZ}$  mass requirements are applied as (-6%, +%) of the mass hypothesis.

Mass [GeV]	Signal	Z+Jets	$tt$	ZZ/WZ/WW	TotalBkg
200	$2.17 \pm 0.21$	$24.34 \pm 4.45$	$2.78 \pm 0.80$	$4.68 \pm 0.45$	$31.79 \pm 4.55$
210	$1.97 \pm 0.18$	$20.52 \pm 2.46$	$6.14 \pm 2.14$	$4.08 \pm 0.43$	$30.75 \pm 3.29$
220	$2.61 \pm 0.21$	$22.55 \pm 2.66$	$8.81 \pm 2.39$	$4.06 \pm 0.43$	$35.42 \pm 3.60$
230	$3.49 \pm 0.24$	$21.14 \pm 2.23$	$14.12 \pm 3.14$	$3.44 \pm 0.37$	$38.70 \pm 3.87$
250	$3.71 \pm 0.24$	$17.53 \pm 3.21$	$12.54 \pm 3.11$	$4.00 \pm 0.42$	$34.08 \pm 4.49$
275	$3.81 \pm 0.22$	$13.58 \pm 3.19$	$8.56 \pm 2.51$	$2.94 \pm 0.33$	$25.08 \pm 4.07$
275	$3.81 \pm 0.22$	$13.58 \pm 3.19$	$8.56 \pm 2.51$	$2.94 \pm 0.33$	$25.08 \pm 4.07$
300	$4.25 \pm 0.22$	$8.93 \pm 1.32$	$6.28 \pm 1.84$	$2.39 \pm 0.30$	$17.59 \pm 2.28$
325	$4.69 \pm 0.24$	$8.45 \pm 1.27$	$3.18 \pm 1.41$	$2.48 \pm 0.36$	$14.11 \pm 1.93$
350	$5.21 \pm 0.25$	$5.99 \pm 1.02$	$0.31 \pm 0.23$	$2.23 \pm 0.33$	$8.53 \pm 1.09$
375	$4.90 \pm 0.24$	$5.15 \pm 0.90$	$0.21 \pm 0.21$	$1.71 \pm 0.26$	$7.07 \pm 0.96$
400	$4.57 \pm 0.21$	$4.09 \pm 0.80$	$0.01 \pm 0.01$	$1.32 \pm 0.20$	$5.41 \pm 0.82$
450	$3.01 \pm 0.15$	$2.86 \pm 0.61$	$0.00 \pm 0.00$	$1.16 \pm 0.19$	$4.02 \pm 0.64$
500	$1.72 \pm 0.08$	$1.49 \pm 0.41$	$0.00 \pm 0.00$	$0.57 \pm 0.10$	$2.06 \pm 0.42$
550	$1.01 \pm 0.06$	$1.12 \pm 0.50$	$0.00 \pm 0.00$	$0.50 \pm 0.13$	$1.62 \pm 0.51$
600	$0.65 \pm 0.04$	$0.79 \pm 0.25$	$0.00 \pm 0.00$	$0.35 \pm 0.10$	$1.14 \pm 0.27$

tribution of the full mass for those events is needed to actually validate the expectation so it would be required to be “unblinded” before the final calculations are done. It should be noted that reproducing the global shape of that distribution is not really something that can bias our optimization that is not based on that distribution. We just use that distribution for defining the signal-region events for which the angular analysis is performed.

On the other hand, only the “no tag” category contains enough statistics and can be considered insensitive to the presence of signal due to the overwhelming background. Therefore the validation of the sideband method will be performed in this region only. The others will remain blind until the end.

As described in Section 5.4, the final discriminant is not the full invariant mass, but some multivariate analysis of kinematics variables obtained from events in the specific signal regions. These signal regions are defined by a mass window for the hadronic Z and, for each assumption of the Higgs mass, a mass window for the full invariant mass.

In this respect, both, the distributions of the events AND the definition of the signal region for each Higgs mass under study will be considered sensitive to signal and therefore will be “blind” in the analysis. The events in the decided signal regions will not be looked at until the unblinding is approved (at preapproval). The exact definition of the mass windows will be defined a priori as part of the blind optimization process. In fact they will be defined before the events in the signal region are used for anything, specifically for validating the background estimation from the sidebands.

The validation of the final variables and of the optimization process will be done using the events in the sidebands that enter in the signal region, when MC is not available for getting the background estimation. Therefore we do not need to use the data in the signal region (i.e. the selected window for the hadronic Z and the window in the full invariant mass for each assumption of the Higgs mass) for anything and will be blind regarding this purpose.

Those events will be kept blind until the analysis is preapproved. At that moment nothing will be changed in the analysis strategy to avoid being biased by the oversampling in the signal region of the data.

## 5.6 Unblinded results

Figures 12 and 13 show the final selected events in the signal region compared to the MC expectation. The agreement is overall very good and no obvious hint of a possible signal appears in the data sample. In Figures ?? and 15 a comparison of angular distribution in data and MC for electron and muon events in the signal region is shown. The pre-selection described in section is applied together with a cut on the di-jet invariant mass to select events in the signal region.

## 6 Background Determination from Data

The  $m_{ZZ}$  distribution is used as discriminant variable in the subsequent analysis, which is performed separately for the three  $b$ -tag categories: 0, 1, and 2  $b$ -tag. These distributions, one per category, do not depend on the Higgs mass hypothesis. They are displayed in Figs. 12 and 13 for the data and the simulated backgrounds.

In order to avoid a possible bias of the experimenter in the process of developing the analysis strategy and optimizing the selection requirements a so called *blind analysis* is performed, which means data in the signal region are never examined at this stage of the analysis. Once the

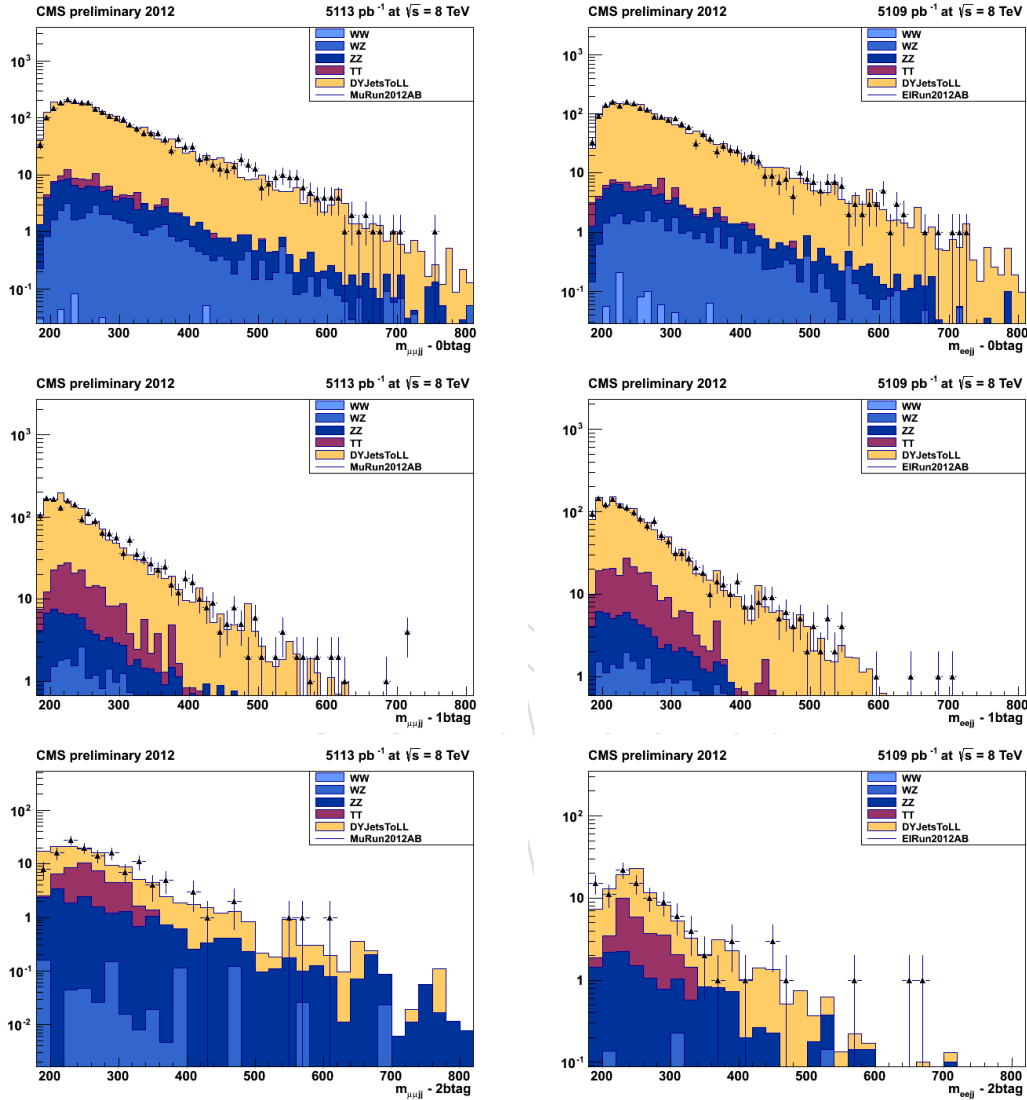


Figure 12: The  $m_{ZZ}$  distribution in the signal  $m_{JJ}$  range in the three  $b$ -tag categories from top to bottom: 0  $b$ -tag (top), 1  $b$ -tag (middle), and 2  $b$ -tag (bottom). Left-side plots refer to events in the muon channel and those on the right side to events in the electron channel. Points with error bars show data after final selection, histograms show MC prediction with the dominant contributions shown separately.

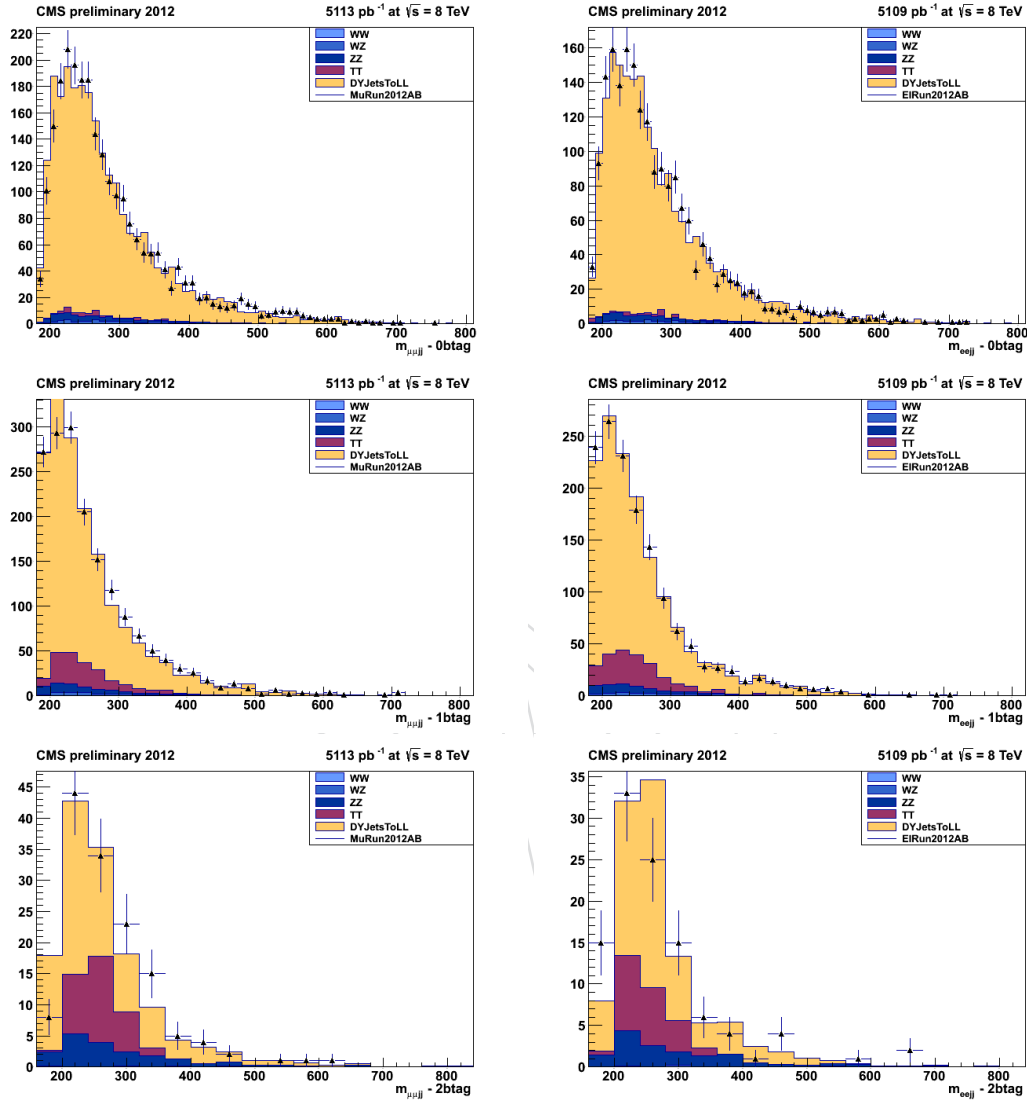


Figure 13: The  $m_{ZZ}$  distribution in the signal  $m_{JJ}$  range in the three  $b$ -tag categories from top to bottom: 0  $b$ -tag (top), 1  $b$ -tag (middle), and 2  $b$ -tag (bottom). Left-side plots refer to events in the muon channel and those on the right side to events in the electron channel. Points with error bars show data after final selection, histograms show MC prediction with the dominant contributions shown separately.

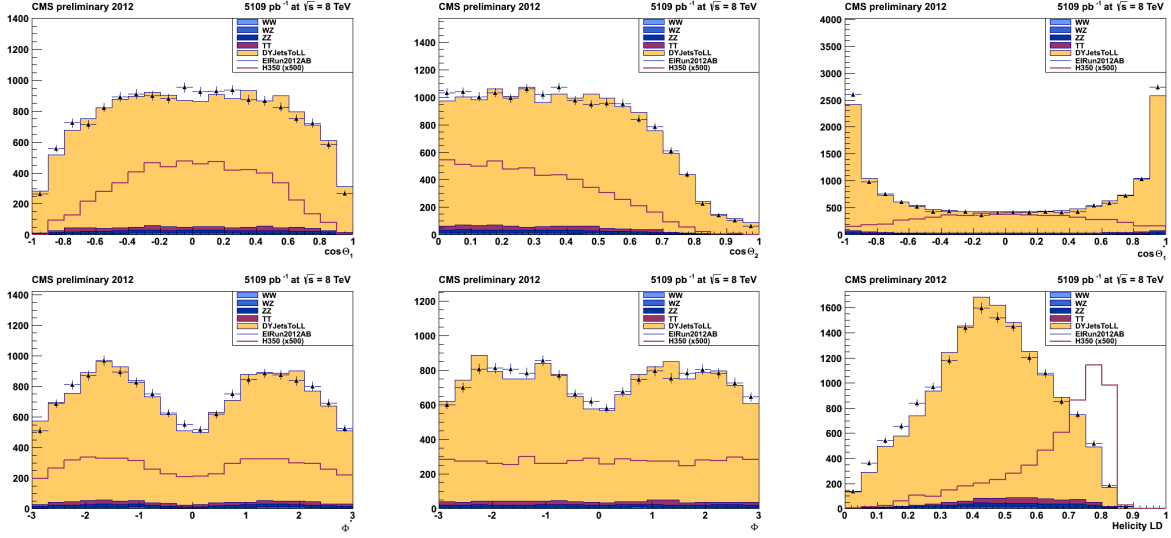


Figure 14: Five angular distributions of  $\cos \theta_1$ ,  $\cos \theta_2$ ,  $\cos \theta^*$ ,  $\Phi$ ,  $\Phi_1$  and the helicity likelihood discriminant for 2012 electron data (points) and Summer 12 Monte Carlo samples (histogram) in the signal region. Open histograms indicate the expected distribution for a Higgs boson with mass 350 GeV, multiplied by a factor of 500 for illustration.

event selection criteria are settled data in the signal-depleted region, in the sidebands of the  $m_{JJ}$  distribution, are investigated. Only when the data distributions in the sideband regions confirm the robustness of the analysis data in the signal region are inspected.

## 6.1 Background estimation from the $m_{JJ}$ sidebands

The dominant background to a potential Higgs boson signal stems from Z events produced in association with jets, so called Z+jets events, and from  $t\bar{t}$  events. Di-boson events are a less relevant background source, contributing less than 10% to the total. Jets from the dominant backgrounds originate in the hadronisation of both light and heavy quarks in Z+jet events, and mainly in the hadronisation of b quarks in the  $t\bar{t}$  processes.

Dependence of the analysis results on the details of the background modeling and simulation is reduced by extracting the background prediction from data events selected in control regions where the signal is expected to be negligible. The distribution of the dijet invariant mass,  $m_{JJ}$ , provides a handle to define such control region, as signal events cluster around the true Z boson mass, unlike background events, which show a broad spectrum. Events in a  $m_{JJ}$  region close to  $m_Z$ , but different from the region containing most of the expected signal,  $75 \text{ GeV} < m_{JJ} < 105 \text{ GeV}$ , present a 4-fermion mass distribution,  $m_{ZZ}$ , similar to that expected for the events selected in the analysis. This signal-depleted region is referred to as sideband. Data events selected in the sideband region provide a good prediction of the background in the signal region ( $N_{\text{bkg}}(m_{ZZ})$ ), both in shape and normalisation, apart from subtle kinematic differences in the two  $m_{JJ}$  regions. These differences, which are mostly independent of the theoretical calculation of the cross-sections of background processes, are accounted for using simulated events. The resulting background prediction can be expressed in terms of the distribution observed in the data in the sideband,  $N_{\text{sb}}(m_{ZZ})$ , as

$$N_{\text{bkg}}(m_{ZZ}) = N_{\text{sb}}(m_{ZZ}) \times \frac{N_{\text{bkg}}^{\text{MC}}(m_{ZZ})}{N_{\text{sb}}^{\text{MC}}(m_{ZZ})} = N_{\text{sb}}(m_{ZZ}) \times \alpha(m_{ZZ}). \quad (3)$$

The function  $\alpha(m_{ZZ})$ , determined from simulated events, accounts for the small kinematic dif-

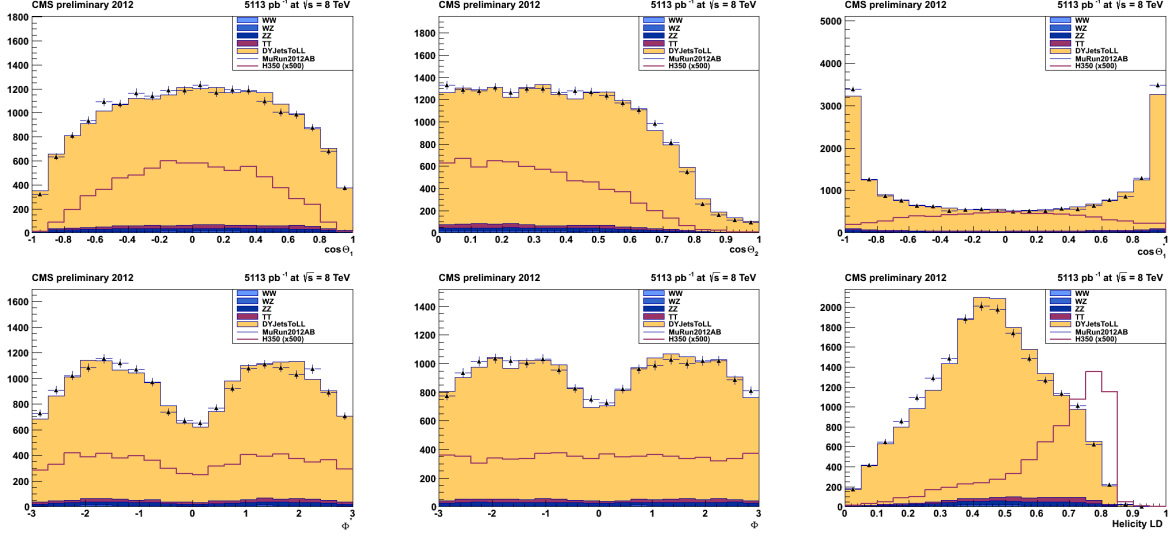


Figure 15: Five angular distributions of  $\cos \theta_1, \cos \theta_2, \cos \theta^*, \Phi, \Phi_1$  and the helicity likelihood discriminant for 2012 muon data (points) and Summer 12 Monte Carlo samples (histogram) in the signal region. Open histograms indicate the expected distribution for a Higgs boson with mass 350 GeV, multiplied by a factor of 500 for illustration.

ferences expected between the  $m_{ZZ}$  obtained in the signal and sideband regions. The background estimated this way includes contributions from the main sources: Z+jets,  $t\bar{t}$  and di-boson processes, which means that  $\alpha(m_{ZZ})$  contains simulation based corrections for all those background processes.

An interesting approach to obtain a proper prediction of the Z+jets background is based on extracting the  $t\bar{t}$  contribution from a different control data sample, which uses  $t\bar{t} \rightarrow e\mu + X$  events to model the background. More details on how to get this sample and its ability to reproduce the  $t\bar{t}$  background are given in the next section (6.2). Subtracting the  $t\bar{t}$  background prior the calculation of  $\alpha(m_{ZZ})$  avoids the unnecessary annoyance of relying on the  $t\bar{t}$  and single top simulation. The function  $N_{\text{bkg}}^{Z+\text{jets}}(m_{ZZ})$  becomes an estimate of the Z+jets and di-boson backgrounds alone:

$$N_{\text{bkg}}^{Z+\text{jets}}(m_{ZZ}) = \left( N_{\text{sb}}(m_{ZZ}) - N_{\text{sb}}^{t\bar{t} \rightarrow e\mu}(m_{ZZ}) \right) \times \alpha^{Z+\text{jets}}(m_{ZZ}), \quad (4)$$

where  $\alpha^{Z+\text{jets}}(m_{ZZ})$  is the ratio of simulated Z+jets plus di-boson events in the signal region to the sideband region.  $N_{\text{sb}}^{t\bar{t} \rightarrow e\mu}$  is the  $m_{ZZ}$  distribution of  $t\bar{t} \rightarrow e\mu + X$  events in the sideband obtained from data. The background prediction in the signal region is the sum of two contributions, one corresponding to Z+jets and di-boson processes, which contain a correction based on the simulation, and the  $t\bar{t}$  background, estimated from a pure data sample of  $t\bar{t} \rightarrow e\mu + X$  events,  $N_{\text{sb}}^{t\bar{t} \rightarrow e\mu}$ , selected in the signal region:

$$N_{\text{bkg}}(m_{ZZ}) = N_{\text{bkg}}^{Z+\text{jets}}(m_{ZZ}) + N_{\text{sb}}^{t\bar{t} \rightarrow e\mu}(m_{ZZ}) \quad (5)$$

This study is currently well advanced and is expected to be adopted as the main analysis for the final publication.

### 6.1.1 Uncertainty on the background prediction

As a first approximation, the  $N(m_{ZZ})$  functions are calculated using binned distributions of  $m_{ZZ}$ , obtained both from data and simulations in the different regions. Lack of events in the

high mass tail of the distributions make difficult to get a proper estimate of the  $\alpha(m_{ZZ})$  function, particularly at high mass. Exploiting the fact that the  $m_{ZZ}$  distribution of background events stems from random combination of jets, unrelated by the decay of a potential heavier particle, the  $N(m_{ZZ})$  functions in Eqs. 3 and 4 are expected to decay exponentially at high mass. Hence, they are parametrized using a modified Moyal function, which tends to an exponential function with a negative constant at high mass and reproduces the background shape at low mass:

$$f(x) = \frac{N}{\sqrt{2\pi\sigma_1\sigma_2}} \exp\left(-\frac{1}{2}e^{-\frac{x-\mu}{\sigma_1}} - \frac{x-\mu}{2\sigma_2}\right) \quad (6)$$

where for readability  $m_{ZZ}$  is denoted by  $x$  and the exponential function by  $\exp$ .  $N$  is a normalisation constant,  $\mu$  is the peak position and  $\sigma_1$  and  $\sigma_2$  are the widths of the left and right branches of the function. These parameters are determined, for each b-tag category and  $N(m_{ZZ})$  function in Eqs. 3 and 4 separately, performing a fit to the corresponding binned distributions in the sideband and signal regions. For a given set of fitted parameters its covariance matrix,  $C_{ij}$ , is used to calculate the variance of  $f(x)$ :

$$\text{var}(f(x)) = \vec{J}(x)C\vec{J}^T(x) = \sum_{i=1}^4 \sum_{j=1}^4 J_i(x)C_{ij}J_j(x) \quad (7)$$

where  $\vec{J}(x)$  is the Jacobian of  $f(x)$ :

$$\vec{J}(x) = \left( \frac{\partial f(x)}{\partial N}, \frac{\partial f(x)}{\partial \mu}, \frac{\partial f(x)}{\partial \sigma_1}, \frac{\partial f(x)}{\partial \sigma_2} \right) \quad (8)$$

The partial derivatives are calculated analytically:

$$\begin{aligned} \frac{\partial f(x)}{\partial N} &= \frac{f(x)}{N} & ; & \quad \frac{\partial f(x)}{\partial \sigma_1} = -\frac{f(x)}{2\sigma_1} \left( 1 + \frac{x-\mu}{\sigma_1} \cdot e^{-\frac{x-\mu}{\sigma_1}} \right) \\ \frac{\partial f(x)}{\partial \mu} &= \frac{f(x)}{2} \left( \frac{1}{\sigma_2} - \frac{e^{-\frac{x-\mu}{\sigma_1}}}{\sigma_1} \right) & ; & \quad \frac{\partial f(x)}{\partial \sigma_2} = -\frac{f(x)}{2\sigma_2} \left( 1 - \frac{x-\mu}{\sigma_2} \right) \end{aligned} \quad (9)$$

The function  $\alpha(m_{ZZ})$  calculated as the ratio of analytical functions is more precise than the ratio of binned distributions, and more easily extrapolated to high mass based on the simple basic assumption of the exponential tails. The uncertainty on  $\alpha(m_{ZZ})$  and on the background prediction in the signal region is obtained by simple error propagation, assuming the uncertainties on the fits to the background distributions in the sideband (sb) and signal region (sr) are uncorrelated:

$$\alpha(x) = \frac{N_{bkg}^{MC}(x)}{N_{sb}^{MC}(x)} \Rightarrow \frac{\sigma_\alpha^2(x)}{\alpha^2(x)} = \frac{\sigma_{N_{bkg}^{MC}}^2(x)}{N_{bkg}^{MC^2}(x)} + \frac{\sigma_{N_{sb}^{MC}}^2(x)}{N_{sb}^{MC^2}(x)} \quad (10)$$

$$N_{bkg}(x) = \alpha(x) \cdot N_{sb}(x) \Rightarrow \frac{\sigma_{N_{bkg}}^2(x)}{N_{bkg}^2(x)} = \frac{\sigma_\alpha^2(x)}{\alpha^2(x)} + \frac{\sigma_{N_{sb}}^2(x)}{N_{sb}^2(x)}$$

The uncertainties,  $\sigma(x)$ , of the  $N(x)$  functions are extracted from the variance of the fitted functions (Eqs. 7-9):

$$\sigma(x) = \sqrt{\text{var}(f(x))} \quad (11)$$



430 The results of the fits to the  $m_{ZZ}$  binned distributions are depicted in Fig. 16 for the three b-  
 431 tag categories, along with the uncertainties from the fits. The  $\alpha(x)$  functions are displayed in  
 432 Fig. 17 and the background prediction in the signal region in Fig. 18, also for the three b-tag  
 433 categories.

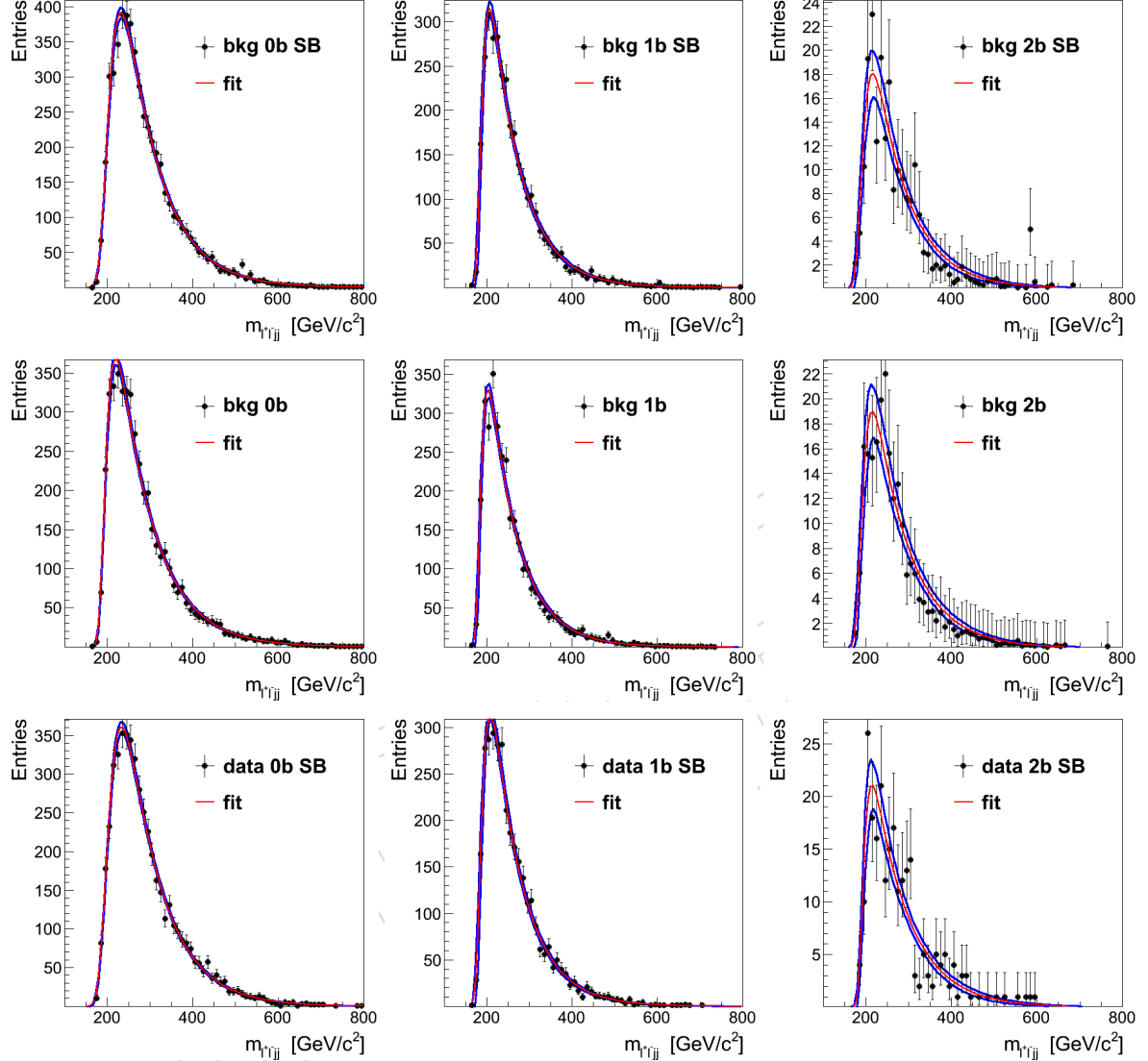


Figure 16: Results of the fits to the  $m_{ZZ}$  binned distributions for the three b-tag categories: (upper) for the background in the sideband region and (middle) in the signal regions, and (bottom) for the data in the sideband region. The outer solid lines (blue) correspond to the  $\pm 1\sigma$  uncertainty on the fit, calculated using the covariance matrix.

434 The uncertainty on the background predicted with this method is treated as a systematic un-  
 435 certainty in the statistical procedure designed to quantify a possible signal of a Higgs boson.  
 436 These systematic uncertainties are statistical in nature and scale with the size of the data sam-  
 437 ples in the sideband and signal regions. The errors of the fit parameters yield normalization  
 438 uncertainties of the order of 5% in the 0 and 1 b-tag categories and 17% in the 2 b-tag category,  
 439 significantly lower than those expected from plain simulation. In particular, only the uncer-  
 440 tainty in the b-tag efficiency obtained from simulation amounts to 20% in the 2 b-tag category.

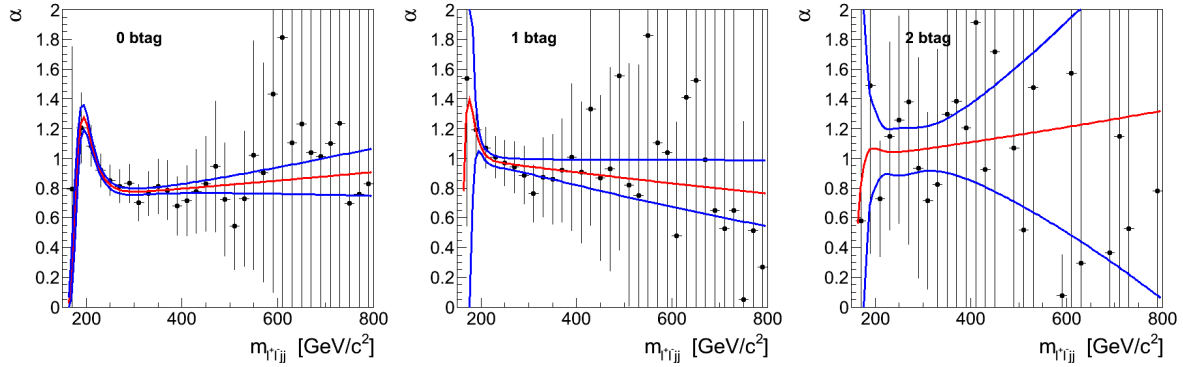


Figure 17:  $\alpha(x)$  functions (inner line in red) together with their uncertainties (outer lines in blue), for the three b-tag categories. For comparison, the ratio of binned distributions is shown as dots with error bars. No fit is performed to these points.

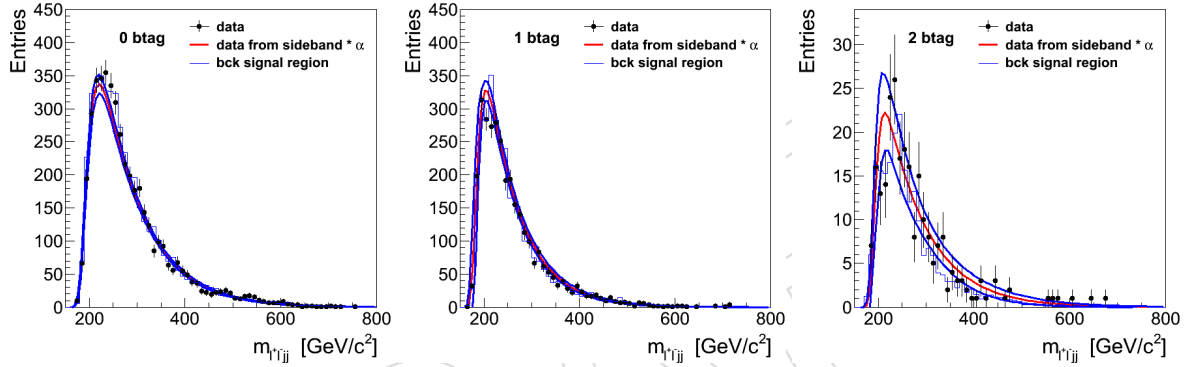


Figure 18: Mass distribution,  $m_{\ell\ell jj}$ , of data events (dots with error bars) in the signal region, for the three b-tag categories. The background prediction from the  $\alpha$  function is depicted as the inner red line and its uncertainty represented by the outer thick blue lines. For comparison, the background prediction in the signal region from the simulation is shown as a thin blue histogram. No fit is performed to this distribution.

Table 23 and 24 show the expected yield when applying the above procedure and the observed yields in each of the 6 channels. The errors on these expectations are on the order of 5% for the 0 and 1 b-tag categories and about 20% for the 2 b-tag category.

## 6.2 $t\bar{t}$ Background Determination from Data

The  $t\bar{t}$  background is estimated from the data using  $e\mu$  events passing the same cuts as the signal. This method accounts for other small backgrounds (as  $WW$  + jets,  $Z \rightarrow \tau\tau$  + jets, single top, fakes) where the lepton flavour symmetry can be invoked as well.

In this study we use the Powheg + Pythia  $t\bar{t} \rightarrow 2l2\nu X$  Monte Carlo sample. Other top MC samples as the Madgraph  $t\bar{t}$  inclusive sample or Powheg + Herwig  $t\bar{t} \rightarrow 2l2\nu X$  produce consistent results.

Top-pair Monte Carlo studies show that the  $e\mu$  vs.  $ee + \mu\mu$  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

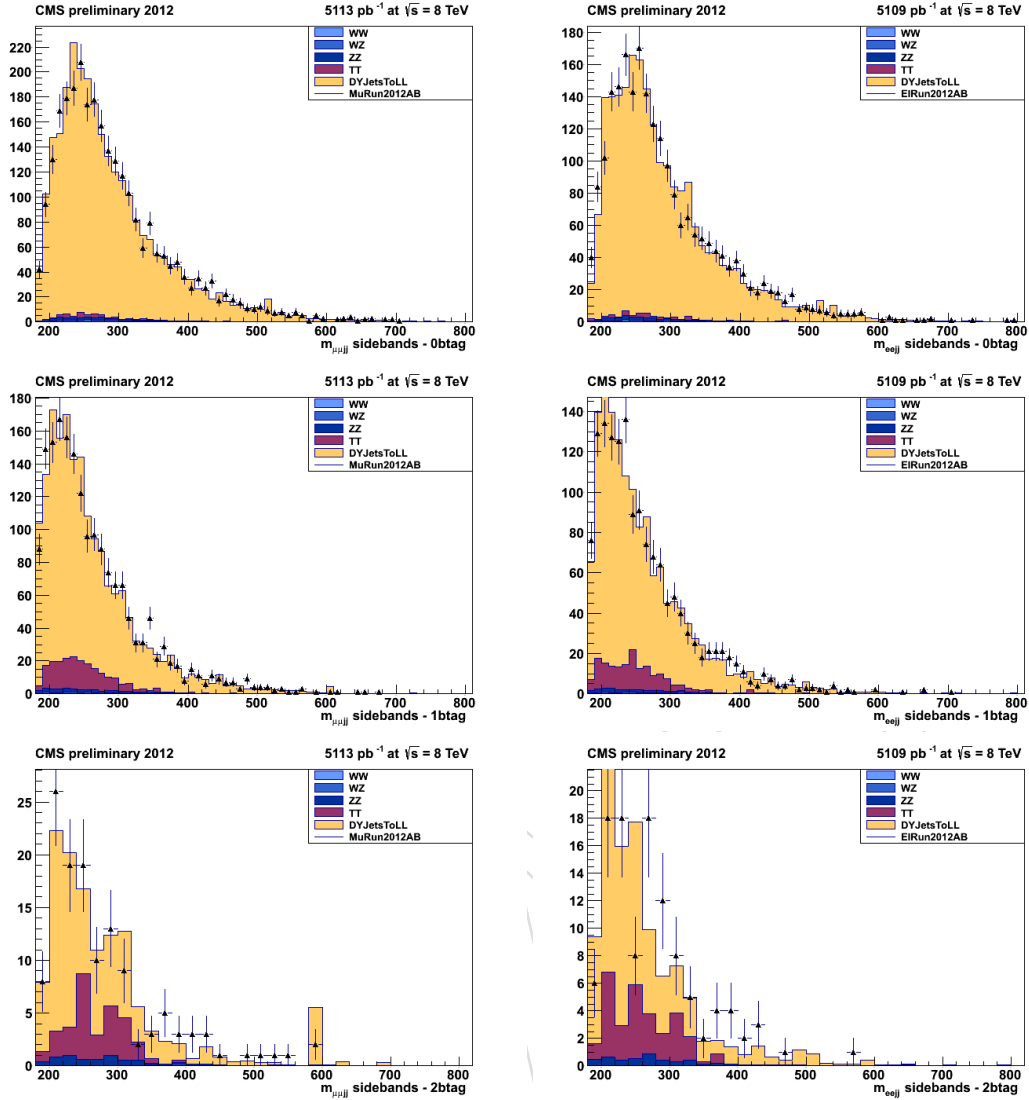


Figure 19: The  $m_{ZZ}$  distribution in the sideband  $m_{JJ}$  range in the three  $b$ -tag categories from top to bottom: 0  $b$ -tag (top), 1  $b$ -tag (middle), and 2  $b$ -tag (bottom). left: Muon channel; right: Electron channel. Points with error bars show data after final selection, histograms show MC prediction with the dominant contributions shown separately.

case of Powheg + Pythia top MC, the  $e\mu/(ee + \mu\mu)$  relative event normalizations are  $1.0036 \pm 0.0018$  after selection and kinematical cuts, and  $1.0045 \pm 0.0076$  after  $b$ -tagging.

Figure 20 shows a comparison of the  $ee + \mu\mu$  and  $e\mu$  top MC distributions of several relevant variables, for events with at least two leptons and two jets passing selection cuts. Only the hardest- $\sum P_T$  dilepton combination, and the dijet combination with largest TCHE discriminator values are considered. The selection step is specified in each plot. The category “ $\geq 1$   $b$ -tag” includes events with at least one jet tagged using TCHEM prescription. The category “2  $b$ -tag” includes events with one jet tagged using TCHEM prescription and one jet tagged using TCHEL. The normalization is arbitrary.

The 2012  $e\mu$  data yields are compared to the sum of top MC prediction and other small backgrounds in Table 25, while distributions of relevant variables are superimposed in Figure 21.

Table 23: Observed and expected event yields with  $5.1 \text{ fb}^{-1}$  of data in the electron channel. The yields are quoted in the range  $0 < () < 1000$ , depending on the Higgs boson mass hypothesis. The expected background is quoted from the sideband procedure (data) and from simulation (MC). The errors on the expected background from simulation include only statistical uncertainties.

		0 b-tag	1 b-tag	2 b-tag
$\in [0, 1000]$				
observed yield		$1975 \pm 44$	$1436 \pm 38$	$106 \pm 10$
expected background (data)		2382.76	1798.91	141.24
expected background (MC)		$1866.12 \pm 58.52$	$1288.06 \pm 48.70$	$96.25 \pm 11.87$
signal expectation (MC)				
Higgs	$=200$	$6.64 \pm 0.32$	$7.47 \pm 0.37$	$1.55 \pm 0.16$
	$=250$	$15.65 \pm 0.49$	$10.24 \pm 0.38$	$2.83 \pm 0.19$
	$=300$	$19.25 \pm 0.47$	$11.45 \pm 0.36$	$3.88 \pm 0.20$
	$=350$	$22.53 \pm 0.49$	$13.30 \pm 0.38$	$5.44 \pm 0.25$
	$=400$	$19.53 \pm 0.41$	$11.05 \pm 0.31$	$5.10 \pm 0.22$
	$=450$	$13.38 \pm 0.30$	$8.10 \pm 0.24$	$3.68 \pm 0.16$
	$=500$	$8.48 \pm 0.18$	$4.52 \pm 0.13$	$2.11 \pm 0.09$
	$=550$	$5.17 \pm 0.13$	$2.98 \pm 0.10$	$1.45 \pm 0.07$
	$=600$	$3.27 \pm 0.08$	$1.77 \pm 0.06$	$0.96 \pm 0.04$

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 TCHE discriminant 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  $WW, Z \rightarrow \tau\tau$ , and single top contributions from Monte Carlo. The fake component is estimated from  $e\mu$  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  $e - \mu$  symmetry holds in reasonable approximation for the non-isolated lepton data.

The sample composition before b-tagging is 85%  $t\bar{t}$ , 9% fakes, and 7% other small backgrounds. After requiring 1 TCHEM tag (1 TCHEM and 1 TCHEL tags) the relative fractions change to 91%(94%)  $t\bar{t}$ , 5%(4%) fakes, and 4%(2%) other small backgrounds.

Now, we test the  $e\mu$  vs.  $ee + \mu\mu$  symmetry using a top-enriched subsample of the data. Figure 22 shows the MET significance distribution after requiring 1 TCHEM tag (left) and 1 TCHEM and 1 TCHEL tags (right). For values sufficiently large of MET significance the number of events of the  $e\mu$  and  $ee + \mu\mu$  samples are equal within statistical errors. In order to test the agreement in shape, Figure 23 contains the di-jet invariant mass and "Higgs" invariant mass distributions after requiring 1 TCHEM tag, MET significance  $> 6$ , and  $|M_{ll} - M_Z| > 30 \text{ GeV}/c^2$ . One can observe agreement on both the normalization and shape of the  $e\mu$  and  $ee + \mu\mu$  distributions.

Table 24: Observed and expected event yields with  $5.1 \text{ fb}^{-1}$  of data in the muon channel. The yields are quoted in the range  $0 < () < 1000$ , depending on the Higgs boson mass hypothesis. The expected background is quoted from the sideband procedure (data) and from simulation (MC). The errors on the expected background from simulation include only statistical uncertainties.

		0 b-tag	1 b-tag	2 b-tag
$\in [0, 1000]$				
observed yield		$2524 \pm 50$	$1731 \pm 42$	$138 \pm 12$
expected background (data)		1923.35	1479.1	107.35
expected background (MC)		$2395.48 \pm 70.32$	$1744.59 \pm 58.26$	$134.34 \pm 16.50$
signal expectation (MC)				
Higgs	$=200$	$8.12 \pm 0.39$	$9.69 \pm 0.45$	$2.37 \pm 0.22$
	$=250$	$19.23 \pm 0.53$	$13.54 \pm 0.47$	$4.14 \pm 0.26$
	$=300$	$23.94 \pm 0.53$	$15.45 \pm 0.44$	$5.22 \pm 0.24$
	$=350$	$27.85 \pm 0.54$	$17.13 \pm 0.44$	$6.73 \pm 0.28$
	$=400$	$25.36 \pm 0.48$	$13.82 \pm 0.35$	$6.57 \pm 0.25$
	$=450$	$17.69 \pm 0.36$	$10.01 \pm 0.27$	$4.46 \pm 0.18$
	$=500$	$10.90 \pm 0.21$	$5.49 \pm 0.14$	$2.80 \pm 0.11$
	$=550$	$6.62 \pm 0.15$	$3.69 \pm 0.11$	$1.82 \pm 0.08$
	$=600$	$4.22 \pm 0.09$	$2.12 \pm 0.06$	$1.21 \pm 0.05$

Cuts	Top MC	Total MC	$e\mu$ data
$M_{ll} > 50 \text{ GeV}/c^2$	6602.1	7566.5	7504
$70 \text{ GeV}/c^2 < M_{ll} < 110 \text{ GeV}/c^2$	2244.2	2606.9	2582
$70 \text{ GeV}/c^2 < M_{jj} < 110 \text{ GeV}/c^2$	463.0	551.8	563
$\geq 1$ TCHEL b-tagged jet	423.5	474.2	481
$\geq 1$ TCHEM b-tagged jet	371.0	411.4	398
$\geq 1$ TCHEL & $\geq 1$ TCHEM	223.8	238.9	220
1BL+1BM MET Sig $< 10$	76.9	80.9	80

Table 25: Comparison of 2012  $e\mu$  data to Powheg + Pythia top MC event yields, corresponding to an integrated luminosity of  $5.2 \text{ fb}^{-1}$ . “Total MC” contains the top,  $WW$ ,  $Z \rightarrow \tau\tau$ , single top, and fakes contributions. Every cut in a line assumes all cuts in lines above.

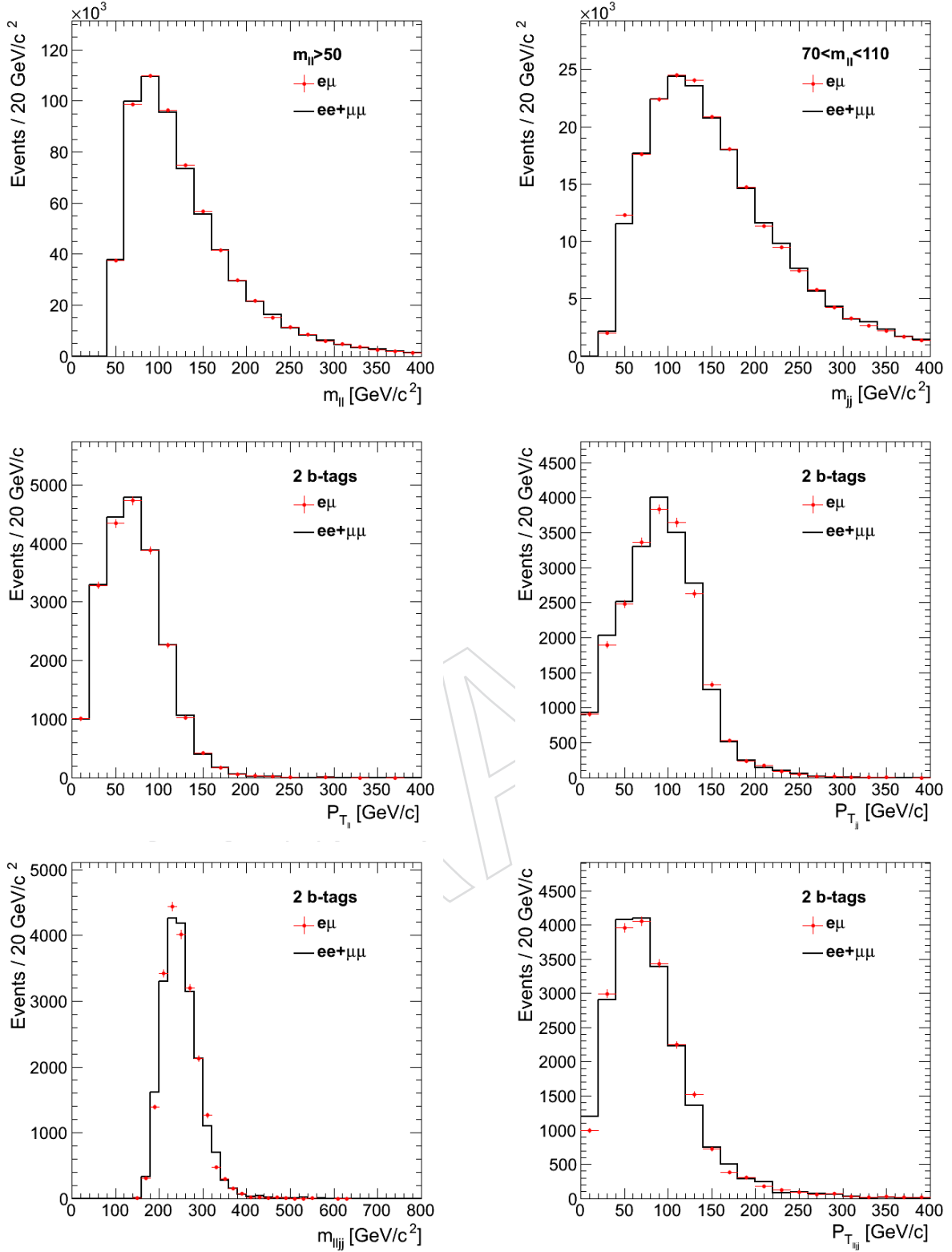


Figure 20: Powheg + Pythia top MC  $e\mu$  to  $(ee + \mu\mu)$  comparison for several variables after different step of the selection, as specified in the legends. Top: di-lepton invariant mass (Left) and di-jet invariant mass. Center: di-lepton (left) and dijet transverse momentum (right). Bottom: di-lepton + di-jet "Higgs" invariant mass (left) and transverse momentum (right).

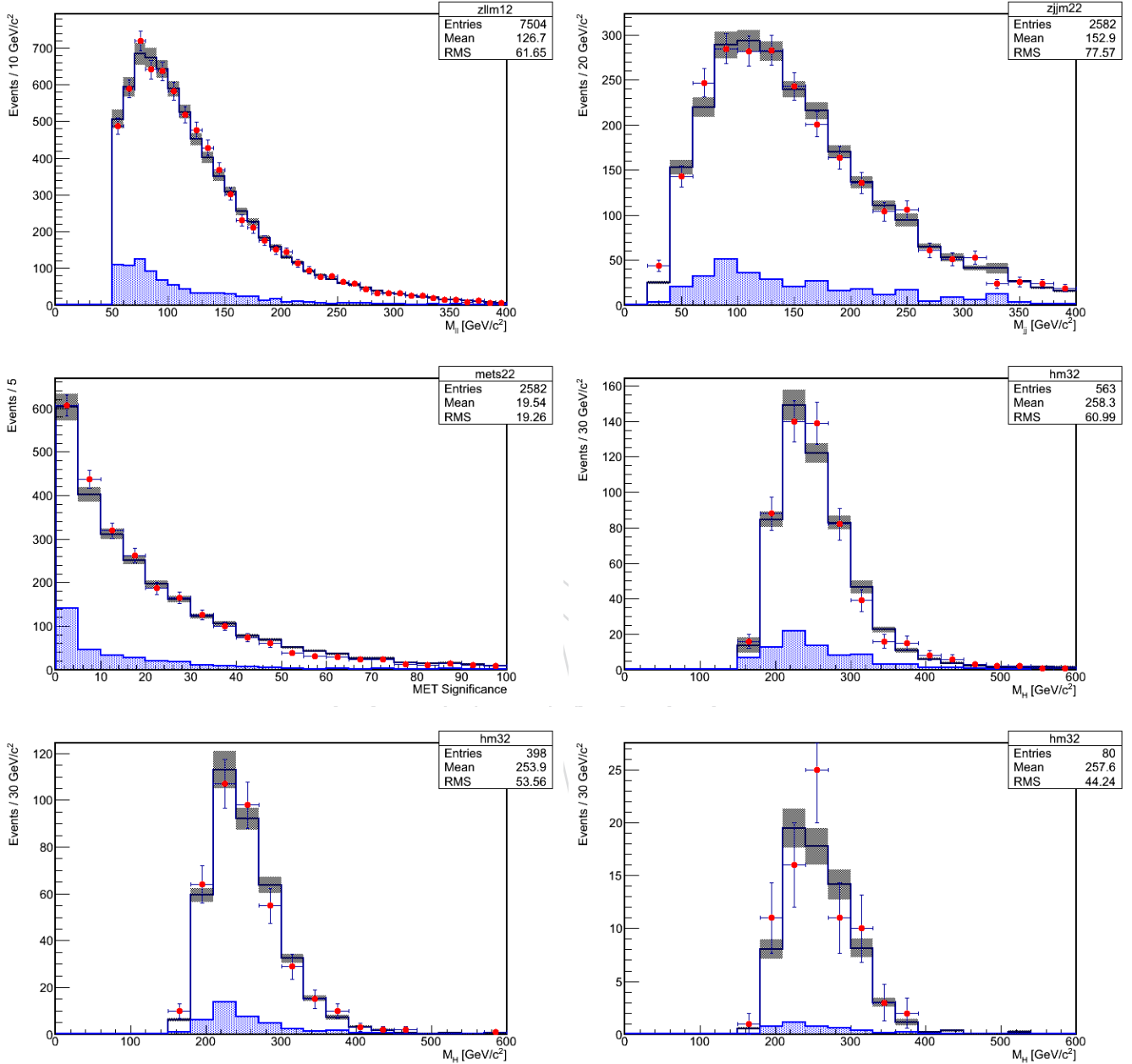


Figure 21: Comparison of 2012  $e\mu$  data to Powheg + Pythia top MC, corresponding to an integrated luminosity of  $5.2 \text{ fb}^{-1}$ . Red dots are  $e\mu$  data; white histogram top Monte Carlo; blue histogram other small backgrounds. Top: dilepton invariant mass (left) and dijet invariant mass (right). Center: MET significance (left) and "Higgs" invariant mass (right). Bottom: "Higgs" invariant mass for events with 1 TCHEM b-tag (left), and 1 TCHEM + 1 TCHEL b-tags and MET significance  $< 10$ .



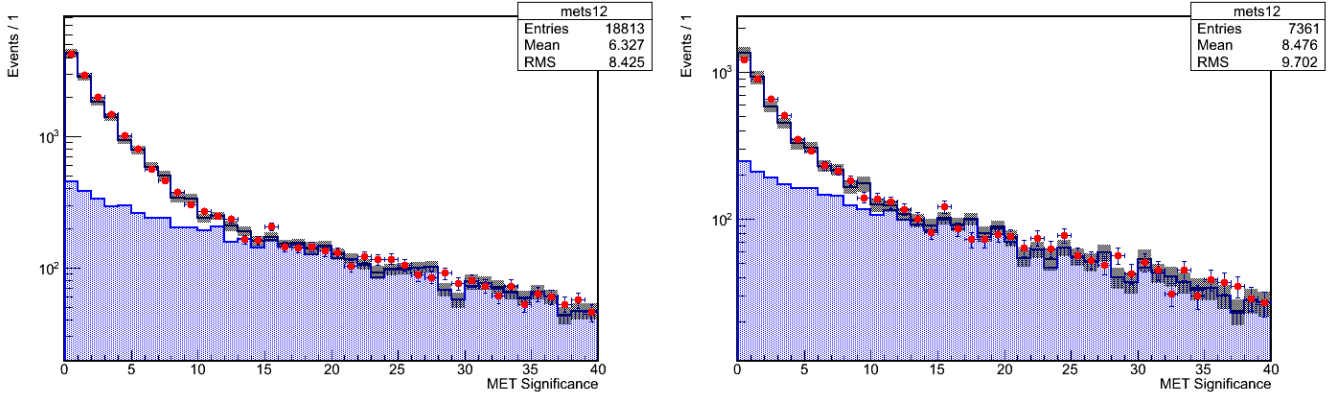


Figure 22: MET significance distribution for dilepton data compared to the sum of Drell-Yan Monte Carlo plus  $e\mu$  data after 1 TCHEM b-tag (left) and two 1 TCHEM + 1 TCHEL b-tags (right). Red dots are  $ee + \mu\mu$  data; white histogram Drell Yan Monte Carlo; blue histogram  $e\mu$  data (plus other small backgrounds).

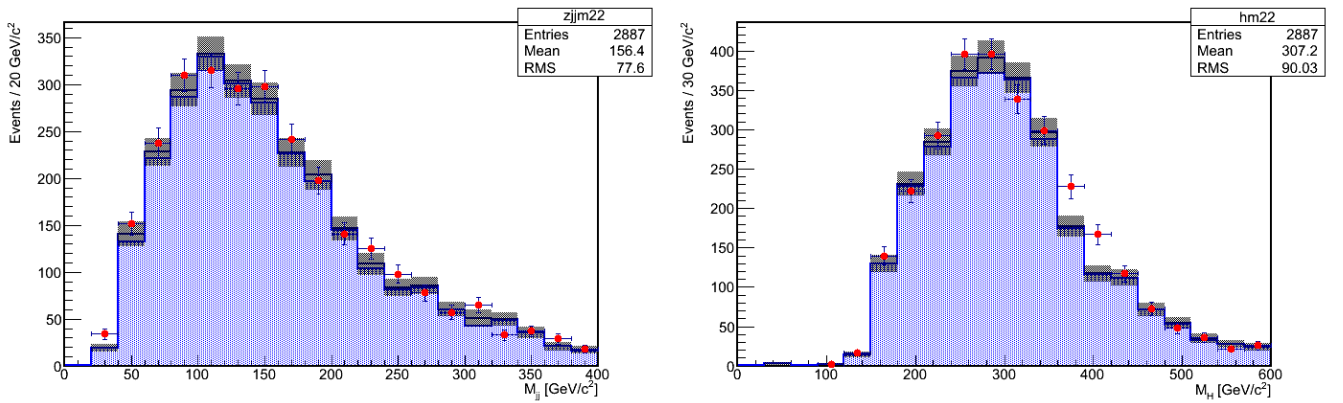


Figure 23: Dijet (left) and “Higgs” (right) invariant masses for  $ee + \mu\mu$  and  $e\mu$  data outside the leptonic Z mass window. Other cuts are detailed in the text. Red dots are  $ee + \mu\mu$  data; white histogram Drell Yan Monte Carlo; blue histogram  $e\mu$  data (plus other small backgrounds).

### 6.3 Diboson background

The less relevant di-boson background amounts to less than 5% in the 0 and 1  $b$ -tag categories and about 10% in the 2  $b$ -tag category. A conservative uncertainty of 30% on the prediction from the simulation results in a 1-3% uncorrelated contribution to the total background prediction error. This uncertainty is suppressed in the analysis as this background is estimated along with the other background sources using the  $\alpha(m_{ZZ})$  method.

## 7 Systematics

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

### 7.1 Luminosity uncertainty

The latest recommendation for the 2012 datasamples is the uncertainty on LHC luminosity of 4.4% [26].

### 7.2 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.

### 7.3 Uncertainties in the background prediction

Uncertainty on background is considered separately and is one of the dominant effects on the exclusion limits. Systematic uncertainties on background affect the expectation of the shape and yield extrapolated from the sidebands as described in section 6.1. Since most of the sources of systematic uncertainties affecting the MC predictions cancel their effect in the computed ratio (to obtain the estimated  $\alpha(m_{ZZ})$ ), the main uncertainty is actually coming from the parameterization of the final sideband distribution.

The way the uncertainties affecting the background shape are discussed in Section 6.1.1. The errors on the fitted parameters are properly propagated into the uncertainty on the shape. The full uncertainty is then taken into account in the statistical treatment of the result.

### 7.4 Uncertainties in the signal expectation

Uncertainty on the signal shape parameterization arise from two sources: theory uncertainty in the BW width ( $\Gamma$ ) parameterization, such as its mass-dependence, and experimental uncertainty in the CB resolution function. The former is taken from comparison of the mass dependent and mass-independent width parameterization used in the POWHEG and JHU generators. The latter is estimated from various sources discussed below, such as jet energy uncertainty, resolution, etc. In general signal shape uncertainties are not the dominant sources of systematics with expected signal yield of several events.

The main systematic uncertainties on signal normalization are summarized in Table 26, and are discussed in more detail in the subsection below. Lepton efficiencies are evaluated with a tag-and-probe approach when one lepton from an inclusive sample of  $Z$  decays serves as a tag

and efficiency for the other lepton is calculated. Effects of jet reconstruction are evaluated with variation of the jet energy and resolution within calibration uncertainties. Effects of pile-up are taken as a difference between reconstruction efficiency with pileup below and above the average expected value, otherwise distributed according to observed values in data. Requirement on the MET significance translates into about 3% inefficiency and the resulting uncertainty does not surpass this value. Uncertainty on the  $b$ -tagging has been evaluated with inclusive sample of  $b$ -jets. Uncertainty on quark-gluon LD selection efficiency was evaluated with predominantly quark jets in the  $\gamma$ +jets sample. Uncertainties in the production mechanism affect both longitudinal momentum of the Higgs, due to PDFs, and transverse momentum of the Higgs, due to QCD initial-state radiation effects. We follow the PDF4LHC recommendation to estimate uncertainty due to PDF knowledge and calculate uncertainty on signal acceptance. We rescale the transverse momentum distribution of the Higgs using HQT as a reference and take the full change in efficiency as systematic uncertainty. Uncertainties on the Higgs cross-section are taken from the Yellow Report which includes uncertainties from QCD renormalization and factorization scales, PDFs, and  $\alpha_s$ . These uncertainties are separated between the gluons fusion and VBF production mechanisms, but gluon fusion uncertainties dominate in the total production cross-section.

Table 26: 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  $R$ ). See text for more details.

source	0 $b$ -tag	1 $b$ -tag	2 $b$ -tag	comment
muons reco		2.7%		tag-and-probe study
electrons reco		4.5%		tag-and-probe study
jet reco		1%–8%		JES-uncert., JER uncert. negligible; correlated between categ
pileup		1-2%		correlated between categ
$b$ -tagging	2-7%	3-5%	10-11%	anti-correlated between categ.
MET	–	–	3-4%	loose requirement
production mechanism (PDF)		2-4%		PDF4LHC, acceptance only
production mechanism (WBF)		1%		
production mechanism (lineshape)		0-3%		only for $M_H > 400$
luminosity		4.4%		same for all analyses
Higgs cross-section (for $R$ )		13–18%		detailed table from YR available

#### 7.4.1 Lepton energy scale, resolution, selection, and trigger

Lepton trigger and selection is common among several  $H \rightarrow ZZ$  analyses and we benefit from common study based on tag-and-probe techniques. In particular, recent studies within the framework of Ref. [4] 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, 1.0(2.0)% due to muon (electron) momentum/energy scale.

#### 7.4.2 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. Preliminary estimate with 2010 data show that jet energy uncertainty could be kept within about 4% and resolution within 10%. For more details see Refs. [27, 28]. Since the background

is extracted from sidebands in data, the systematics due to JES and jet resolution uncertainty affects only signal efficiency and potentially  $m_{ZZ}$  distribution.

Our preliminary estimates show that JES variation by  $\pm 1\sigma$  changes reconstruction efficiency of a 400 GeV Higgs by about 5%. In Fig. 24 the effect of a possible JES bias ( $\pm 1\sigma$ ) is shown on some fundamental variables. The effect on the jets transverse momentum and dijet invariant mass is sizable and it drives the bias on the acceptance while the effect is very small on the Higgs candidate invariant mass, thanks to the  $m_{JJ}$  kinematical fit. The small effect on the angular LD is due to the bias in the boosts applied to compute the angles in the Higgs and Z reference frames. The bias on the QG discriminant and  $b$ -tag categorization is negligible, as expected. Detailed study as a function of Higgs mass hypothesis is provided in Table 27.

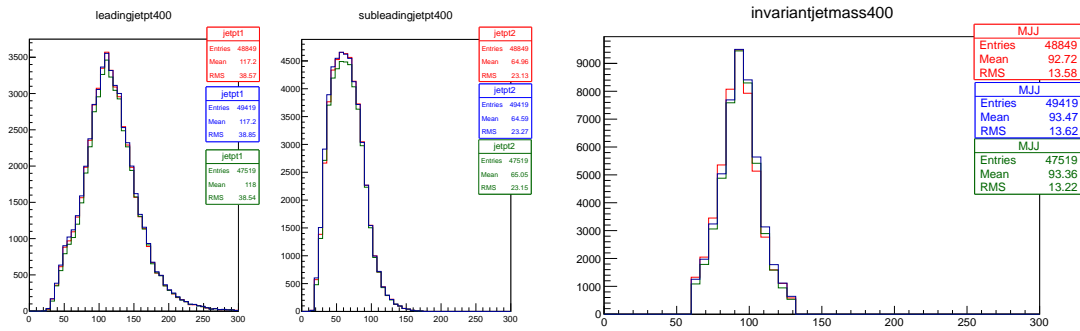


Figure 24: Distribution of leading (left) and subleading (middle) jet transverse momentum, and Higgs candidate mass (right) after final selection requirements with nominal Jet Energy Scale (JES) and with JES changed by  $\pm 1\sigma$ .

The bias on the signal shape due to the jet energy scale uncertainty was evaluated in 2011, but found to be small. The effect of the jet resolution uncertainty on the signal was evaluated in 2011 by applying an additional smearing to the jets and comparing to the same sample without additional smearing. As the background is evaluated directly from the data it is not expected that the jet energy resolution will have a significant effect.

### 7.4.3 Pile-up reweighing

As mentioned in Section 3.1, the number of true interactions per bunch crossing in the simulated samples was re-weighted 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; we use 65.93 mb and 72.87 mb as recommended by the CMS pileup group [29], which is  $\pm 5\%$  difference with respect to the central value 69.40 mb. The re-estimated distributions and the central value are compared in Fig. 25. Figure 26 shows that the difference in the distributions of the number of reconstructed good vertexes between data and MC is covered by the  $\pm 5\%$  shift of pileup distributions in data. Here, a good vertex must have at least three tracks and must be within 24 cm (2 cm) of the nominal center of the detector along (perpendicular to) the beam axis.

We reweighed the number of true interactions in the MC to match the shifted distributions in data in Fig. 25 and re-compute the signal efficiency. This leads to a change of  $\approx 1\text{--}2\%$  in the signal efficiency for  $m_H < 600 \text{ GeV}/c^2$ , approximately independent of lepton channel and  $b$ -tag category, as shown in Table 28.

Table 27: Variations in efficiency due to Jet Energy Scale Uncertainties for several signal samples and categories

$m_H$ [GeV]	Category	Nominal	JES+	JES-
200	ele	0.0196	0.0203	0.0184
	mu	0.0255	0.0264	0.0239
	0tag	0.0186	0.0237	0.0184
	1tag	0.0223	0.0220	0.0200
	2tag	0.0044	0.0044	0.0040
300	ele	0.0537	0.0550	0.0525
	mu	0.0678	0.0694	0.0663
	0tag	0.0561	0.0573	0.0548
	1tag	0.0502	0.0514	0.0490
	2tag	0.0153	0.0156	0.0149
400	ele	0.0726	0.0735	0.0707
	mu	0.0903	0.0913	0.0877
	0tag	0.0732	0.0741	0.0691
	1tag	0.0651	0.0659	0.0646
	2tag	0.0245	0.0247	0.0247
525	ele	0.0796	0.0797	0.0793
	mu	0.0951	0.0951	0.0947
	0tag	0.0802	0.0804	0.0799
	1tag	0.0660	0.0660	0.0657
	2tag	0.0285	0.0284	0.0284
600	ele	0.0732	0.0774	0.0764
	mu	0.0871	0.0922	0.0909
	0tag	0.0759	0.0759	0.0793
	1tag	0.0576	0.0610	0.0599
	2tag	0.0268	0.0281	0.0280

#### 7.4.4 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  $\eta$  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  $\eta$  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 30 and 29 give the  $b$ -tagging systematic uncertainty for the signal, for muons and electrons. The systematic effect is computed as the ratio of the number of tagged jets with a  $SF$  varied by plus and minus its uncertainty to the number of tagged jets with the nominal  $SF$ . The uncertainty is reported 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.

Table 28: The relative systematic uncertainty in percentage (%) from PU reweighing.

Electron			
$m_H [\text{GeV}/c^2]$	$N_{\text{btag}} = 0$	$N_{\text{btag}} = 1$	$N_{\text{btag}} = 2$
200	+1.57 -1.91 +1.04	+0.34 -0.14 +0.99	+0.01 +0.08 +1.07
300	-0.93 +0.38	-0.77 +0.26	-1.01 +1.64
400	-0.37 +0.13	-0.31 +1.07	-1.50 -0.11
525	-0.15 +1.03	-1.06 +0.95	-0.06 +0.16
600	-0.99 +0.39	-0.96 +0.22	-0.27 +0.51
700	-0.39 +0.98	-0.23 +0.05	-0.60 +0.97
800	-0.83 +0.61	-0.22 +1.59	-1.25 +0.70
900	-0.69 +0.99	-1.39 +0.35	-0.58 +0.24
1000	-1.10	-0.53	-0.45

Muon			
$m_H [\text{GeV}/c^2]$	$N_{\text{btag}} = 0$	$N_{\text{btag}} = 1$	$N_{\text{btag}} = 2$
200	-0.08 -0.31 +0.76	+1.06 -1.05 +1.20	+1.15 -1.26 +0.37
300	-0.66 +0.46	-1.11 +0.31	-0.40 +2.31
400	-0.46 +0.39	-0.39 +0.63	-2.04 +0.25
525	-0.40 +0.94	-0.52 +0.69	-0.45 +0.28
600	-0.87 +0.68	-0.78 +0.20	-0.61 +0.95
700	-0.76 +0.62	-0.28 +1.42	-1.19 +1.14
800	-0.70 +0.67	-1.25 +0.17	-1.35 +2.21
900	-0.65 +1.19	-0.25 +1.10	-1.94 +3.09
1000	-1.00	-0.94	-2.76

Table 29: Systematic uncertainty on the signal in the electron channel.

$H_{mass}$	0-tag		1-tag		2-tag	
	$\frac{SF_{Up}}{SF_{Central}}$	$\frac{SF_{Down}}{SF_{Central}}$	$\frac{SF_{Up}}{SF_{Central}}$	$\frac{SF_{Down}}{SF_{Central}}$	$\frac{SF_{Up}}{SF_{Central}}$	$\frac{SF_{Down}}{SF_{Central}}$
	$\frac{SF_{Up}}{SF_{Central}}$	$\frac{SF_{Down}}{SF_{Central}}$	$\frac{SF_{Up}}{SF_{Central}}$	$\frac{SF_{Down}}{SF_{Central}}$	$\frac{SF_{Up}}{SF_{Central}}$	$\frac{SF_{Down}}{SF_{Central}}$
200	0.95	1.02	1.08	1.01	1.08	0.87
250	0.97	1.03	1.05	0.94	1.06	0.96
300	0.97	1.01	1.02	0.99	1.11	0.91
350	0.97	1.03	1.02	0.98	1.09	0.92
400	0.97	1.02	1.03	1.00	1.08	0.91
425	0.98	1.04	1.02	0.96	1.08	0.94
525	0.96	1.03	1.04	0.98	1.07	0.91
600	0.97	1.03	1.01	0.96	1.09	0.91

Table 30: Systematic uncertainty on the signal in the muon channel.

$H_{mass}$	0-tag		1-tag		2-tag	
	$\frac{SF_{Up}}{SF_{Central}}$	$\frac{SF_{Down}}{SF_{Central}}$	$\frac{SF_{Up}}{SF_{Central}}$	$\frac{SF_{Down}}{SF_{Central}}$	$\frac{SF_{Up}}{SF_{Central}}$	$\frac{SF_{Down}}{SF_{Central}}$
	$\frac{SF_{Up}}{SF_{Central}}$	$\frac{SF_{Down}}{SF_{Central}}$	$\frac{SF_{Up}}{SF_{Central}}$	$\frac{SF_{Down}}{SF_{Central}}$	$\frac{SF_{Up}}{SF_{Central}}$	$\frac{SF_{Down}}{SF_{Central}}$
200	0.96	1.03	1.04	1.00	1.05	0.82
250	0.97	1.03	1.01	0.99	1.12	0.83
300	0.98	1.03	1.01	0.98	1.09	0.90
350	0.97	1.02	1.00	0.98	1.13	0.92
400	0.97	1.03	1.02	1.01	1.07	0.86
425	0.98	1.03	1.01	0.99	1.10	0.91
525	0.97	1.03	1.01	0.97	1.11	0.95
600	0.97	1.02	1.04	1.01	1.06	0.89

#### 7.4.5 MET uncertainty

MET affects directly only the 2  $b$ -tag category. 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. The uncertainty is computed following the method also used in the low mass analysis (see section ??). Requirement on the MET significance translates into about 3% inefficiency and the resulting uncertainty does not surpass

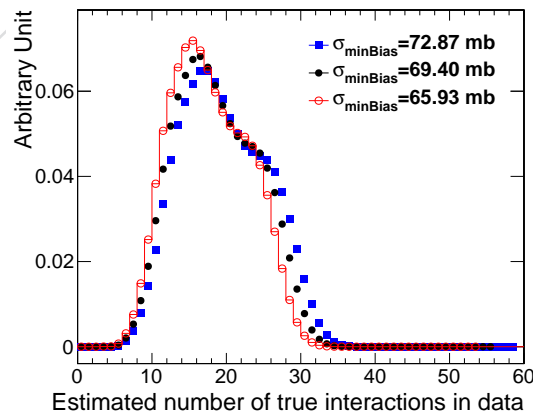


Figure 25: The estimated number of true interactions in 2012 data, assuming different values of minimum-bias cross section. The central value is 69.4 mb (solid circles).



this value, as shown in figure 27.

#### 7.4.6 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 CT10 [30], MSTW2008NLO [31] and NNPDF2.1 [32] and their error sets. Table 7.4.6 summarizes the resulting acceptance uncertainties. The envelope of the various PDF sets is used as the total uncertainty, as recommended and amounts to 2-4% without strong dependence on  $b$ -tag category. The uncertainty noticeably increases for very high Higgs masses.

Table 31: Summary of systematic uncertainties on the signal acceptance following PDF4LHC recommendations.

PDF	$M_H = 200 \text{ GeV}$			$M_H = 600 \text{ GeV}$			$M_H = 900 \text{ GeV}$		
	0 $b$ -tag	1 $b$ -tag	2 $b$ -tag	0 $b$ -tag	1 $b$ -tag	2 $b$ -tag	0 $b$ -tag	1 $b$ -tag	2 $b$ -tag
CT10	+1.8% -2.3%	+1.9% -2.4%	+2.1% -2.1%	+1.3% -2.0%	+1.4% -2.2%	+2.1% -3.5%	+2.8% -3.8%	+3.0% -3.9%	+4.0% -6.0%
all categories		+1.9% -2.3%			+1.5% -2.4%			+3.0% -4.2%	
MSTW2008NLO	1.1% -0.1%	+1.2% -0.1%	+1.4% -0.4%	+1.2% +0.4%	+1.2% +0.5%	+2.0% +0.9%	+2.3% +0.4%	+2.4% +0.3%	+3.4% +0.7%
all categories		1.2% -0.1%			+1.3% -0.5%			+2.5% +0.4%	
NNPDF2.1	+2.3% +0.7%	+2.5% +0.7%	+2.7% +0.7%	+2.1% +1.0%	+2.0% +1.0%	+3.5% +1.7%	+3.9% +1.6%	+4.2% +2.5%	+5.8% +2.3%
all categories		+2.4% +0.7%			+2.3% +1.1%			+4.3% +1.7%	
Total	+2.3% -2.3%	+2.5% -2.4%	+2.7% -2.1%	+2.1% -2.0%	+2.0% -2.2%	+3.5% -3.5%	+3.9% -3.8%	+4.2% -3.9%	+5.8% -6.0%
all categories		+2.4% -2.3%			+2.3% -2.4%			+4.3% -4.2%	

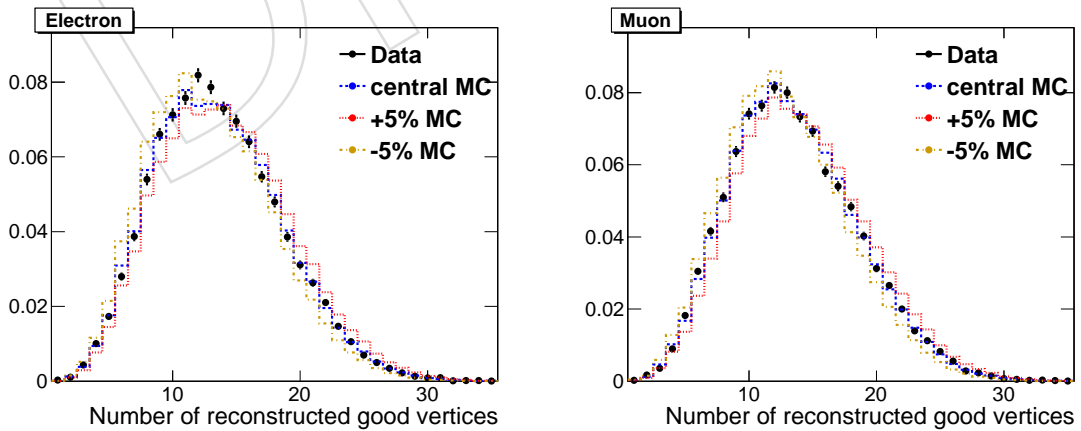


Figure 26: The reconstructed number of good vertexes in data (points with error bars) and in reweighted MC (various lines), from the electron channel (left) and the muon channel (right). The number of true interactions in MC has been reweighted to match the estimated distributions in data, assuming three different values of minimum-bias cross section.

Figure 27: Efficiency in data and Monte Carlo (after  $t\bar{t}$  subtraction) for various MET significance cuts in the 2 btag category after the preselection cuts. The difference between data and Monte Carlo efficiencies (magnified by 10) is also shown.

We additionally estimate the uncertainty that originates from the fact that the analysis has been tuned using gluon fusion based simulation while a real signal contains a mixture of events produced by gluon fusion and VBF. Here we compute the difference in signal acceptance between the two production mechanisms in Monte Carlo and multiply this difference with the expected fraction of VBF production, leading to a global uncertainty on the production cross section.

Table 32: Summary of systematic uncertainties due the VBF.

$M_H$	$M_H = 200 \text{ GeV}$			$M_H = 400 \text{ GeV}$			$M_H = 600 \text{ GeV}$		
	0 $b$ -tag	1 $b$ -tag	2 $b$ -tag	0 $b$ -tag	1 $b$ -tag	2 $b$ -tag	0 $b$ -tag	1 $b$ -tag	2 $b$ -tag
$\Delta_{eff}$	20%	15%	6%	-2%	4%	4%	-8%	6%	-0.3%
total	7%			11%			13%		
uncertainty	2.4%	1.8%	0.7%	-0.16 %	0.3 %	0.3 %	-1.4 %	1 %	0%
total	2%			0.1%			-1.4%		

Additional uncertainties arise due to uncertainties on the Higgs signal shape. The shape uncertainty is evaluated as described in [33] and contributes in two ways: Due to the mass-dependence of the selection efficiency, the total signal efficiency is affected by the line shape. Efficiency curves for the nominal and alternative line shapes are shown in figure 28. The uncertainty is negligible below 400 GeV and rises to  $\sim 3\%$  at 600 GeV, with onl small dependence on btag category.

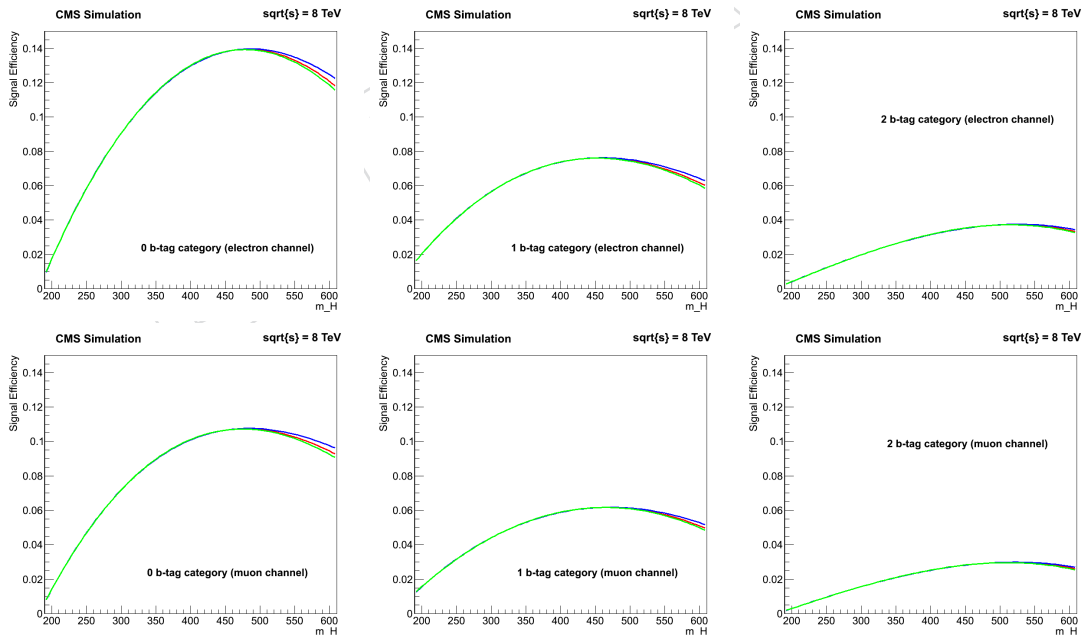


Figure 28: Signal selection efficiency for the nominal line shape (red) and alternative shapes (green/blue) for electrons (top) and muons (bottom) for 0/1/2 btags (left/middle/right).

Additionally the line-shape used in the CLs procedure is re-extracted with the alternative line-shape models (See Figure 29). 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 by the uncertainty. In the worst case (the highest mass we consider), the peak-position shifts by  $\sim 2$  GeV (compared to a sigma of 60 GeV) and the sigma changes by  $\sim 1$  GeV. Due to the miniscule effect of this uncertainty, it is not propagated further.

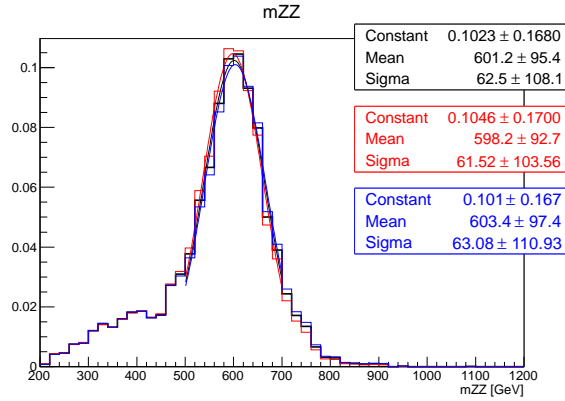


Figure 29: Reconstructed  $M=600$  GeV Higgs signal (area normalized) with the nominal line-shape (black) and systematic variations (blue/red). Gaussian fits to the core of the distribution are overlaid.

## 8 Signal Shape Parameterization

The expectations for signal are obtained in the analysis by applying all the selection criteria to the signal events obtained with the official MC samples, properly scaled by the reconstruction efficiencies and the Pile-Up distribution.

Those expectations can be understood as the yield normalization, which is described in terms of signal efficiency, and the shape of any interesting variable. In our analysis the final variable used to distinguish the signal on top of the expected background is the invariant mass of the two leptons forming the leptonic Z and the two jets forming the hadronic Z. This 4-object mass is directly related to the mass of the Higgs candidate and therefore is significantly different for events containing a real Higgs (signal event) or just coming from random combination of leptons and jets produced in the event (background event).

The background shape is obtained as described in Section 6.1.

The signal distribution varies depending on Higgs mass hypothesis and it requires a more detailed and mass-dependent treatment. For each MC sample available we need to obtain the distributions of events and obtain the correct parameterization for it in order to include it in the interpretation of the results.

The narrow width approximation used in the 2011 analysis breaks down at high Higgs mass (typically  $> 400$  GeV) due to the very large Higgs width ( $> 70$  GeV). The problem has been discussed in details in Ref. [34] 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 [35]. In 2011, 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  $gg \rightarrow ZZ$  background becomes very large, as recently discussed in Ref. [36]. 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. [36] to estimate the uncertainty due to missing higher perturbative order on the interference and the simulated line shape is reweighted 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 in the analysis in order to account for the correct distributions.

Regarding the shape, the parameterization of the signal distribution is done by splitting the accepted events in two sets:

- **“Matched events”** are the events in which the four reconstructed objects that are used to build the Higgs candidate are correctly identified from the original underlying objects at the generator level (in the MC samples).
- “Unmatched events”** are the events in which any of the used objects, usually a jet, is not actually coming from the original underlying decay products of the Higgs.

since the second component is not actually expected to give the proper mass of the generated Higgs in the event, the behaviour of the two subsamples is slightly different, being the first one an actual resonance and the second the shape of the combinatorial background, although still related to the actual Higgs mass since most of the objects used in the mass reconstructions are actually the correct ones.

To parameterize the shape of the total distribution, it is easier to parameterize each component independently and add the resulting functions. This dramatically reduces the complexity of the fits to be performed. In fact it has been checked that without the splitting the fitting of the shapes becomes impossible to perform due to the big difference in size of the two samples.

The parameterization is performed using a Double Crystall-Ball function (i.e. a Gaussian core with powerlaw tails on both sides) for the matched subsample. A single Crystal-Ball function has turned to be a realistic representation of the unmatched subsample. The addition of the two functions gives a very good parameterization of the obtained shape, as shown in several examples from Fig. 30 to 34 where the expected distributions are compared to the addition of the two components, shown with separated colours in the plots.

The fitted parameterizations for the available MC samples are used to produce the interpolated parameterization in the 73 mass points for which the statistical interpretation of the result is performed, as described in Section 9.

## 9 Statistical Analysis and Results

We determine the expected upper limits to the Standard Model Higgs production cross section as a function of the Higgs boson mass. We use the official tool developed by the CMS Higgs combination group [37] that supports different methods, including Bayesian approach,

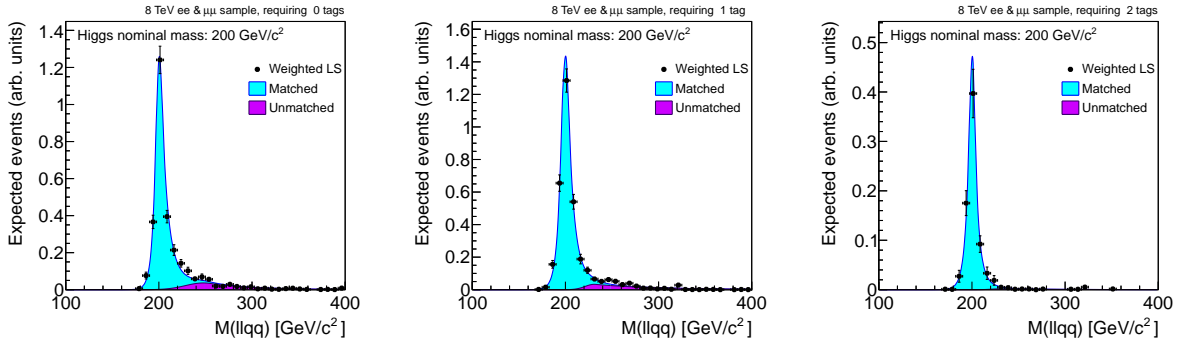


Figure 30: Reconstructed 4-object mass distribution of the accepted signal events for nominal Higgs mass of 200 GeV for the 2012 analysis. In order to parameterize the shape, we split the sample in two samples and fit the components as described in the text. The resulting shape is compared here to the actual obtained distribution and the two components are shown with distinctive colours.

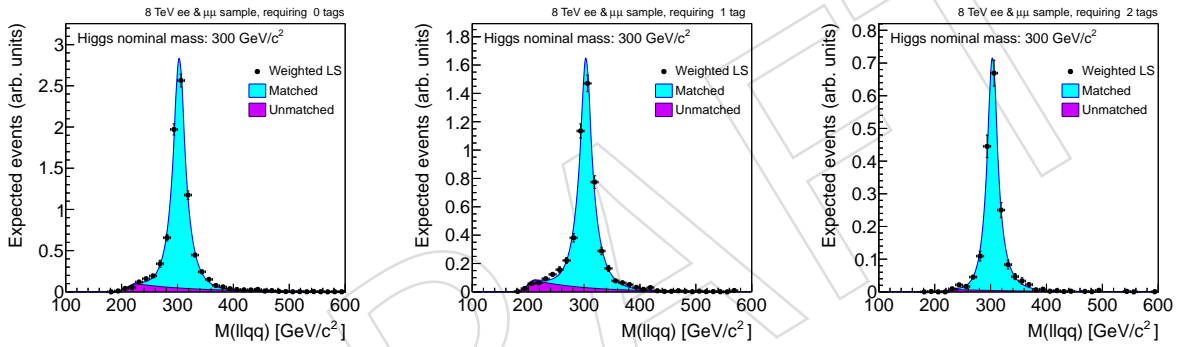


Figure 31: Reconstructed 4-object mass distribution of the accepted signal events for nominal Higgs mass of 300 GeV for the 2012 analysis. Other details are as described in caption to Fig. 30.

frequentist profile likelihood, and Feldman-Cousins [38] methods, and the modified frequentist CLs method [39] with Cousins-Highland integration of nuisance parameters for the treatment of systematic uncertainties [40]. The tool uses the `Roostats` [41] engine from ROOT as internal implementation. We determine the expected upper limit to the Higgs boson production cross section times branching fraction to  $\ell\ell b\bar{b}$ . The limit is expressed as ratio  $r$  of the determined upper limit to the cross section times branching ratio 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.

There are two main possible general approaches to determine the limit to the Higgs cross sections.

The simplest approach, referred to as “cut and count” analysis, uses only the number of events selected within a given window in the reconstructed Higgs mass,  $M_{\ell\ell jj}$ , around an assumed value  $M_H$  of the Higgs boson mass. The Higgs cross section limit is determined from the expected number of signal and background events passing the selections  $s$  and  $b$  respectively. We combine using Poissonian statistics the counting information from the two channels with electrons and muons. The sources of systematic uncertainties on  $s$  and  $b$  are taken into account

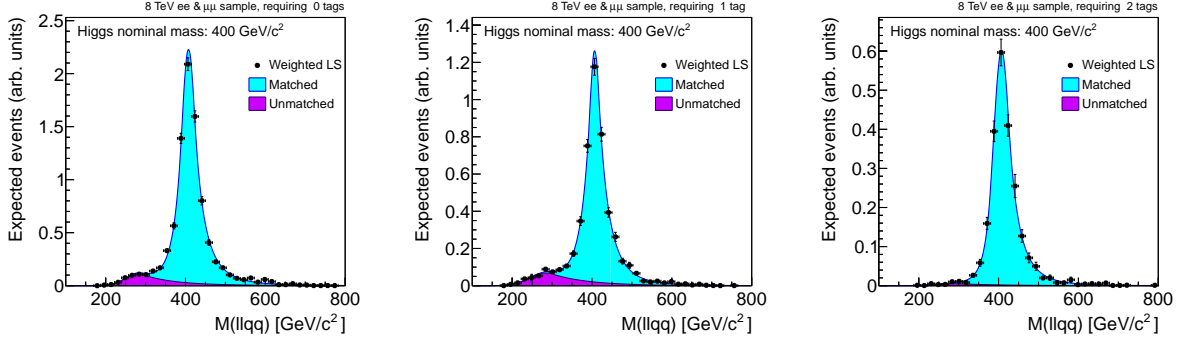


Figure 32: Reconstructed 4-object mass distribution of the accepted signal events for nominal Higgs mass of 400 GeV for the 2012 analysis. Other details are as described in caption to Fig. 30.

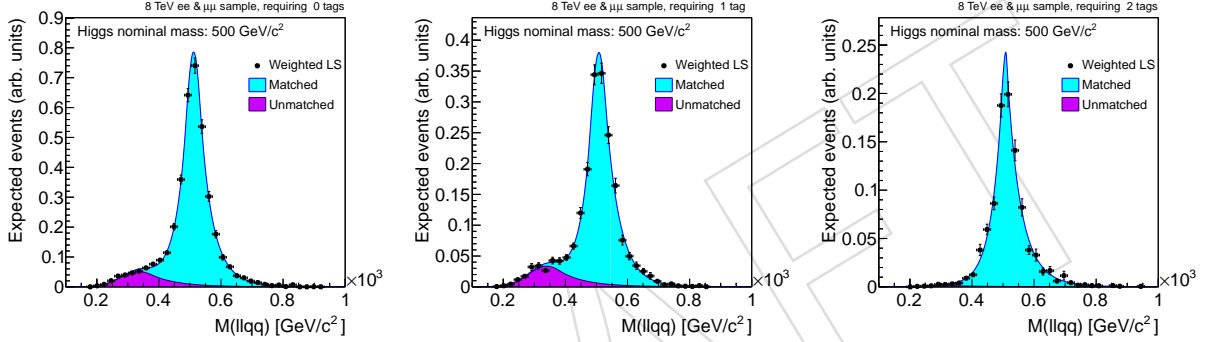


Figure 33: Reconstructed 4-object mass distribution of the accepted signal events for nominal Higgs mass of 500 GeV for the 2012 analysis. Other details are as described in caption to Fig. 30.

in the determination of the limit assuming log-normal distributions of the nuisance parameters.

Another possible approach, referred to as “shape analysis”, takes into account the measured distribution of variables that discriminate Higgs signal events against background events. In the case of the  $H \rightarrow ZZ \rightarrow b\bar{b}\ell^+\ell^-$  analysis, the main discriminant variable is the reconstructed Higgs boson mass,  $M_{\ell\ell jj}$ , which is peaked around the true Higgs boson mass,  $M_H$ , for the signal and has a broader distribution for background processes. We can develop an analysis that does not select signal events cutting on  $M_{\ell\ell jj}$ , and we can use the distribution of  $M_{\ell\ell jj}$  for selected events, comparing to the expected distribution from signal and background. 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. The shape analysis can be implemented in the Higgs combination tool using an unbinned approach with an extended likelihood function defined by:

$$-\ln \mathcal{L} = \sum_{j=1}^k \left[ s_j + b_j - \sum_{i=1}^{n_j} \left( s_j \mathcal{P}_s^{(j)}(m_i^{(j)}) + b_j \mathcal{P}_b^{(j)}(m_i^{(j)}) \right) \right], \quad (12)$$

where  $k$  is the number of channels (in our case  $k = 2$ , and  $j = 1, 2$  correspond to the electron and muon channels),  $n_j$  is the number of selected candidate events in the channel  $j$ ,  $s_j$  and  $b_j$  are



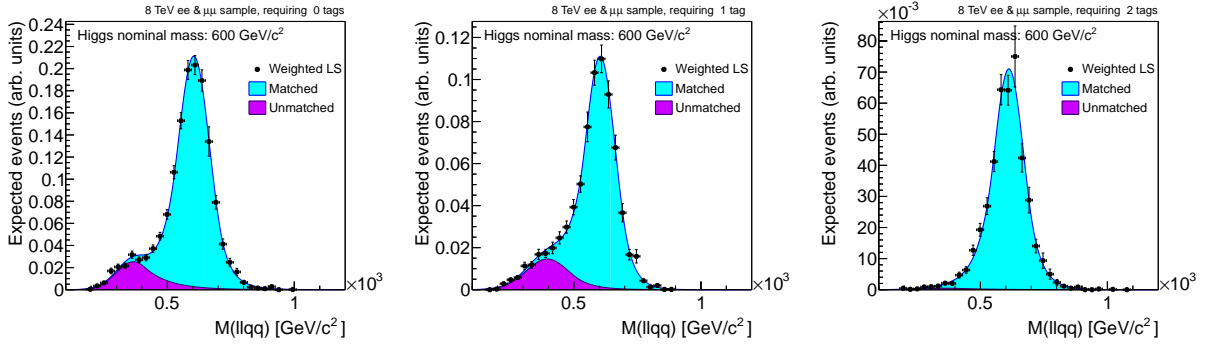


Figure 34: Reconstructed 4-object mass distribution of the accepted signal events for nominal Higgs mass of 600 GeV for the 2012 analysis. Other details are as described in caption to Fig. 30.

the expected signal and background events,  $\mathcal{P}_s^{(j)}$  and  $\mathcal{P}_b^{(j)}$  are the probability density functions for signal and background, and  $m_i^{(j)}$  are the  $n_j$  values of  $M_{\ell\ell jj}$  of the selected candidates in the channel  $j$ . As alternative, the distribution of  $M_{\ell\ell jj}$  can be sub-divided into bins, and in that cases the number of selected events in each bin is considered with the corresponding expected number of events from signal and background in that bin, and the information from all bins is combined using a Poissonian likelihood, similarly to the “cut and count” case, but with more channels.

Since in our final state we are able to fully reconstruct the decay products of the Higgs boson, the sensitivity of the analysis is largely increased by using a shape-based treatment of the expected and observed distribution of the invariant mass of the Higgs candidate (i.e. the 4-object mass). Therefore this sets the method to be used.

For that purpose, we parameterize the background distribution by using the technique given in Section 6.1, by fitting an empirical shape as described there.

The signal distribution is obtained by the parameterization of the signal shape obtained as described in section 8.

The signal efficiency of the selection described in section 5 is evaluated as the ratio between the number of selected events in each of the six channels 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 polinomial in order to be estimatated for those Higgs mass hypothesis where no Monte-Carlo sample is available, as shown in Fig. 35.

Based on the background and signal expectations and their correspondent uncertainties, we estimate the expected limit on the ratio of the SM Higgs boson production cross section to the SM Higgs expectation between 200 and 600 GeV as shown in Fig. 36. Limits on the SM product cross section times branching fraction for  $H \rightarrow ZZ$  are presented in Fig. 37.

According to the result after unblinding, the result of this analysis exclude the existence of a resonance with properties as those of the SM Higgs in the mass range between 295 and 460 GeV.



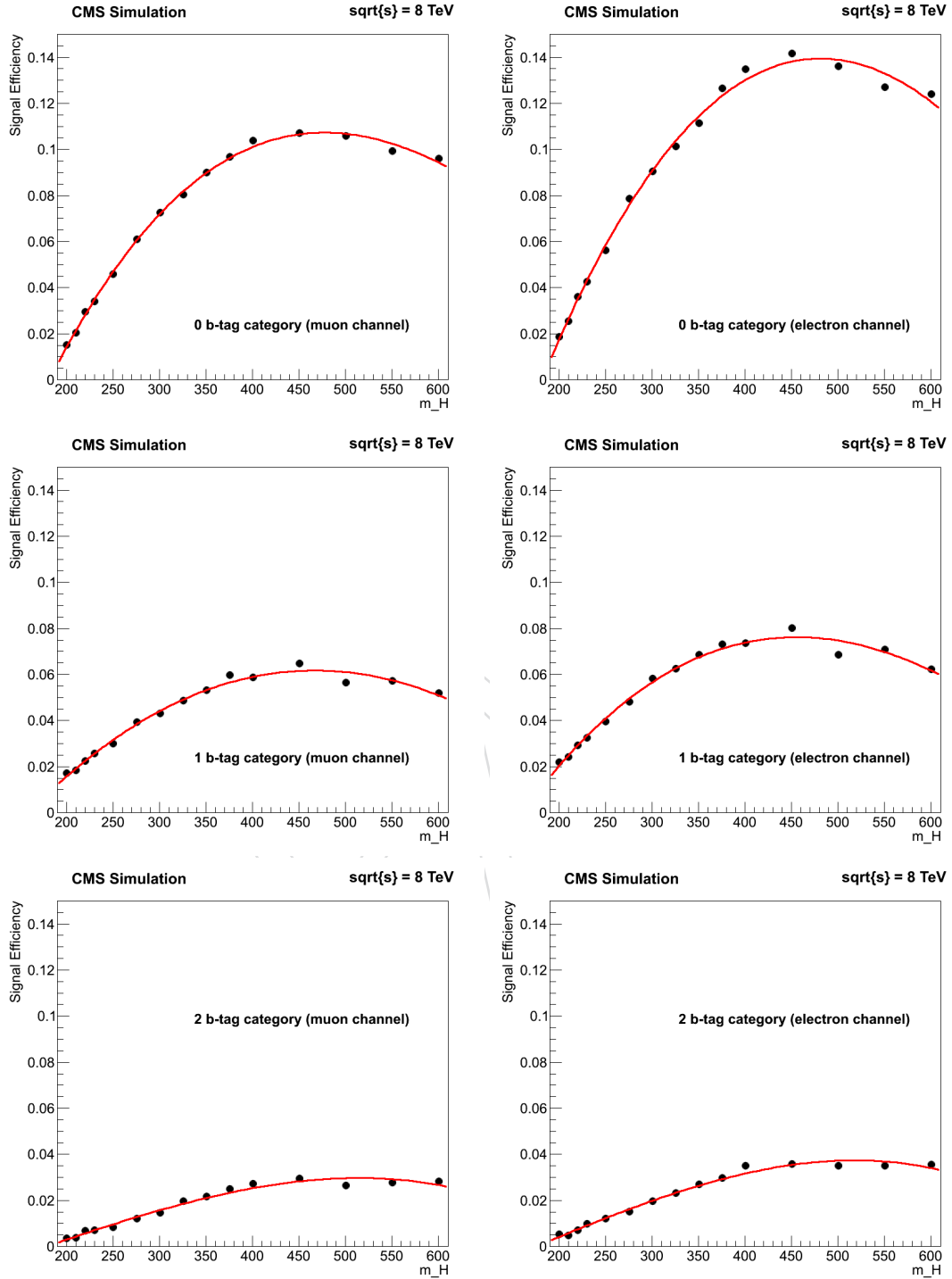


Figure 35: Parameterization of signal efficiency as a function of Higgs mass hypothesis in 0  $b$ -tag (top), 1  $b$ -tag (middle), 2  $b$ -tag (bottom) categories and in the muon (left) and electron (right) channels.

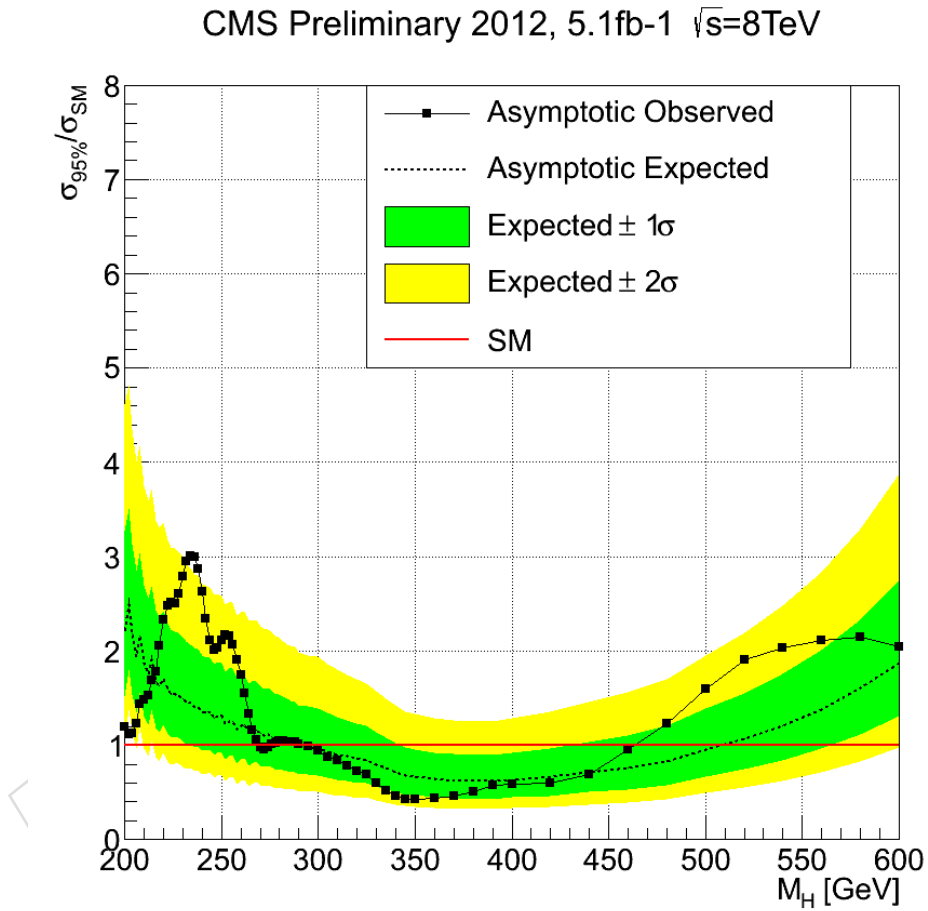


Figure 36: Limit on the expected 95% CL upper limit on the product of the Higgs boson production cross section and the branching fraction of  $H \rightarrow ZZ$  (dash line) and observed upper limit (black dots.) Yellow and Green bands represent the 68% and 95% ranges of expectation.

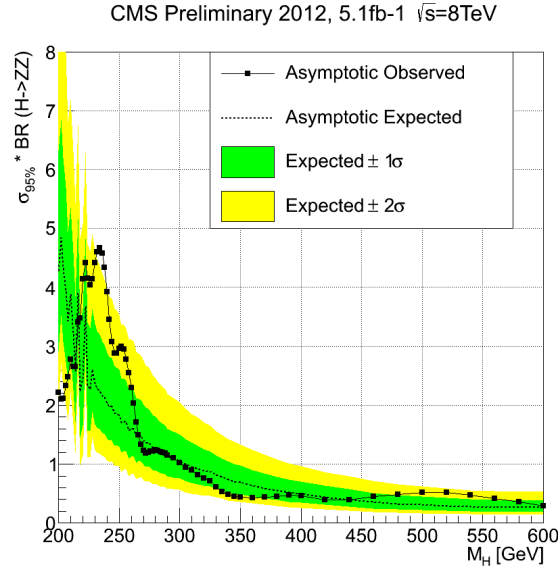


Figure 37: Observed (dashed) and expected (solid) 95% C.L. upper limit on the product of the production cross section and branching fraction for  $H \rightarrow ZZ$  obtained with the  $\text{CL}_s$  technique. The 68% and 95% ranges of expectation for the background-only model are also shown with green and yellow bands, respectively.

## 10 Updating the 7-TeV Results

As discussed in section 8, the Higgs mass distribution for nonimal masses beyond 400 GeV is not correctly modeled in the MC samples used in the analysis. This also applies to the results obtained with the 2011 dataset taken at 7 TeV collisions.

Therefore, we have worked on updating the 2011 result, that was published in [42], to incorporate the lineshape reweighting, with the method described in the 2012 analysis. It should be noted that since the time the 2011 analysis was published, some mistakes were found on the way the datacards (used in the limit calculation) were constructed, which implies the result was not completely correct.

We have taken the opportunity to correct the mistake and at the same time improve the calculated limit by including the correct mass lineshape. It should be remarked that these are the only changes that are implemented to the analysis described in [42] apart from some minor bugs found since the previous analysis was released. We are still using all the selection and objects from the analysis as they were originally developed.

The first step to update the 2011 limit was to compute the shapes for signal used in the limit extraction, as done in section 8. We are using the same strategy, splitting the sample between “matched” and “unmatched” events.

The “matched” events are fitted with a Double CrystalBall function, and the “unmatched” part it is fitted with a simple CrystalBall. It should be noted that in the original analysis, an additional triangle function was included, but it does not seem to be needed any longer. Figures from 38 to 42 show some examples of the fits performed to parameterize the signal shape. Fits for samples with nominal Higgs masses of 400 GeV and higher include the proper lineshape reweighting. As can be seen in the plots, the lineshape is well described by the parameterization.

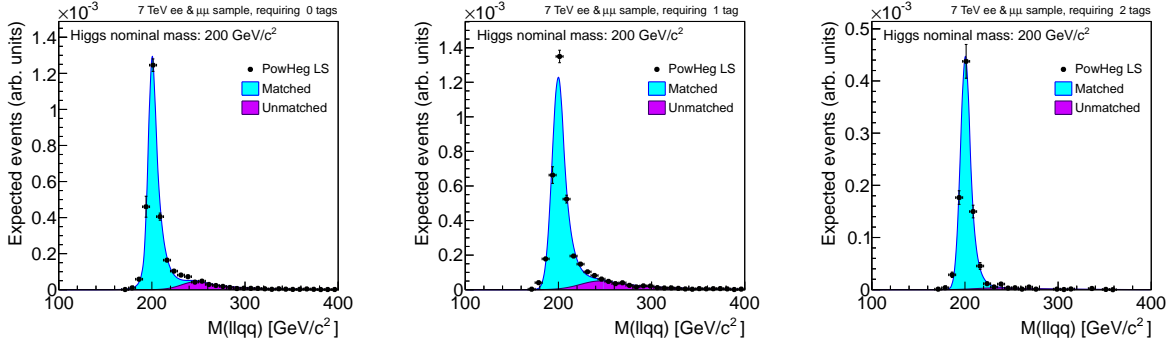


Figure 38: Reconstructed 4-object mass distribution of the accepted signal events for nominal Higgs mass of 200 GeV for the 2011 analysis. In order to parameterize the shape, we split the sample in two samples and fit the components as described in the text. The resulting shape is compared here to the actual obtained distribution and the two components are shown with distinctive colours.

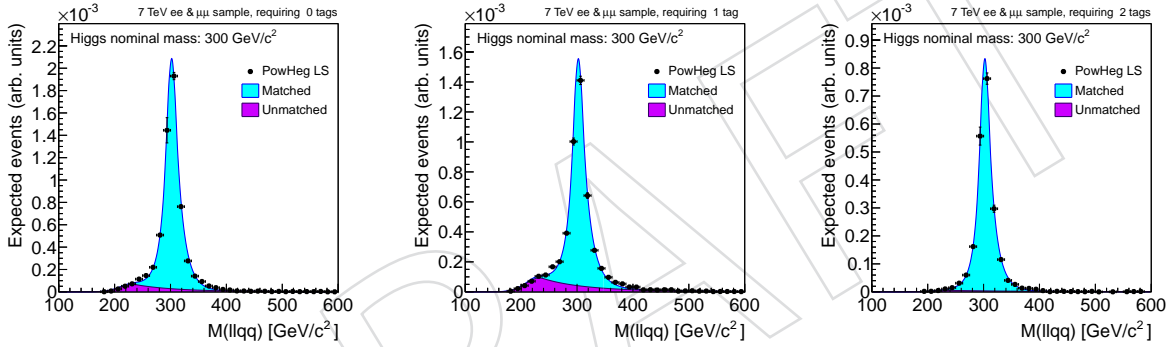


Figure 39: Reconstructed 4-object mass distribution of the accepted signal events for nominal Higgs mass of 300 GeV for the 2011 analysis. Other details are as described in caption to Fig. 38.

The second step is to use the new shapes and the modified yields to compute the limit related to the 2011 results similarly as how it is done in section 9 for the 2012 analysis.

The updated expected limit is shown in Fig. 43. It should be remarked that it significantly differs from the one already published [42] due to the mistakes in the limit calculation. The result presented here overrides the previous one.

## 11 Conclusion

In this report we present the description of a search for a SM-like Higgs boson with a mass between 200 and 600 GeV in the decay  $H \rightarrow ZZ$  when one of  $Z$  decays as  $Z \rightarrow \ell^- \ell^+$  and the other as  $Z \rightarrow q\bar{q}$  using the data collected between March and June of 2012 with  $pp$  collisions at 8 TeV.

The result of this analysis exclude the existence of a resonance with properties as those of the SM Higgs in the mass range between 295 and 460 GeV.

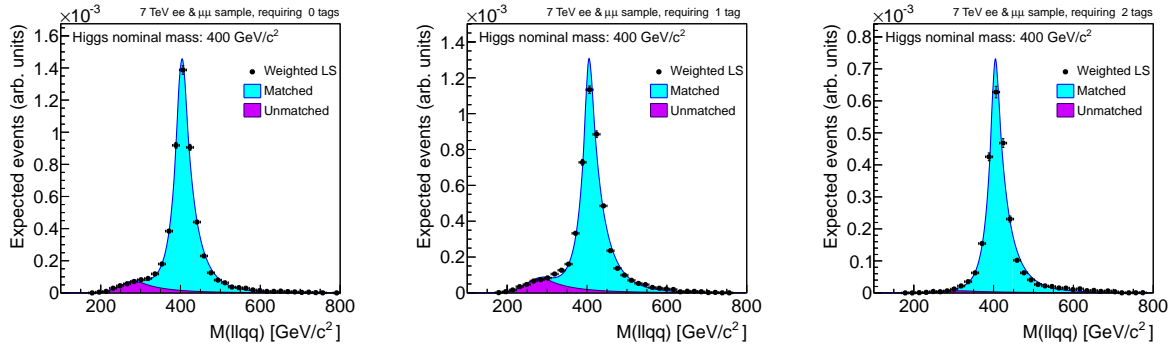


Figure 40: Reconstructed 4-object mass distribution of the accepted signal events for nominal Higgs mass of 400 GeV for the 2011 analysis. Other details are as described in caption to Fig. 38.

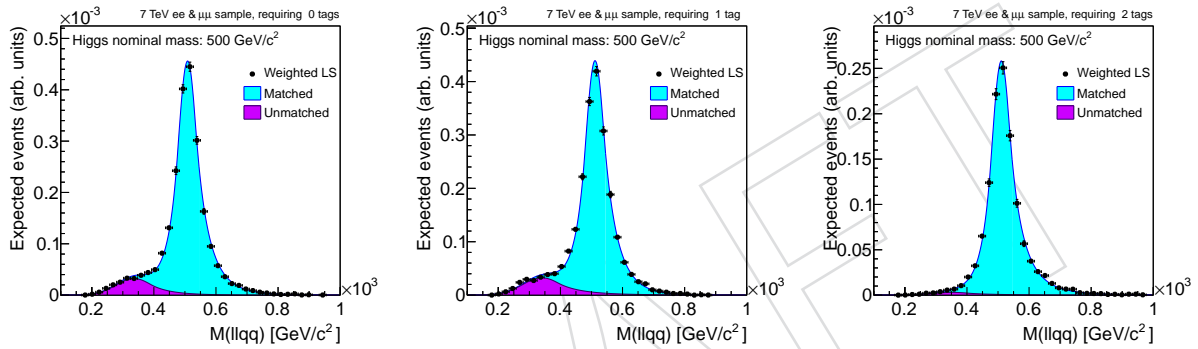


Figure 41: Reconstructed 4-object mass distribution of the accepted signal events for nominal Higgs mass of 500 GeV for the 2011 analysis. Other details are as described in caption to Fig. 38.

792 In addition, the result of the previous analysis performed with the data collected at 7 TeV has  
 793 been updated to include a more suitable description of the mass lineshape of the Higgs. The  
 794 results shown here are expected to override the one published in [42].

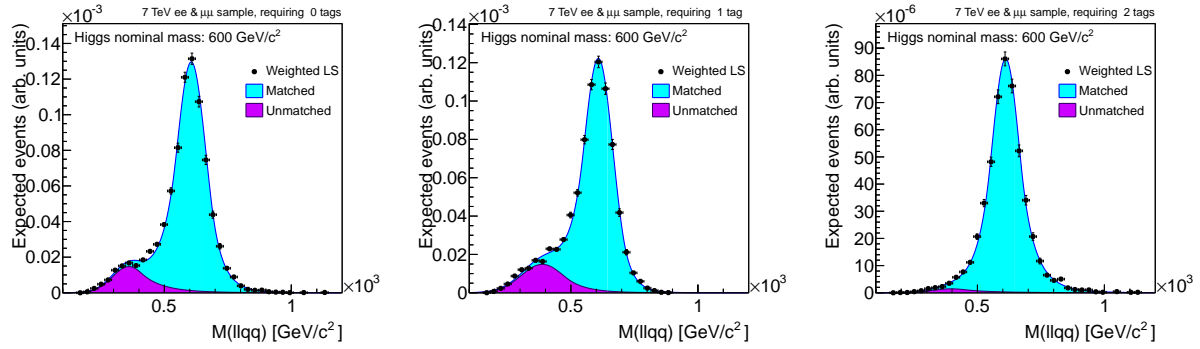


Figure 42: Reconstructed 4-object mass distribution of the accepted signal events for nominal Higgs mass of 600 GeV for the 2011 analysis. Other details are as described in caption to Fig. 38.

Asymptotic CL for  $H \rightarrow ZZ \rightarrow 2l2q$  7TeV 4.9f-1

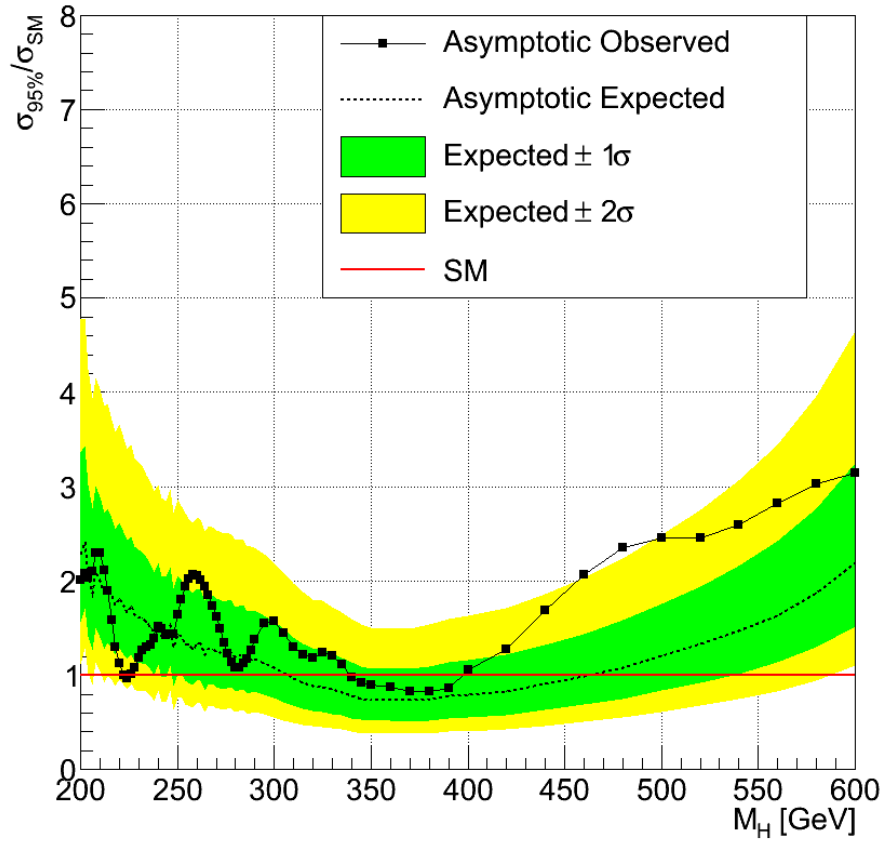


Figure 43: Limit on the expected 95% CL upper limit on the product of the Higgs boson production cross section and the branching fraction of  $H \rightarrow ZZ$  (black dots.), for the recorded luminosity in 2011 of  $4.9 \text{ fb}^{-1}$  at 7TeV. Yellow and Green bands represent the 68% and 95% ranges of expectation.

## References

- [1] CMS Collaboration 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 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] S. Bolognesi et al., "Search for the standard model Higgs Boson in the decay channel  $H \rightarrow ZZ \rightarrow 2\ell 2q$  at CMS", *CMS Note* **CMS-AN-2011/388** (2011).
- [4] CMS Collaboration, "Search for the Standard Model Higgs Boson in the Decay Channel  $H \rightarrow ZZ \rightarrow 2l2b$ ", CMS Analysis Note CMS-AN-2011-399, (2011).
- [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] CMS, "https://twiki.cern.ch/twiki/bin/viewauth/CMS/StandardModelCrossSectionsat8TeV", twiki, (2012).
- [9] CMS, "https://twiki.cern.ch/twiki/bin/view/CMS/HiggsZZ2l2q2012Skims", twiki, (2012).
- [10] 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).
- [11] CMS Collaboration, "Performance of CMS muon identification in pp collisions at  $\sqrt{s} = 7$  TeV", *CMS PAS* **MUO-2010-002** (2010).
- [12] CMS Collaboration, "Jet Performance in pp Collisions at  $\sqrt{s}=7$  TeV", CMS Physics Analysis Summary CMS-PAS-JME-10-003, (2010).
- [13] 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, (2010).
- [14] CMS, "Pileup Studies", twiki, (2012).
- [15] Michael Hildreth, "Pileup Scenarios", twiki, (2012).
- [16] CMS Collaboration, "Trigger strategies for Higgs searches in 2011", CMS Analysis Note CMS-AN-2011-065, (2011).



- [17] CMS, “<https://twiki.cern.ch/twiki/bin/view/CMS/EgammaCutBasedIdentification>”, twiki, (2012).
- [18] CMS, “<https://twiki.cern.ch/twiki/bin/view/CMSPublic/SWGuideMuonId>”, twiki, (2012).
- [19] 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.
- [20] CMS Collaboration, ““Jet Energy Corrections determination at 7 TeV””, CMS Physics Analysis Summary CMS-PAS-JME-10-010, (2010).
- [21] S. Bolognesi et al., “Search for a Semileptonic Decay of a SM Higgs or BSM Boson  $H \rightarrow ZZ \rightarrow 2l2j$ ”, CMS Note **CMS-AN-2011/100** (2011).
- [22] CMS Collaboration, “Commissioning of  $b$ -jet identification with pp collisions at  $\sqrt{s}=7$  TeV”, CMS Physics Analysis Summary **CMS-PAS-BTV-10-001** (2010).
- [23] 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.
- [24] 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.
- [25] CMS Collaboration, “Angular Analysis of Resonances  $pp \rightarrow X \rightarrow ZZ$ ”, CMS Analysis Note CMS-AN-2010-351, (2010).
- [26] CMS Collaboration, “CMS Luminosity Based on Pixel Cluster Counting - Summer 2012 Update”, CMS Physics Analysis Summary CMS-PAS-LUM-12-001, (2012).
- [27] CMS Collaboration, “Absolute jet energy correction uncertainty”, CMS Analysis Note CMS-AN-2010-304, (2010).
- [28] CMS Collaboration, ““Jet Energy Resolution in CMS at  $\sqrt{s} = 7$  TeV””, CMS Physics Analysis Summary CMS-PAS-JME-10-014, (2010).
- [29] Michael Hildreth, “Estimating Systematic Errors Due to Pileup Modeling”, twiki, (2012).
- [30] H.-L. Lai et al., “New parton distributions for collider physics”, *Phys.Rev.* **D82** (2010) 074024, doi:10.1103/PhysRevD.82.074024, arXiv:1007.2241.
- [31] 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.
- [32] 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.
- [33] C. Collaboration, “Discovery of a new boson in the search for the standard model Higgs boson in the HZZ2l channel in pp collisions at  $\sqrt{s}=7$  and 8 TeV”, CMS Note 2012/367, (2012).
- [34] S. Gorla, 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.

- [35] “LHC Higgs Cross Section Working Group web page”.
- [36] G. Passarino, “Higgs Interference Effects in  $gg \rightarrow ZZ$  and their Uncertainty”, *JHEP* **1208** (2012) 146, doi:10.1007/JHEP08(2012)146, arXiv:1206.3824.
- [37] C. H. C. Group, “Documentation of the RooStats-based statistics tools for Higgs PAG”, *CMS TWiki SWGuideHiggsAnalysisCombinedLimit* (2011).
- [38] G. J. Feldman and R. D. Cousins, “Unified approach to the classical statistical analysis of small signals”, *Phys. Rev D* **57** (1998) 3873–3889, doi:10.1103/PhysRevD.57.3873.
- [39] A. L. Read, “Modified Frequentist Analysis of Search Results (The CLs Method)”, *CERN OPEN* **2000-205** (2000).
- [40] R. D. Cousins and V. L. Highland, “Incorporating systematic uncertainties into an upper limit”, *Nucl. Instr. Meth A* **320** (1992) 331–335, doi:10.1016/0168-9002(92)90794-5.
- [41] L. Moneta et al., “The RooStats Project”, in *13<sup>th</sup> International Workshop on Advanced Computing and Analysis Techniques in Physics Research (ACAT2010)*. SISSA, 2010. arXiv:1009.1003. PoS(ACAT2010)057.
- [42] CMS Collaboration, “Search for a Higgs boson in the decay channel  $H \rightarrow ZZ(*) \rightarrow q \bar{q} l l$  in pp collisions at  $\sqrt{s} = 7$  TeV”, *JHEP* **1204** (2012) 036, doi:10.1007/JHEP04(2012)036, arXiv:1202.1416.
- [43] A. Hocker et al. arXiv:0703039.

## A Neural Network

As a cross check of the base line analysis optimization, we present in this Appendix a MVA optimization. The performance obtained by both methods is pretty similar giving extra confidence on our results based on the helicity likelihood discriminant. However the MVA techniques open the possibility to include new variables in addition to the helicity angles in a cleaner, faster and better-performing way for future iterations of this analysis.

### A.1 Signal Optimization Based on Helicity Neural Network

In order to check the effectiveness of the currently used helicity likelihood discriminant a parallel multivariate analysis has been performed. A Neural Network is applied to separate Higgs signal from background using the five helicity angles of the final objects in the analysis as inputs. A training and test evaluation has been performed with the framework of the TMVA package [43] using the real mixture of MC processes as background and Higgs MC. Both the training and testing samples for both background and signal are constructed and events randomly mixed outside of the Neural Network. There are 37,874 signal events and 38,378 background events evenly split between testing and training. The Neural Network was trained on the Monte Carlo generated for a hypothetical Higgs mass of  $400 \text{ GeV}/c^2$ . This training was on events that passed the previously explained preselection, the Z boson mass window cuts previously explained, but before the cut on MET significance.

#### A.1.1 Neural Network Architecture

The architecture of the Neural Network consists of two hidden layers with N and N neurons respectively, where N is the number of variables, and one output node as shown in Figure 44.

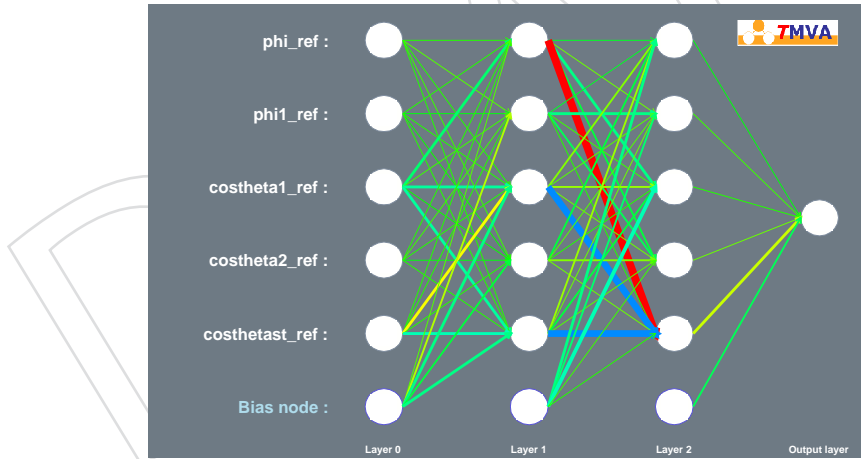


Figure 44: Neural Network architecture used for the training.

#### A.1.2 Input Variables and Training

The input variables for the Neural Network are the same variables that are used for the Helicity Likelihood Discriminant,  $\phi$ ,  $\phi_1$ ,  $\cos\theta_1$ ,  $\cos\theta_2$ , and  $\cos\theta_{\text{astar}}$ . The input variable distributions for both signal and background are shown in Figure 45. The default values are used when booking the Multi Layer Perceptron Neural Network except for the Hidden Layers as described in the previous subsection, and the number of cycles which is set to 1000. The background rejection versus signal efficiency for the input signal and background is shown in Figure 46.

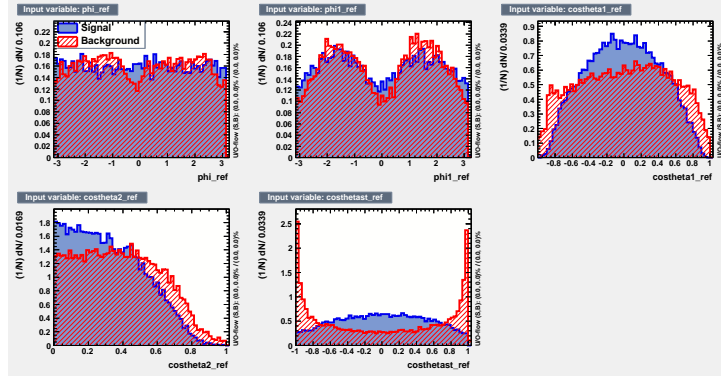


Figure 45: Signal and Background for the the Input Variables for the Neural Network training.

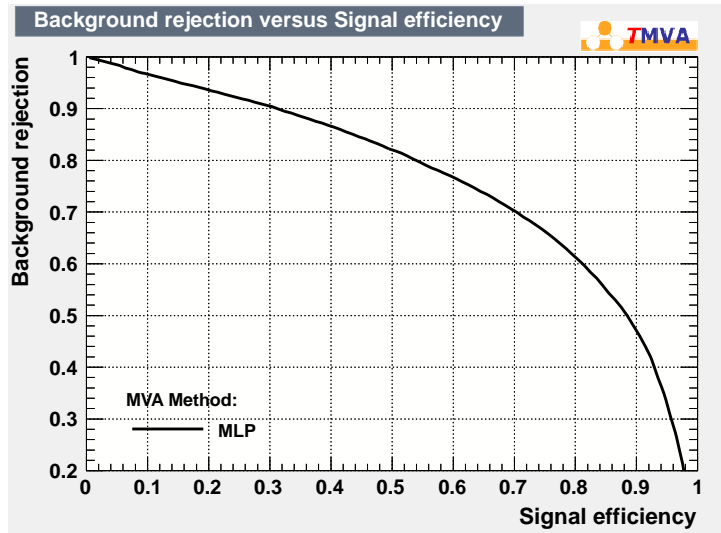


Figure 46: Background Rejection Versus Signal Efficiency for the input and output of the Neural Network.

### A.1.3 Results Based on Helicity Neural Network

The training and testing of the Neural Network can be seen in Figure 47. While there is separation between the signal and background, and good agreement between testing and training, there is a defined spike in background at the same place as the signal spikes in the MLP training, with a similar but less noticeable effect for the Likelihood training. The training was done with Monte Carlo generated for a hypothetical Higgs mass of 400 GeV but then applied to the Monte Carlo for hypothetical Higgs masses of 200, 300, 400, and 500 GeV. This is because the angular components should not depend on Higgs mass. In addition to the MLP neural network done each training and testing was additionally performed with a Likelihood. The performance of the two analysis is almost identical in all cases, especially when looking in tight hypothetical Higgs mass regions.

In the 1tag region after applying the additional MET significance cut the separation between signal and background looks similar to the training for Higgs 300, 400, and 500 GeV, but does not have much discrimination power for a Higgs of 200 GeV. The discrimination power is virtually the same after applying an additional cut of  $-6\%/+10\%$  of the Higgs mass for all

four cases as seen in Figure 48. When comparing background rejection versus signal efficiency between the Neural Network and Likelihood Discriminate the performance is for all practical purposes the same. See Figure 50.

In the 2tag region after applying the additional MET significance cut and a cut of  $-6\%/+10\%$  of the Higgs mass, the discriminating power of the Neural Network is similar to the 1tag case with poor ability for a Higgs of 200 GeV, but good separation for Higgs of 300, 400, and 500 GeV. This is shown in Figure 49. When comparing background rejection versus signal efficiency between the Neural Network and Likelihood Discriminate the performance is roughly the same, except for the Higgs of 200 GeV case where the Neural Network is almost consistently better than the Likelihood Discriminate. This is shown in Figure 51.

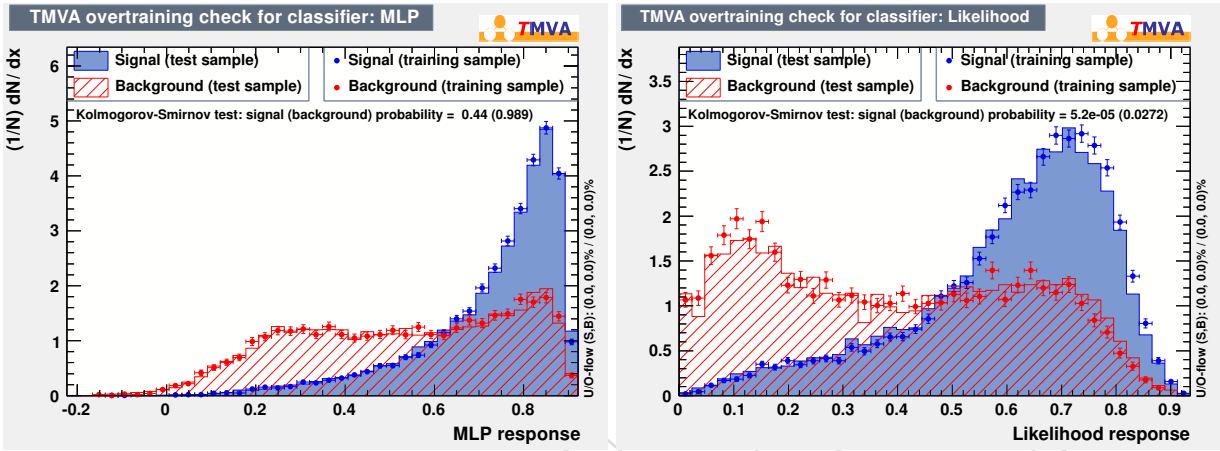


Figure 47: The trainings are done after preselection and additionally require at least one TCHEM jet (Track Counting High Efficiency  $> 1.9$ ) Left: Training 400 GeV Higgs boson with a MLP neural network. Right: Training 400 GeV Higgs boson with a Likelihood.

#### A.1.4 Potential Improvement to to MVA

A neural network or likelihood should be able to take advantage of extra information then simply using the 5 decay angles as previously shown. One such variable that offers good discrimination power in the preselection region is  $Z_{ll}pt / \sum pt$ , where  $\sum pt = l_0pt + l_1pt + j_0pt + j_1pt + met$ , and  $Z_{ll}pt$  and  $\sum pt$  are scalar quantities. This variable distribution is shown in Figure 53. While adding this variable to a MLP training improves the discrimination power in the inclusive pretag region once we look in a mass window of  $-6\%/+10\%$  around Higgs 400 adding  $Z_{ll}pt / \sum pt$  to the training actually lowers the discrimination power of the neural network. This same performance drop is seen when training is done with the addition of the reconstructed higgs mass as well.

While training on each Higgs mass can give additional discrimination power over training on one Higgs mass and applying it to multiple hypothetical Higgs masses there are some difficulties. It is more convenient to be able to train on just one higgs mass and then apply this training to a range of hypothetical higgs masses. An example of training on a Higgs 400 GeV sample and then applying this training to various hypothetical Higgs masses is show in Figure 54. This training is for a MLP neural network trained on the 5 decay angles. This training is applied after preselection and requiring at least one TCHEM jet. This shows that good performance can be achieved using the MLP by only training on one Higgs mass sample, without the need for additional trainings, or parametrizing a discriminate as a function of the reconstructed Higgs

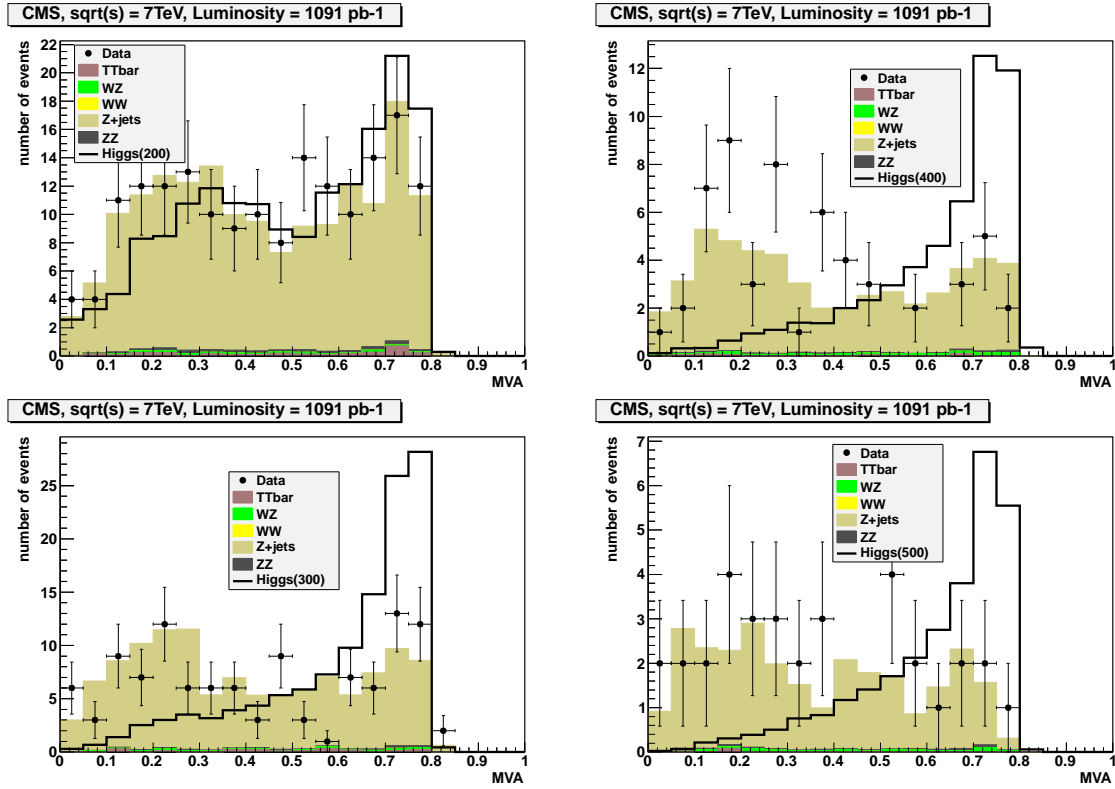


Figure 48: Signal and background Neural Network output in the 1tag region after preselection, the Z boson mass window cuts, cut on MET significance, and a  $-6\%/+10\%$  Higgs mass window. The signal is scaled to the sum of the Monte Carlo background.

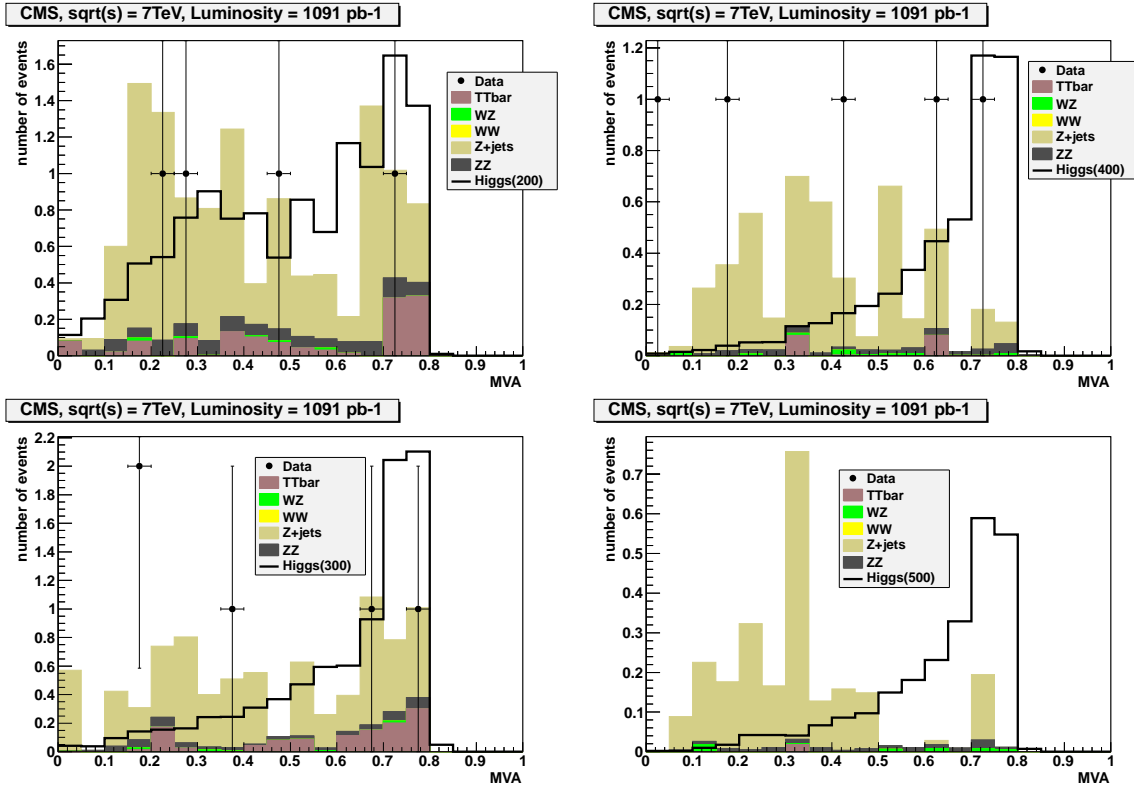


Figure 49: Signal and background Neural Network output in the 2 tag region after preselection, the Z boson mass window cuts, cut on MET significance, and a  $-6\%/+10\%$  Higgs mass window. The signal is scaled to the sum of the Monte Carlo background.

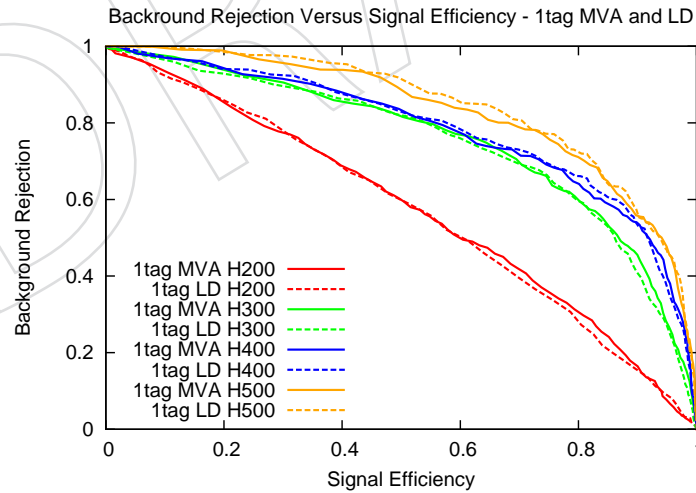


Figure 50: Background Rejection Versus Signal Efficiency in the 1tag region comparing the Multi Variant Analysis output to the the Helicity Likelihood Discriminant for a Higgs mass of 200, 300, 400, and 500 GeV. This is calculated after preselection cuts, Z boson mass cuts, cut on MET significance, in a  $-6\%/+10\%$  Higgs mass window.



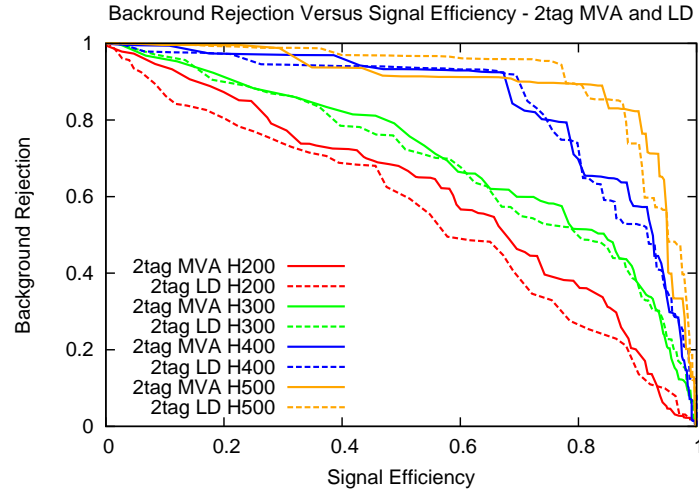


Figure 51: Background Rejection Versus Signal Efficiency in the 2tag region comparing the Multi Variant Analysis output to the the Helicity Likelihood Discriminant for a Higgs mass of 200, 300, 400, and 500 GeV. This is calculated after preselection cuts, Z boson mass cuts, cut on MET significance, in a  $-6\%/+10\%$  Higgs mass window.

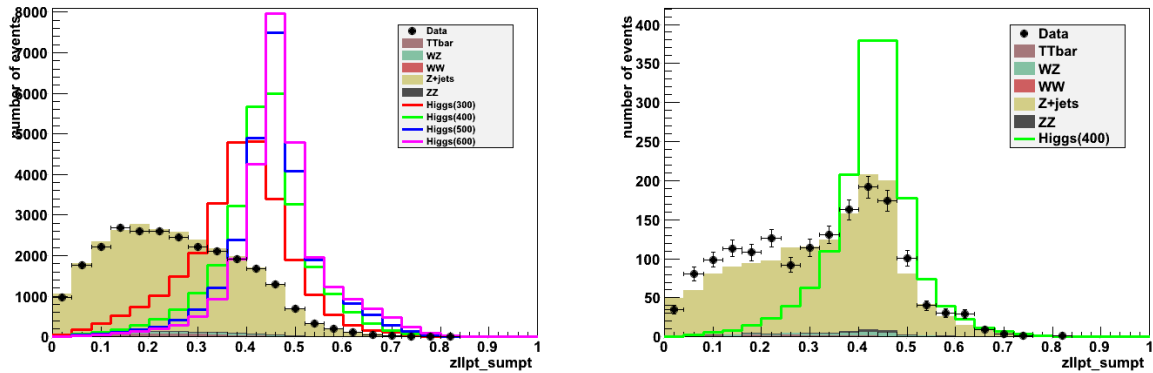


Figure 52: Signal samples are normalized to background. Left:  $\frac{Z_{ll}pt}{\sum pt}$  after preselection. Right:  $\frac{Z_{ll}pt}{\sum pt}$  after preselection and  $376 < m_{ZZ} < 440$  GeV.

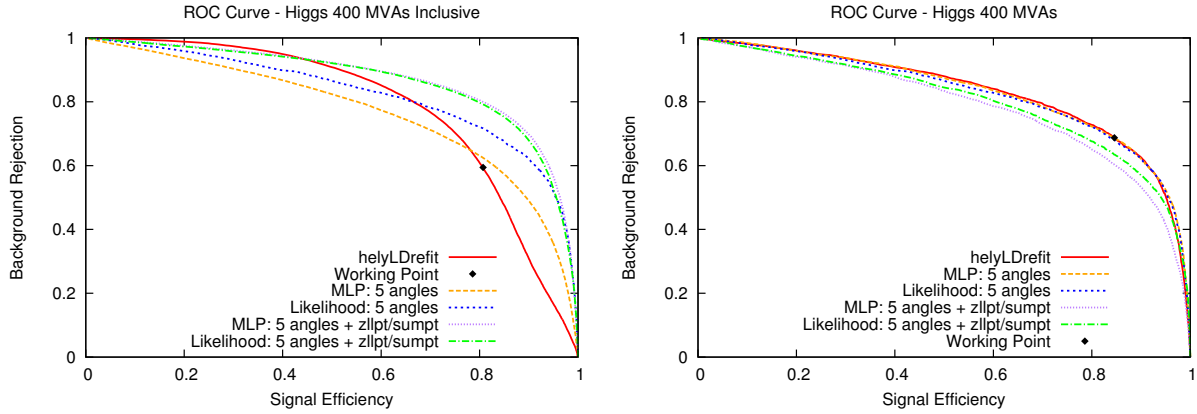


Figure 53: Applying a MLP training to preselection and at least one TCHEM jet. The working point is the equivalent performance of current analysis that we apply in the two tag region. For comparison the helyLDrefit variable is also shown. Left: ROC curves after preselection. Right: ROC curves after preselection and  $376 < m_{ZZ} < 440$  GeV.

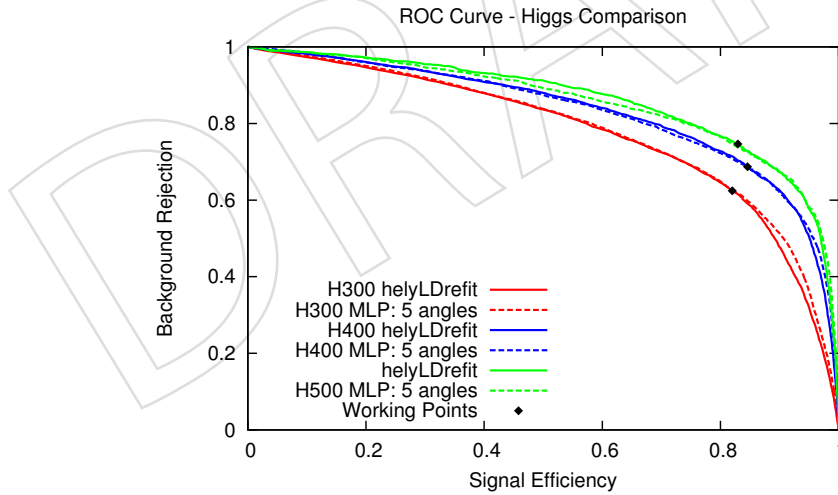


Figure 54: This training is for a MLP neural network trained on the 5 decay angles of a Higgs 400 GeV sample. This training is applied after preselection and requiring at least one TCHEM jet to samples with a Higgs mass of 300,400, and 500 GeV. The working point is the background rejection point that we currently achieve in the two tag region in our analysis. The helyLDrefit variable is shown for comparison.