CMS Analysis Note

The content of this note is intended for CMS internal use and distribution only

September 12, 2008

Search for the Higgs boson in the ZZ$^{(*)}$ decay channel with the CMS experiment

S. Baffioni [a], C. Botta [i], N. Chanon [g], C. Charlot [a], R. Clare [d], M. Costa [i], N. De Filippis [a], A. Drozdetskiy [b], C. Farrel [b], F. Ferri [c], D. Fortin [d], S. Gascon-Shotkin [g], N. Godinovic [k], A. Graziano [i], A. Korytov [b], K. Lelas [k], M. Lethuillier [g], C. Mariotti [i], G. Mitselmakher [b], M. Musich [i], A. Nikitenko [e], Y. Pakhotin [b], A. Pompili [h], I. Puljak [k], R. Salerno [c], Y. Sirois [a], D. Trocino [i], R. Wilken [d]

Abstract

A prospective search for the inclusive production of Standard Model Higgs bosons decaying in ZZ$^{(*)}$ pairs is presented with the CMS experiment at the CERN LHC pp collider. The analysis is performed for the leptonic decay channels $Z \to ll$ with $l = e, \mu$. Signal and background datasets obtained with a detailed Monte Carlo simulation of the detector response, including the limited inter-calibration and alignment precision expected at startup luminosities, are treated using a complete reconstruction chain. The data corresponding to an integrated luminosity of up to 1 fb^{-1} is analysed. If a Standard Model Higgs boson exists with a mass m_{H}, a signal evidence can be established with a significance above 2 standard deviations for some favourable m_{H} values. In absence of significant deviations from Standard Model background expectations, upper limits on the production cross-section of Standard Model-like Higgs bosons can be established and lie beyond existing constraints.

[a] LLR Ecole Polytechnique

[b] U. of Florida

[c] U. Milano Biccoca

[d] U.C. Riverside

[e] I.C. London

[g] IPN Lyon

[h] U. di Bari

[i] U. of Torino

[k] FESB Split

1 Introduction

The Standard Model (SM) of electroweak and strong interactions predicts the existence of a unique physical Higgs boson, the quanta of the scalar field responsible for electroweak symmetry breaking. The mass m_H of this scalar boson is a free parameter of the theory.

Direct searches for the SM Higgs particle at the LEP $\mathrm{e^+e^-}$ collider have lead to a lower mass bound of $m_\mathrm{H} > 114.4\,\mathrm{GeV/c^2}$ (95% CL) [1]. Ongoing direct searches at the TeVatron $\mathrm{p\bar{p}}$ collider by the D0 and CDF experiments set constraints on the production cross-section for a SM-like Higgs boson in mass range extending up to about $200\,\mathrm{GeV/c^2}$ [2]. A consistency fit including all the measured electroweak observables which are sensitive to the existence of a Higgs boson through virtual processes, favours the mass range $m_\mathrm{H} \lesssim 182\,\mathrm{GeV/c^2}$ (95% CL) [3].

The inclusive production of SM Higgs bosons followed by the decay $\mathrm{H} \to \mathrm{ZZ}^{(*)} \to l^\pm l^\mp l'^\pm l'^\mp$, with $l, l' = e$ or μ is expected to be a main discovery channel at the CERN LHC pp collider over a wide range of possible m_H values. Detailed prospective studies of the discovery potential in the $4l$ channels with the CMS experiment for nominal low collider luminosities of $2 \cdot 10^{33}\ \mathrm{cm^{-2}s^{-1}}$ have been performed previously [4, 5, 6, 7, 8]. In this paper, an analysis strategy is presented in the context of the startup luminosity at the LHC. Emphasis is put on the reduction of distinguishable background rates and on methods allowing a data-driven derivation of experimental and background systematic uncertainties. The expected sensitivity of the CMS experiment for the observation of a SM-like Higgs boson is studied using a sequential set of cuts, and for a m_H hypothesis in the mass range from $130\,\mathrm{GeV/c^2}$ to $250\,\mathrm{GeV/c^2}$.

A general description of the CMS detector can be found in Ref. [9]. This analysis relies mostly on the tracker [11] made of silicon pixel detectors and silicon strip detectors, the electromagnetic calorimeter (ECAL) [12] made of quasi-projective $\mathrm{PbW0_4}$ crystal, and the muon spectrometer [13] hosted in the iron magnet return yoke and consisting of Drift Tubes (DTs), Cathode Strip Chambers (CSCs), and Resistive Plate Chambers (RPCs). The CMS inner tracking and ECAL detectors are immersed in a 4 T magnetic field parallel to the z axis. The experiment is assumed to be operated in the trigger configuration foreseen for the LHC start-up luminosity of $\mathcal{L} = 10^{32}$ $\mathrm{cm^{-2}s^{-1}}$.

2 Physics Processes and Their Simulation

Signal and background datasets obtained with a detailed Monte Carlo simulation [14] of the detector response have been produced, taking into account the limited inter-calibration and alignment precision expected for an integrated luminosity of $100\mathrm{pb^{-1}}$. The data has been subject to full reconstruction [15].

The general multi-purpose Monte Carlo event generator PYTHIA [16] is used for the various signal and background processes described in details in the following, either to generate a given hard process at leading order (LO), or only for the showering and hadronization in cases where the hard processes are generated at next-to-leading order (NLO). All signal and background processes are re-weigthed to NLO.

PYTHIA incorporates Multi Parton Interaction (MPI) models to overlay underlying events due to additional soft interactions between the partons of the proton remnants. The so-called "DWT" tune available in PYTHIA 6.2 is used in all cases for the MPI with parameters adapted to the CTEQ5L parton density functions [17].

The main signal and background processes considered are:

- $\mathrm{H} \to \mathrm{ZZ}^{(*)} \to 4l$,
- $t\bar{t} \to 2\mathrm{W}b\bar{b}$,
- $\mathrm{Z}b\bar{b} \to 2lb\bar{b}$,
- $\mathrm{ZZ} \to 4l$.

Table 1 and sections 2.1 to 2.4 give more details on the production of the corresponding samples. Here and henceforward, Z stands for Z, Z^*, and γ^* (where possible). For the event generation, l is to be understood as being any charged lepton, e, μ or τ.

Additional background samples are used in this analysis and are contained in a dataset which was conceived to mimic (to a certain degree) a "real"' data stream from CMS in situ, the so-called SM soup. The soup mixes a large variety of Electroweak and QCD-induced SM processes, e.g. Z+jets and W+jets. It is used to measure the rate

Process	ME	$\sigma_{NLO} BR$
$\mathrm{H} \to \mathrm{ZZ} \to 4l$	PYTHIA	4-50 fb
$t\bar{t} \to 2Wbb$	ALPGEN/TopRex	840 pb
$\mathrm{Z}bb \to 2lbb$	CompHEP	573 pb
$\mathrm{ZZ} \to 4l$	CompHEP	1.2 pb

Table 1: Monte Carlo simulation datasets used for cut optimization; Z stands for Z, Z*, γ^*; l means e, μ, τ. The detector intercalibration and alignement precision expected for an integrated luminosity of $\mathcal{L} = 100\ \mathrm{pb}^{-1}$is considered for the simulation.

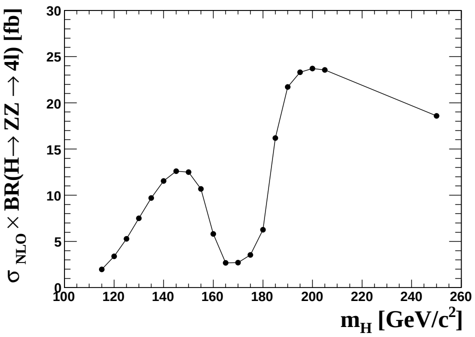


Figure 1: The NLO cross section for $\mathrm{H} \to 4l$ as a function of the mass m_{H} calculated as $\sigma_{NLO}(pp \to \mathrm{H}) \times BR(\mathrm{H} \to \mathrm{ZZ}) \times BR(\mathrm{Z} \to 2l)^2$, where l stands for e and μ.

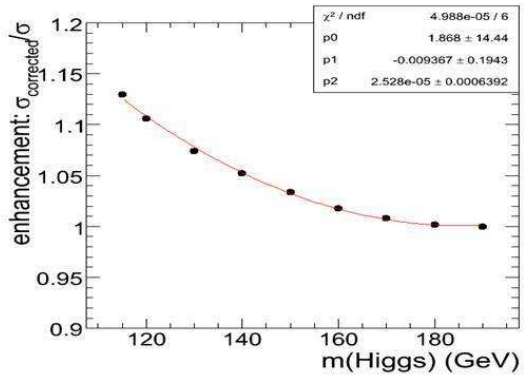


Figure 2: Enhancement of the Higgs production in the 4μ and $4e$ decay channels due to an interference of amplitudes with permutations of identical leptons originating from different Z-bosons.

of some of the "signal-like events" as well as to provide various "control samples". The control samples selected from the soup are used to derive directly from "data" a normalization for the main backgrounds, and to control efficiencies and systematics. The soup otherwise allow to verify that other plausible backgrounds involving "fake" primary leptons from QCD jets are suppressed to a negligible level in comparison to the three main backgrounds.

2.1 Signal: $\mathrm{H} \to \mathrm{ZZ}^{(*)} \to 4l$

The Higgs boson samples are generated with PYTHIA 6.225 [16] (LO gluon and weak-boson fusion, $gg \to \mathrm{H}$ and $q\bar{q} \to q\bar{q}\mathrm{H}$). The parton density function (PDF) set CTEQ5L with the QCD scale set at PYTHIA's default values is used. The Higgs boson is forced to decay to two Z-bosons, which are allowed to be off-shell, and both Z-bosons are forced to decay via $\mathrm{Z} \to 2l$ (where l stands for e, μ, and τ).

Events are then re-weighted to correspond to the total NLO cross-section $\sigma_{NLO}(pp \to \mathrm{H}) \cdot BR(\mathrm{H} \to \mathrm{ZZ}) \cdot BR(\mathrm{Z} \to 2l)^2$, where $\sigma(pp \to \mathrm{H})$ and $BR(\mathrm{H} \to \mathrm{ZZ})$ were taken from [18] and $BR(\mathrm{Z} \to 2l) = 0.101$ [19]. Figure 1 shows this $\mathrm{H} \to 4l$ cross section as a function of the Higgs boson mass m_{H}. In this analysis, a total 20 Monte Carlo samples corresponding to different Higgs boson masses: from 115-205 GeV/c^2 with a step of 5 GeV/c^2, and an additional mass point of 250 GeV/c^2, is used.

In comparison to $\sigma(pp \to \mathrm{H}) \cdot BR(\mathrm{H} \to \mathrm{ZZ}^{(*)}) \cdot BR(\mathrm{Z} \to 2\mu) \cdot BR(\mathrm{Z} \to 2e)$, the 4μ and $4e$ channel cross sections are enhanced due to an interference of amplitudes with permutations of identical leptons originating from different Z-bosons. PYTHIA cannot account for such an enhancement, but the correction can be calculated with CompHEP—see Fig. 2. This correction is taken into account in the 4μ and $4e$ analyses. The effect is not very large and completely vanishes as both Z-bosons go on-shell for large Higgs boson masses.

The analyses to be described below require, *in fine*, four generated leptons in the final states within the detector acceptance and with matching flavours and opposite signs: $\mu^+\mu^-\mu^+\mu^-$, $e^+e^-e^+e^-$, $\mu^+\mu^-e^+e^-$. Figure 3 show the efficiency for finding four such leptons with respect to the Higgs boson decays $\mathrm{H} \to 4l$. This calculation does not take into account events where one or both Z-bosons decay to $\tau\tau$ (although a few percent of such events will be reconstructed as four-lepton $\mu^+\mu^-\mu^+\mu^-$, $e^+e^-e^+e^-$, $\mu^+\mu^-e^+e^-$ final states and will contribute to the off-peak tail in the m_{4l}-distribution for a signal.

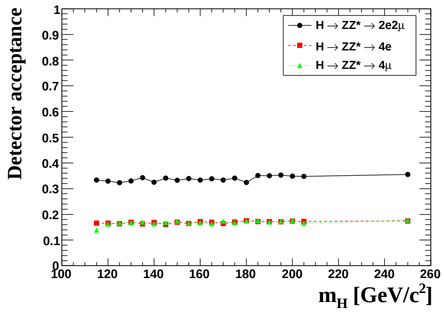


Figure 3: Efficiency to observe the final states $\mu^+\mu^-\mu^+\mu^-$, $e^+e^-e^+e^-$, and $\mu^+\mu^-e^+e^-$ within the CMS basic fiducial acceptance. The efficiency is calculated with respect to all Higgs boson decays $\mathrm{H} \rightarrow \mathrm{ZZ}$, and Z-bosons decaying to ee and $\mu\mu$.

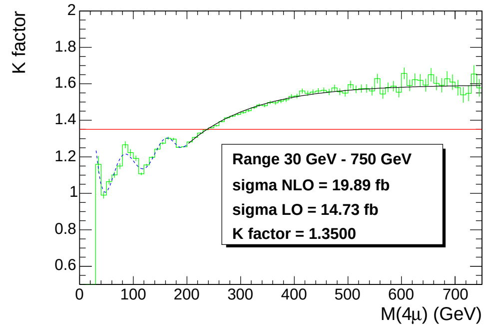


Figure 4: Mass-dependent Next-to-Leading-Order K-factor $K_{NLO}(m_{4l})$ for the ZZ $\rightarrow 4l$ process as evaluated with MCFM [24].

2.2 Background: $t\bar{t} \rightarrow 2Wb\bar{b}$

The $t\bar{t} \rightarrow 2Wb\bar{b}$ sample is generated with ALPGEN [20], as a part of the MC soup. A full inclusive sample $t\bar{t} + n$jets ($n = 0, 1, 2, 3, 4$) is used with statistics corresponding to the integrated luminosity of about 2 fb^{-1}. Parton showering and hadronization is done by PYTHIA 6.409. Number of events corresponds to the total NLO cross-section $\sigma(pp \rightarrow t\bar{t})\, BR(W \rightarrow l\nu)^2$, where $\sigma(pp \rightarrow t\bar{t}) = 840$ pb is taken from [21]. In order to increase statistics of MC events and to study systematics due to different MC generation models another sample is produced with TopRex [22] and preselected with following requirements: at least four leptons (electrons and/or muons) with $p_T > 2$ GeV/c within $|\eta| < 2.7$. Efficiency of this preselection on the inclusive sample is 8.6%.

2.3 Background: $\mathrm{Z}b\bar{b} \rightarrow 2lb\bar{b}$

The $\mathrm{Z}b\bar{b} \rightarrow 2lb\bar{b}$ samples are generated with CompHEP 4.2p1 [23] matrix element generator (PDF CTEQ5L, QCD scales $\mu_R = \mu_F = M_\mathrm{Z}$, b-quark mass $m_b = 4.85$ GeV/c^2, and a di-lepton mass cut $m_{ll} > 5$ GeV/c^2), interfaced to PYTHIA 6.225 for showering and hadronization. Included sub-processes are: $q\bar{q}/gg \rightarrow \mathrm{Z}b\bar{b} \rightarrow 2lb\bar{b}$, where q can be any of the light quarks, u, d, s, c. Initial states with b quarks were also considered at the generator level and found to be negligible. No restrictions on b-quark decays are applied. The corresponding CompHEP LO cross section is 345 pb. To obtain the NLO cross section, we calculated the NLO K-factor using MCFM [24]: $K_{NLO} = 1.66 \pm 0.03$. The conditions for the MCFM NLO and LO calculations were as follows: CTEQ6, $\mu_R^2 = \mu_F^2 = \hat{s}$, $m_b = 0$ GeV/c^2, $M(\mathrm{Z}^{res}) > 10$ GeV/c^2, $|\eta_l| < 3$,$p_T(l) > 2$ GeV/c, $p_T(jets) > 5$ GeV/c, $|\eta_{jets}| < 6$, $m_{b\bar{b}} > 9.24$ GeV/c^2. In this analysis preselected sample is used with same requirements as for the $t\bar{t}$ background. Preselection efficiency for this sample is 1.6%.

2.4 Background: $q\bar{q} \rightarrow \mathrm{ZZ} \rightarrow 4l$

This Monte Carlo sample is generated with CompHEP 4.2p1 matrix element generator (PDF CTEQ5L, QCD scales $\mu_R = \mu_F = \hat{s}$, and where the q-quark can be u, d, s, c or b). A cut $m_{ll} > 5$ GeV on the invariant mass of all possible pairs of same-flavor opposite-sign di-leptons is applied. Both t- and s-channel diagrams were included. The s-channel diagram, not available in PYTHIA, gives a large peak at $m_{4l} = m_\mathrm{Z}$. It contributes about 10% to events with $120 < m_{4l} < 180$ GeV and can be safely neglected for higher $4l$ invariant masses. The interference between t- and s-channels is found to be always negligible. More details on the relative role of the s-channel can be found elsewhere [25]. The CompHEP events are further interfaced to PYTHIA 6.225 for showering and hadronization.The CompHEP LO cross section is 846 fb.

To account for contributions to all the NLO diagrams and the NNLO gluon fusion process ($gg \rightarrow \mathrm{ZZ}$, known to contribute $\approx 20\%$ with respect to the LO [26]), events are re-weighted with a m_{4l}-dependent K-factor $K(m_{4l}) = K_{NLO}(m_{4l}) + 0.2$. The NLO K-factor $K_{NLO}(m_{4l})$, obtained with MCFM [24], is shown in Figure 4. The average correction is $< K >= 1.35 + 0.2 = 1.55$. All details on calculation of this m_{4l}-dependent K-factor and other dynamic differences between NLO and LO processes are summarized elsewhere [27].

3 Trigger selection

For Higgs boson masses m_{H} above 100 $\mathrm{GeV/c^2}$, the intermediate state in the cascade $\mathrm{H} \rightarrow \mathrm{ZZ}^{\star} \rightarrow 4l$ is expected to be dominantly produced with a least one Z boson on the mass shell, which then decay in pair of leptons carrying each a p_T of about $m_{\mathrm{Z}}/2$. The triggering of the CMS detector on the Higgs boson $4l$ signal relies on the presence of one or two high p_T leptons.

For the LHC start-up luminosity of $\mathcal{L} = 10^{32}\ cm^{-2}s^{-1}$, the High Level Trigger (HLT) configuration foreseen in CMS allows for single lepton p_T thresholds well below 20 GeV/c . Hence a very high selection efficiency is expected for the Higgs boson when the four final state leptons are within the triggering acceptance.

A global "OR" between different High Level Trigger sequences (*trigger-paths*) is chosen to maximize the signal detection efficiency. The triggers-paths taken into consideration are: single muon isolated, single muon non isolated, double muon non isolated, single electron isolated, single electron relaxed, double electron isolated, double electron relaxed and their combinations. The double muon isolated path is not used as it is essentially redundant with the corresponding non isolated path for what concerns signal selection. The expected trigger selection efficiency is determined for the purpose of this prospective analysis by applying the global "OR" of the HLT paths to Monte Carlo simulated event samples.

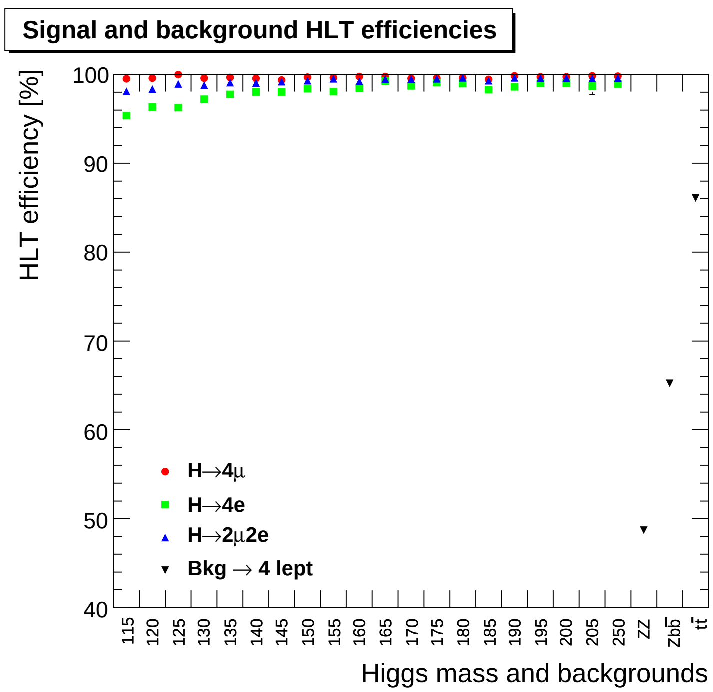


Figure 5: Expected trigger efficiencies for the three $H \rightarrow 4l$ channels and the main backgrounds for a logical OR of all the single and double lepton trigger paths of the CMS High Level Trigger at LHC start-up luminosities.

The HLT path efficiency for signal is defined as $\varepsilon = (\#\ L1 \wedge HLT\ events)/N_{gen}$ and the error is computed as $\delta\varepsilon = \sqrt{(\varepsilon(1-\varepsilon))/(N_{gen})}$, where N_{gen} is the number of generated events in the sample. In the denominator there is the number of events that have 4 leptons of the right flavour and charge ($2\mu^-2\mu^+$, $2e^-2e^+$ or $\mu^-\mu^+e^-e^+$) within the detector acceptance, i.e. $|\eta(\ell)| < 2.5$. The trigger efficiencies for the main backgrounds are evaluated starting from samples produced with the following requirements at generation level:

- $\mathrm{Z}/\gamma^*\mathrm{Z}/\gamma^*$: at least 4 leptons (muons or electrons);
- $(Z/\gamma^*)\mathrm{b\bar{b}}$: at least 4 leptons (muons or electrons) with $p_T > 2$ GeV/c and $|\eta| < 2.7$ (filter efficiency: 1.6%);
- $\mathrm{t\bar{t}}$: at least 4 leptons (muons or electrons) with $p_T > 2$ GeV/c and $|\eta| < 2.7$ (filter efficiency: 31.7%).

The trigger efficiencies are reported in Fig. 5 for the $H \rightarrow 4\mu$, $H \rightarrow 4e$ and $H \rightarrow 2\mu 2e$ channels and for several hypothetical values of the SM Higgs boson mass, as well as for the main backgrounds, $\mathrm{Z}/\gamma^*\mathrm{Z}/\gamma^*$, $(Z/\gamma^*)\mathrm{b\bar{b}}$ and

$\mathrm{t\bar{t}}$. A very high trigger efficiency is obtained for the signal, withy values above 95% and close to 100% for Higgs masses higher than 200 $\mathrm{GeV/c^2}$. In particular, muon triggers show better performances, so the channels with muons have higher efficiencies. While significant overlaps occur between different trigger paths, all the chosen triggers contribute to ensure the highest possible efficiency for the signal.

The rate of main background events which pass the HLT selection can be estimated to be of about $25k$ $\mathrm{t\bar{t}}/\mathrm{fb}^{-1}$+ $540k$ $\mathrm{Zb\bar{b}}/\mathrm{fb}^{-1}$+ 670 $\mathrm{ZZ}^{(*)}/\mathrm{fb}^{-1}$= $566k$ evts/fb^{-1}.

4 Lepton Reconstruction, Identification and Isolation

The reconstruction of the SM Higgs boson in the decay chain $H \rightarrow \mathrm{ZZ}^* \rightarrow 4\mathrm{l}$ imposes high performance lepton reconstruction, identification and isolation as well as excellent lepton energy-momentum measurements. The identification of isolated leptons emerging from the event primary vertex allows for a drastic reduction of QCD-induced sources of misidentified ("fake") leptons. The precision energy-momentum measurements translates in a precision Higgs boson mass measurement. The mass m_H is the single most discriminating observable for the Higgs boson search.

With four leptons in the final state, and in view of the modest fraction of the total production cross-section observable in the $4l$ channels, a very high lepton reconstruction efficiency is mandatory. This turns out to be especially challenging for the reconstruction of leptons at very low p_T^{l}. For Higgs bosons with masses $m_{\mathrm{H}} \lesssim 2m_Z$, one lepton pair at least couples to an off-shell Z^* boson. The softest lepton in that pair typically has $p_T^{\mathrm{l}} \lesssim 10$ GeV/c for masses $m_{\mathrm{H}} < 140$ GeV/c^2. Such very low p_T^{l} values lie at the extreme edge of the domain which will be controlled at the LHC using tag-and-probe methods in inclusive single m_{Z} production. In the low p_T range, a full combination of information provided by the tracker and electromagnetic calorimetry (for electrons) or by the tracker and muon spectrometer (for muons) becomes essential for the reconstruction, identification and isolation of leptons.

The electron measurements are hampered by effects resulting from the presence of a strong magnetic field and the amount of CMS tracker material. Electrons traversing the silicon layers of the tracker radiate bremsstrahlung photons and the energy reaches the ECAL with a significant spread in the azimuthal direction ϕ. The radiated photons can further convert and lead to more complicated hit patterns in the tracker, resulting possibly in a non negligible charge mis-identification probability even for isolated electrons, and inducing non-Gaussian contributions to the event-by-event fluctuations in the calorimeter energy and track momentum measurements. With four electrons in the $H \rightarrow 4e$ final state, each signal event is very likely to have some electron observables affected by electromagnetic "showering" patterns in the tracker volume, with electron reconstructed objects involving several electromagnetic clusters and several track segments. The electron reconstruction starts from electromagnetic clusters in the ECAL. The algorithm combining ECAL and tracker information used in this analysis is described in section 4.1.3.

The muon pattern recognition starts at the level off individual "outer" detectors, with track segments built in the DTs and CSCs, and 3D hit points obtained in the RPCs. The information from each of the muon sub-detector is then combined to form track candidates which are propagated back to the "inner" silicon tracker volume, and then matched and combined with inner tracks. The muon reconstruction algorithm used in this analysis is described in section 4.2.1.

The expected reconstruction performance and strategies for the control of associated uncertainties from data are presented in sections 4.1.4 and 4.2.2.

4.1 Electron reconstruction, identification and isolation

4.1.1 Electron Reconstruction

The standard collection of the "pixelMatchGsfElectrons" from the CMSSW reconstruction version 1.6.12 is used [15]. This incorporates improvements in the detailed steering of the reconstruction algorithm with respect to those used in the CMS Physics Technical Design Report [39], but the essential strategy described in [29] is maintained. The algorithm is briefly described here for completeness.

The reconstruction starts with the reconstruction of superclusters in the ECAL. A threshold of 1 GeV/c is used to initiate cluster building together with an extended road in ϕ (with respect to that used in the algorithm for the HLT) for a better collection of bremsstrahlung. The ECAL supercluster is used to drive the seeding of electron tracks in the silicon detector, with hits positions in the pixel layers predicted by the propagation of the energy weighted mean position of the supercluster backward through the magnetic field under both charge hypotheses. The requirements for the search of the first and second pixel hits are loosened with respect to those of the HLT. Starting from the seed, a trajectory is created. The track building relies on the Bethe-Heitler modelling of the electron energy losses and a loose χ^2 cut is used to efficiently collect tracker hits up to the ECAL front face. A Gaussian Sum Filter (GSF) is applied for the forward and backward fits. The track momentum is taken from the most probable value of the mixture of the Gaussian distributions available for each hit position. This procedure allows to efficiently build electron tracks while maintaining good momentum resolution. The relative difference between the momenta measured at both track ends, $f_{\rm brem} = (p_{\rm in} - p_{\rm out})/p_{\rm in}$ is a measure of the fraction of the electron initial energy emitted via bremsstrahlung in the tracker.

The fitted electron track together with the supercluster used to initiate the track are then associated to form an "electron object" if the following requirements are satisfied:

- $0.35 < E_{\rm sc}/p_{\rm in} < 3\ (5\ \text{in the endcaps})$, where $E_{\rm sc}$ is the supercluster energy and $p_{\rm in}$ the track momentum at the innermost track position,
- $|\Delta\eta_{\rm in}| = |\eta_{\rm sc} - \eta_{\rm in}^{\rm extrap.}| < 0.02$, where $\eta_{\rm sc}$ is the energy weighted position in η of the supercluster and $\eta_{\rm in}^{\rm extrap.}$ is the position in η extrapolated to the ECAL,
- $|\Delta\phi_{\rm in}| = |\phi_{\rm sc} - \phi_{\rm in}^{\rm extrap.}| < 0.1$, where $\Delta\phi_{\rm in}$ is a similar quantity in azimuthal coordinates,
- $H/E < 0.2$, where H is the energy deposited in the HCAL towers in a cone of radius $\Delta R = 0.1$ centered on the electromagnetic super cluster position and E the energy of the electromagnetic supercluster,
- $p_T > 5$ GeV/c, where p_T is the transverse momentum of the track at the innermost position.

These requirements are designed to maintain high efficiency at the reconstruction stage while keeping the purity of the reconstructed object at a reasonable level.

Ambiguities in the cases where several GSF tracks are associated to the same supercluster or where two different superclusters are associated with the same GSF track, and which occur with a probability of 1-2% for electrons, are resolved by chosing the best $E_{\rm sc}/p_{\rm in}$ matching.

The electron reconstruction efficiency [1)] is shown in Fig. 6 as a function of $p_T^{\rm e}$ and $\eta^{\rm e}$ for the electrons from Higgs boson events at $m_{\rm H} = 150\,{\rm GeV}/c^2$. The efficiency steeply rises and reaches a plateau around 86% for $p_T^{\rm e} \gtrsim 20$ GeV/c. The efficiency is 90% for $|\eta^{\rm e}| \lesssim 1.1$ and decreases towards the edge of the tracker acceptance when approaching $|\eta^{\rm e}| \simeq 2.5$. At $|\eta^{\rm e}| > 2$, the efficiency loss is mainly due to the limited coverage of the forward pixel disks.

After the preselection of the electron candidates, a non negligible charge mis-identification probability per candidate of about 1.3% is observed, for electrons from 150 GeV/c^2 Higgs boson signal. This is caused by the presence of bremsstrahlung photon conversion close to the primary electron track and in the innermost layers of the silicon tracker. Studies are ongoing to improve the charge determination by using, for this, only the part of the electron track closest to the interaction vertex.

The quantity $f_{\rm brem}$ together with other observables sensitive to the amount of bremsstrahlung radiated along the electron trajectory and to the pattern of photon emission and conversions, are used to classify electrons. Class-dependent electron energy measurement corrections are applied. The initial electron momentum direction is taken from the track. Depending on the class and on the E/p range, the ECAL corrected super cluster energy or the tracker momentum or the weighted mean of both measurements is used to estimate the initial electron momentum. Weights are evaluated using class-dependant errors available for each reconstructed electron.

The purity is then increased by applying an electron identification algorithm on electron candidates coming out of the reconstruction step.

---

1) The electron reconstruction efficiency has been further improved in recent developments made available in version 2 of the CMSSW software and which could not be used in this analysis. The potential gain in detection efficiency was estimated in a private production for Higgs boson signal events. The efficiency gain is obtained by exploiting the tracker layers in addition to the pixel layers for the electron track seeding algorithm, by relaxing the $E_{\rm sc}/p_{\rm in}$ cut, and by replacing the p_T (tracker-based) cut by an E_T (calorimeter-based) cut. Integrated over the full pseudorapidity range, an efficiency gain of $\sim$ 10% on the $4e$ reconstruction efficiency is obtained at $m_{\rm H} = 150$ GeV/c^2.

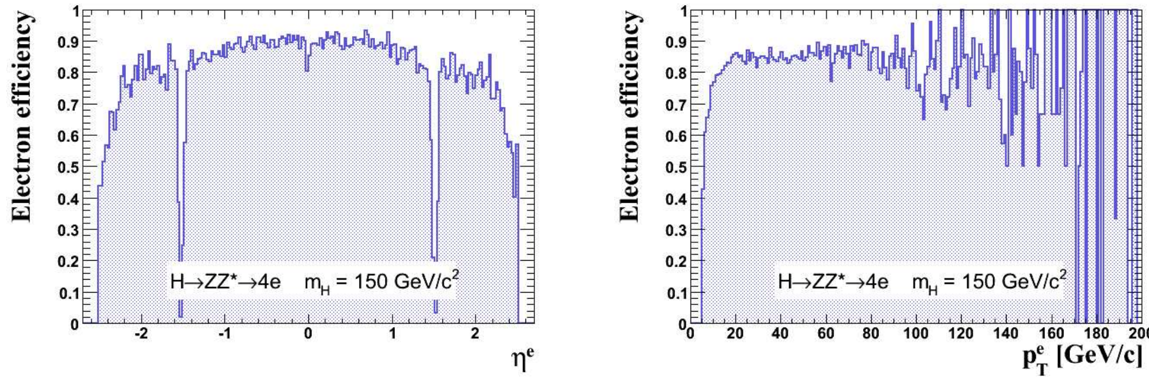


Figure 6: Electron reconstruction and pre-selection efficiency for $m_{\rm H} = 150$ GeV$/c^2$: a) versus $p_T^{\rm e}$; b) versus $\eta^{\rm e}$.

4.1.2 Electron Identification

The final selection of electron candidates is performed using a cut-based approach with tighter requirements on electron identification observables. A standard CMSSW algorithm is used with cuts specifically tuned for the purpose of this analysis where the very highest electron reconstruction efficiency must be preserved, that is compatible with a reduction (through the full analysis chain) of the background sources involving ”fake” primary electron.

The algorithm makes use of matching observables involving track parameters at both the outermost and the innermost track positions as well as shower shape observables. The classification based on f_{brem} and on the observed super cluster pattern in the ECAL is used to apply different set of cuts for the different electron topologies.

The following observables are used in the electron identification algorithm:

- $|\Delta\eta_{\rm in}|$,
- $|\Delta\phi_{\rm in}|$,
- H/E,
- Σ_9/Σ_{25}, where $\Sigma_{9(25)}$ is the sum of the 9(25) crystal energies centered on the hotest crystal of the seed cluster,
- $E_{\rm seed}/p_{\rm out}$, where $E_{\rm sc}$ is the seed cluster energy and $p_{\rm out}$ the track momentum at the outermost track position.

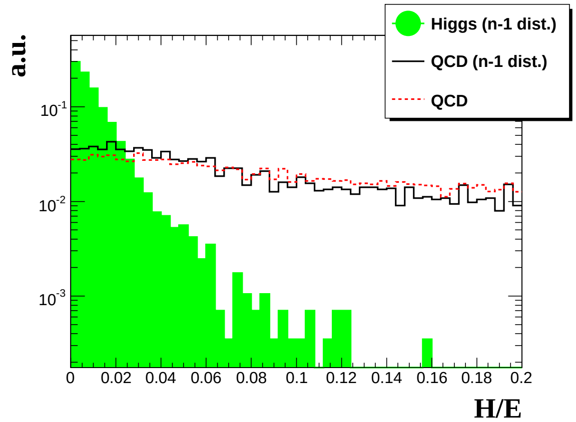
The Figure 7 presents the distributions of the electron identification observables H/E and of $E_{\rm seed}/p_{\rm out}$ for electrons from the Higgs signal and fakes from the QCD background, after cuts on all other electron identification variables have been applied (so called ’N-1 distributions’). The H/E distribution is shown for electrons in the golden class while the $E_{\rm seed}/p_{\rm out}$ is shown for electrons in the showering class. The discriminating power of these variables is clear. The threshold effect visible at 0.5 in the $E_{\rm seed}/p_{\rm out}$ distribution for QCD events is due to the pixel match filtering.

The set of cut values for electron identification used in this analysis are listed in Table 2.

4.1.3 Electron isolation

Track based isolation.

An isolation cone of size $R_{\rm cone} = \sqrt{\Delta\eta^2 + \Delta\phi^2}$ is defined around the electron direction. The tracks falling inside a narrower signal ’internal cone’ with opening $R^{int}_{\rm cone} < R_{\rm cone}$ around the electron direction, as well as the candidate electron track itself are discarded for the calculation of the isolation observable. Within the isolation cone, the reconstructed tracks (excluding any pre-selected electron candidates as defined in section 4.1.1) satisfying the quality requirements listed in Table 3 and originating from the same vertex as the candidate electron are considered. Whether or not a track actually enters in the calculation of the isolation observable is controlled by a lower



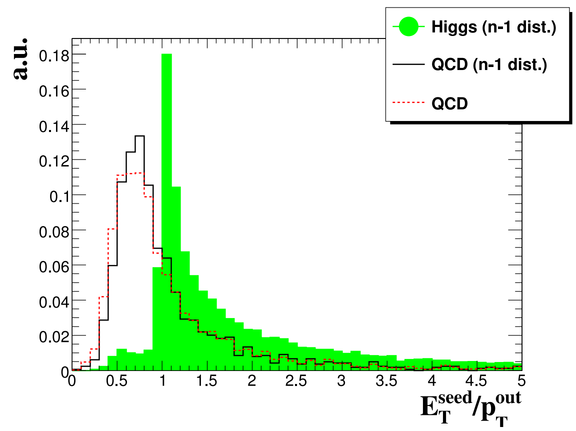


Figure 7: N-1 distributions (see text) of electron identification variables for the Higgs signal (plain-green histograms) and fakes from QCD background (solid-black lines): H/E for the golden class (top) and $E_{\rm seed}/p_{\rm out}$ for the showering class (bottom). Distributions without any electron identification cut are also shown for the background (dashed-red line).

p_T threshold and by a parameter defining the agreement in longitudinal impact parameter (LIP) with the electron candidate.

The isolation requirement is imposed on the transverse momentum sum of the considered tracks divided by the electron transverse momentum, $\sum p_T^{\rm tracks}/p_T^{\rm e}$, which is found as an optimal variable across the transversal momenta spectrum of electrons from Higgs boson decay for a wide range of Higgs boson masses.

All parameters of the isolation, size of internal and external cones, LIP compatibility with primary vertex and the minimal p_T track threshold, have been optimized and the following values are found as optimal: $R^{ext}_{\rm cone} = 0.25$, $R^{int}_{\rm cone} = 0.015$, $\Delta_{LIP} = 0.2$ and $p_T^{thr} = 1$ GeV/c. This values gives a best $\rm t\bar{t}$ and $Zb\bar{b}$ rejection for signal efficiency between 90 and 95%.

Calorimeter based isolation.

Isolated electrons are expected to have their superclusters in the ECAL surrounded by negligible additional deposits of energy, and be accompanied by negligible energy deposits in the HCAL. The usage of the ECAL for electron isolation requires special care given the presence of internal bremsstrahlung (from fermions involved in the interaction process) and external bremsstrahlung (from final state electrons traversing the tracker material) and the high probability of secondary photon conversion within the tracker volume.

The QCD and Z+jet(s) backgrounds can be largely suppressed at pre-selection without the help of the ECAL information for isolation as will be seen in section 5. The baseline selection strategy deployed in this analysis and presented in section 8 further succeeds in suppressing all remaining distinguishable background sources, while maintaining a very high signal detection efficiency, without the help of the ECAL for isolation purposes. One advantage is that the question of the eventual usage of internal bremsstrahlung recovery algorithms can be postponed

barrel	*golden*	*bigbrem*	*narrow*	*showering*
$H/E\ <$	0.06	0.06	0.07	0.08
$\mid \Delta\eta_{\rm in}\mid\ <$	0.005	0.008	0.008	0.009
$\mid \Delta\phi_{\rm in}\mid\ <$	0.02	0.06	0.06	0.08
$E_{\rm seed}/p_{\rm out}\ >$	0.7-2.5	1.7	0.9-2.2	0.6
$\sum_9 / \sum_{25}\ >$	0.8	0.7	0.7	0.5
endcap	*golden*	*bigbrem*	*narrow*	*showering*
$H/E\ <$	0.06	0.06	0.07	0.08
$\mid \Delta\eta_{\rm in}\mid\ <$	0.005	0.008	0.008	0.009
$\mid \Delta\phi_{\rm in}\mid\ <$	0.02	0.06	0.06	0.08
$E_{\rm seed}/p_{\rm out}\ >$	0.7-2.5	1.7	0.9-2.2	0.6
$\sum_9 / \sum_{25}\ >$	0.8	0.8	0.8	0.5

Table 2: Value of cuts on electron identification variables.

n. of hits	$>= 10$	[8, 9]	[5, 7]
$p_T(GeV)$	> 1	> 1	> 1
$d_0(cm)$	< 1	< 0.2	< 0.04
$d_z(cm)$	< 5	< 2	< 0.5
d_0/σ_{d_0}	–	< 10	< 7
d_z/σ_{d_z}	–	< 10	< 7

Table 3: The quality track requirements for tracks considered in electron and muon isolation algorithms.

to later stages (i.e. beyond pre-selection) of the analysis. Further studies for the usage of ECAL isolation in the $\mathrm{H} \rightarrow 4\ell$ analysis are ongoing [2)].

The isolation using information from the HCAL is found to be a powerful tool in complement to tracker-based isolation for the suppression of backgrounds involving "fake" primary electrons from QCD jets. Within the chosen track isolation cone of size $R_{\rm cone} = \sqrt{\Delta\eta^2 + \Delta\phi^2} = 0.25$, all hadronic calorimeter towers satisfying $E_T >$ 0.5 GeV are considered. The hadronic isolation requirement is imposed on the transverse momentum sum of the considered HCAL towers divided by the electron transverse momentum.

Both track based and HCAL based isolations are combined in a single isolation variable for final analysis. Performance of this combination is described in Section 6.2.1.

4.1.4 Electron measurements uncertainties

A possibly important source of systematic error on electron reconstruction and identification comes from the limited knowledge of the material budget in the tracker. A change in the integral amount of the tracker material traversed by electron tracks before reaching the ECAL would affect the electron selection and the identification observables, as well as the energy scale. An uncertainty on the material budget knowledge would also affect acceptance calculation from the Monte Carlo. The possibility to use f_{brem} to control the amount of material budget from data has been studied in [6] where it as shown that one can extract from this variable an estimate of X/X_0 linearly correlated with the true X/X_0. With a luminosity of 1 fb^{-1}a precision on the amount of material budget of the order of 6% per η bin of 0.1 unit can be reached.

The tag-and-probe method to measure the electron reconstruction efficiency from data has been discussed already in several references [6, 30, 31], and is being further studied in the context of the CSA08 exercise. It relies upon $Z \rightarrow e^+e^-$ decays to provide an unbiased and high purity electron sample from which one can measure the efficiency of a particular selection step or cut using the Z mass constraint. One of the electrons from the Z decay, the *tag* is required to satisfy stringent identification requirements, while the other electron, the *probe*, constrained to combine in an invariant mass close to the Z mass, is used to evaluate the efficiency of a given selection or cut.

The efficiencies for electron reconstruction, identification and isolation measured using the tag-and-probe method are shown on Fig. 8 as a function of η and E_T^{SC}, together with the statistical errors corresponding to an integrated luminosity of 100 pb^{-1}. The sample used is the ALPGEN "showder" soup including 10 pb^{-1} misalignement

[2)] New isolation algorithms exploiting the ECAL information have been made available recently in version 2 of the CMSSW software and could not be used in this analysis.

and miscalibration conditions. The *tag* is defined as a *golden* electron passing electron identification and isolation requirements. The tag and probe invariant mass is required to be within a 10 GeV window around the Z mass.

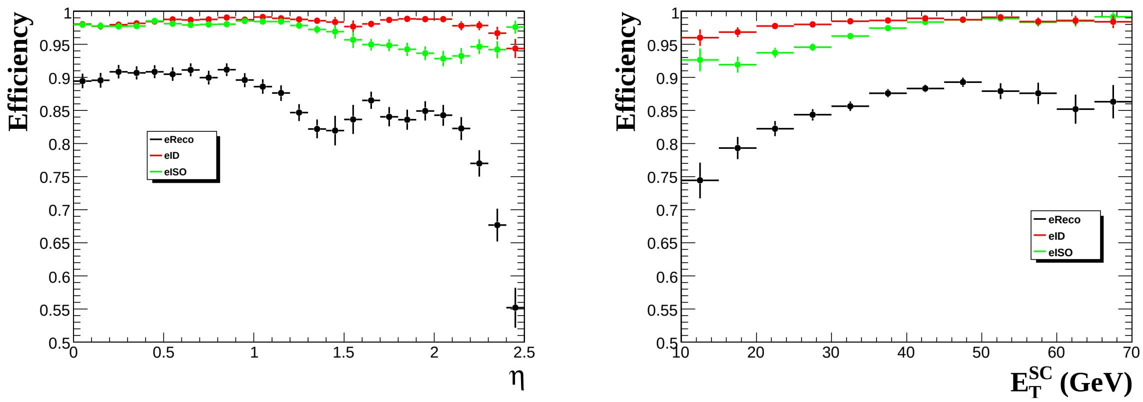


Figure 8: Electron reconstruction, identification and isolation efficiency as measured with the tag and probe method for an integrated luminosity of 100 pb^{-1}: a) versus super cluster η; b) versus super cluster E_T.

The systematic uncertainty associated with the determination of the electron reconstruction and identification are therefore determined by the statistic available for a given luminosity, and increases moving away in p_T from the jacobian peak, as shown on Fig. 9.

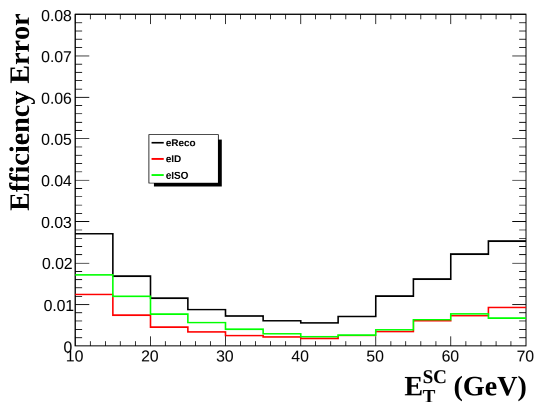


Figure 9: Electron reconstruction, identification and isolation efficiency error versus super cluster E_T as measured with the tag and probe method for an integrated luminosity of 100 pb^{-1}.

The efficiency of reconstructing and identifying an electron can therefore be measured with a systematic uncertainty of less than 2.5% for an integrated luminosity of 100 pb^{-1} with misalignment and miscalibration conditions from 10 pb^{-1}.

Another potential source of systematic uncertainty on the electron measurement comes from the determination of the energy scale. It has been shown [32] that the tag-and-probe method can be used to estimate the associated systematic error. It is expected to be of the order of 1% for an integrated luminosity of 100 pb^{-1} in the ECAL barrel and about 2% in the ECAL endcaps. Similarly, the control and fine tuning of the classification that enters in the determination of electron momentum corrections and in the electron identification can be performed using $Z \to e^+e^-$ early data as was shown in [6]. .

Finally, the limited knowledge of the exact position and orientation of the silicon sensors in the tracker in the first period of data taking will affect the performances of the electron reconstruction. Effect of misalignment on the tracking performances has been studied in details and some results are reported in the muon section. The effect of misalignment is taken into account during the track reconstruction by increasing the alignment position errors, to

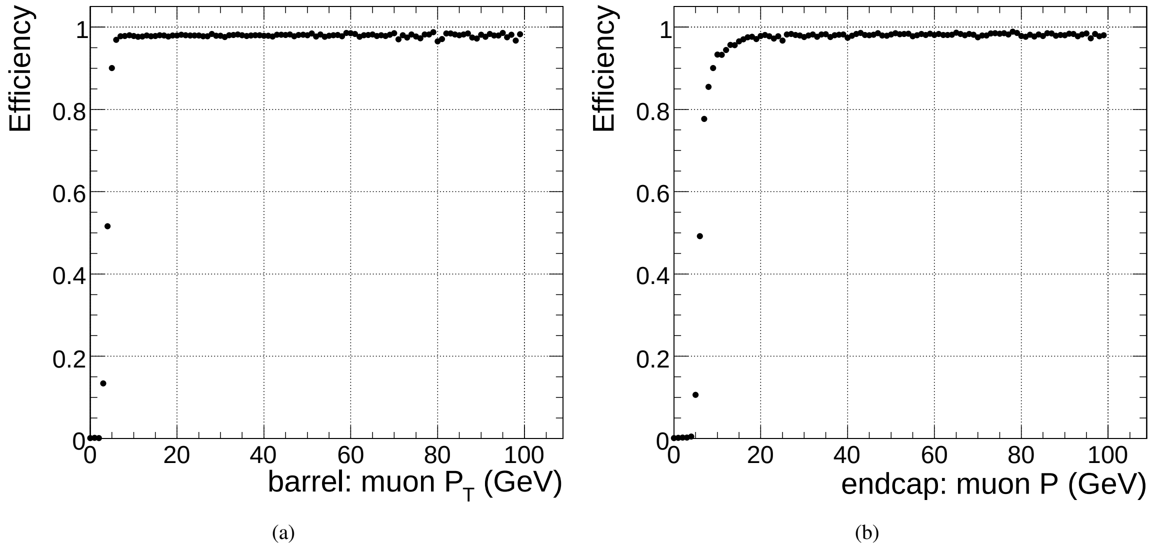


Figure 10: The muon reconstruction efficiency as a function of p_T in barrel (left) and p in endcap (right).

the price of an increased fake rate. In general, the performances obtained in the 100 pb^{-1} scenario are very close to the ideal case. In the case of electron tracking, the seeding stage which dominates the reconstruction efficiency relies in particular on the beam spot position in the transverse plane. Ongoing studies with early data scenario in the context of CSA08 indicates a significant loss of efficiency with very first data (1 pb^{-1} conditions) due to a wrong position of the beam spot as obtained from the misaligned tracker. After a first alignment is applied (10 pb^{-1} conditions) the efficiency loss due to the bad initial knowledge of the beam spot position is essentially recovered. The effect of tracker misalignment on the electron track parameters and therefore on the selection efficiency in a 100-1000 pb^{-1} scenario is expected to be small, although residuals are to be expected in particular in the charge mis-identification rate and in the primary vertex determination. As compared to muon tracks, it is expected that these effects are less important due to the larger uncertainties already introduced in the GSF track fit to account for bremsstrahlung energy loss.

4.2 Muon reconstruction, identification and isolation

The description of the CMS Muon System, of the muon reconstruction and identification and their performace can be found in ref. [10, 33]. For this analysis the so called "global muons" have been used, presenting an efficiency above 97% over most of the η region and for $p_T^{\mu} \gtrsim 5\mathrm{GeV}/c$. In figure 10 the efficiency as a function of the muon p and p_T is shown; in this analysis muons with $p_T > 5$ GeV/c have been used corresponding to an efficiency of 90%. In the region defined by $1.1 < \eta < 2.4$ muons are considered only if their momentum p is larger than 9 GeV/c.

4.2.1 Muon isolation

An isolation cone is drawn around the track of all the reconstructed muon. An inner veto cone is also defined around the muon, in order to subtract the deposits of the muon itself from the overall amount. The radius of a veto cone is defined in the η - ϕ space as $\Delta R = \sqrt{\Delta\eta^2 + \Delta\phi^2}$. The sizes of both the outer and the inner veto isolation cones have been optimized during the analysis; the best values have been found to be $\Delta R_{outer} = 0.3$, $\Delta R_{veto} = 0.015$.

If two or more muons fall in the same isolation cone, the contributions of extra muons must be subtracted. Similarly, the effects of muon bremsstrahlung are properly taken into account and corrections are applied.

Thresholds on the p_T, on the radial and longitudinal impact parameters of the tracks and on the corresponding significancies are also been optimized. Their goal is to suppress ghost tracks and they depend on the number of hits per track: the more the hits, the more the track is likely to be a real one. Impact parameter selections are very effective especially against $Zb\bar{b}$ background. The efficiency of these cuts has been estimated to be at the level of 98%. The sum of the deposits inside an isolation cone runs only over the tracks that survive these cuts.

The calorimeter-based observables refer to the electromagnetic (*ECal*) or to the hadronic (*HCal*) deposits in the cone around the muon track, but an improvement of the isolation efficiency can be achieved by using $CalIso = \alpha\ ECal + HCal$, where $\alpha = 1.5$ has been found as optimal.

Both track based and calorimeter based isolations are combined in a single isolation variable for final analysis. Performance of this combination is described in Section 6.2.1.

4.2.2 Muon measurements uncertainties

A data-driven method for the estimation of systematic uncertainties on the muon reconstruction efficiency in the context of a H $\to 4l$ analysis has been discussed in Ref. [37]. The method relies on the usage of a sample of inclusive W and Z bosons which are expected to be collected with small background and in the approximate ratio W:Z = 10:1 for a single "tag" muon with the requirement of $p_T > 19$ GeV$/c$. About 10^5 (10^6) W/Z events should be collected with the CMS detector for an integrated luminosity of 10 pb^{-1}(100 pb^{-1}) at the LHC. By counting the number of $Z \to 2\mu$ events in the resonance peak of the invariant mass distributions built using the tag muon and either all other tracks, or all other standalone muons, or all other globally reconstructed muons, one can evaluate the efficiency of finding globally-reconstructed muons. Such a measurement automatically accounts for the real detector performance, including intermittent and smooth variations in time. The efficiency of finding globally-reconstructed muons can thus be measured with an uncertainty of better than 1% for the integrated luminosity of 1 fb^{-1}considered for the Higgs boson search in this analysis. Thus, the four-muon efficiency therefore will be known with an absolute error of better than 4%. When deducing the expected $ZZ \to 4\mu$ events from the measured $Z \to 2\mu$ cross section, this uncertainty partially cancels out and becomes 2%.

The uncertainty on the muon p_T resolution directly propagates into the four-muon invariant mass $M(4\mu)$ reconstruction. This almost does not affect the background distribution. However, the $M(4\mu)$ distribution width drives the width of the $M(4\mu)$ window that we use for evaluating the signal excess significance at low Higgs-boson masses. Fortunately, even making a mistake in the $M(4\mu)$ distribution width by as much as 25% has only a small effect on evaluating a significance of an excess of events [7]. Also, the muon p_T resolution is fairly easy to measure from data using the measured J/ψ and Z peak widths with a precision much better than needed.

The uncertainty on the muon p_T scale can be similarly calibrated from data using the measured J/ψ and Z peaks. The effect of these uncertainties on the number of background events in a signal window appears only on steep slopes of the $M(4\mu)$ distribution. For the steepest part of the $M(4\mu)$ distribution in the $180 - 200$ GeV$/c^2$ range, we obtain $\delta b/b \approx 0.1\delta M_{4\mu}$, where $\delta M_{4\mu}$ is in GeV$/c^2$. This implies that to be able to neglect this effect, one needs to know the momentum scale with a precision of 0.1 GeV$/c$ at $p_T \approx 50$ GeV$/c$. This can be easily achieved with just a few hundred $Z \to 2\mu$ events.

Uncertainties from misalignment

The limited knowledge about the exact position and orientation of the CMS tracker and muon subdetectors in the first data taking period affects the performances of the muon/track reconstruction and impacts on physics.

In order to study the impact of the mis-alignment of the CMS tracking devices on the tracking procedures, realistic estimates for the expected displacements of the tracking systems are supplied in several scenarios that are supposed to reproduce the mis-alignment conditions during the first data taking period corresponding to 10 pb^{-1}(i.e. 100 $\text{pb}^{-1}$of integrated luminosity) and 100 pb^{-1}(i.e. 1000 $\text{pb}^{-1}$of integrated luminosity). Most of the estimates results from tests performed during integration, survey measurements taken at different stages during tracker assembly, laser alignment measurements and Monte Carlo studies on track based alignment.

For the strip tracker, we assume that we can align the subdetector position, the layer level and the rod/petal/string/ring level with an accuracy of 100μm. Pixel detector can be only partially aligned in the beginning of operation. A detailed list of expected values of the alignment uncertainties for layers (barrel) and disks (endcaps) for the mentioned scenarios is reported in Ref [**?**].

As a part of preparations for start-up LHC physics, these scenarios serve physics analysis studies as a tool of estimating systematics due to misalignment.

Track reconstruction efficiencies, track parameter and resonance mass resolutions have been evaluated using single muons with $p_T = 100$ GeV$/c$ or muons from $Z^0 \to \mu\mu$ decays. The track parameter resolutions deteriorate with increasing misalignment, as expected. The pT resolution is 1.5 – 2% in the barrel region in the case of no misalignment. In case of the 100 and 10 pb^{-1}scenarios the misalignments are larger and the pT resolution degrades to 2.5-4% and 5.5-8% in the barrel, respectively, as show in Figure **??** (left). The transverse impact parameter resolution is dominated by the misalignment of the pixel detector. It is observed to degrade from 10 μm for the ideal geometry to values $\sim$ 12, $\sim$ 15 and $\sim 50\mu m$ in case of the 1000, 100 and 10 pb^{-1}scenarios, respectively. The longitudinal impact parameter resolution is less strongly affected in the misalignment scenarios. Therefore primary vertex finding efficiency and position resolution are affected by the tracker mis-alignment, too.

Muons in a large spectrum of transverse momentum coming from the Z decay are used to investigate the impact of tracker misalignment on the p_T dependence of track parameter resolution, especially on p_T resolution.

Also shown is the di-muon invariant mass. The width of the reconstructed Z peak increases of a 12 % for the 10 pb^{-1}misalignment scenario. The effect of muon chamber misalignment on the Z peak width is negligible compared to the one of the tracker.

Uncertainties for muon isolation

The main sources of efficiency losses and systematics on lepton isolation are the so-called underlying events (UE) originating from the very same pp collisions, and the pile-up caused by random coincidences with other pp collisions in-phase (from the same proton-bunch crossing) or out-of-phase. The out-of-phase pile-up can be neglected at the LHC startup luminosity of 10^{32} $\mathrm{cm}^{-2}\mathrm{s}^{-1}$. The effects due to in-phase pile-up are tamed by the requirement that the four primary leptons originate some the same primary vertex. The UE are, by nature, unavoidable.

The question of how well can the isolation cut efficiency be predicted on the basis of current Monte Carlo event generators, given the poorly known UE physics, was studied extensively in Ref. [38] in the case of 4μ final states. It was shown that the variation of the isolation cut efficiencies per muon with different UE models, can be as large as 5%. A result which further depends on the details of the lepton isolation cut such as the minimum p_T threshold above which charged tracks are considered.

Given that the theory-induced uncertainty is comparable to other leading systematic effects in this analyses, a strategy has been developed to measure instead the isolation cut efficiency using the experimental data themselves (so called random cone technique). The objective is to demonstrate from associated experimental systematic uncertainties that a data-driven approach allows for a better control of the lepton isolation uncertainties. Random cone technique for measuring muon isolation cut efficiency was proposed and studied in details in the context of the CMS Physics TDR [39] and in Ref. [38].

5 Event Skimming and Pre-selection

The first step of the event selection is performed in two parts, called skimming and pre-selection. The main strategy of this step is to get rid of unuseful events, which will never pass the entire analysis selection, whilst preserving the maximal signal efficiency and the phase space for background systematic studies. The aim of the first part, called skimming, is technical: to reduce event rate to a manageable data volume. The main goal of pre-selection is to eliminate fake events, in particular QCD events. The end results of pre-selection will serve as reference for later stages efficiency calculation.

5.1 Event skimming

A common $\mathrm{H} \rightarrow \mathrm{ZZ}^{(*)} \rightarrow 4\ell$ "skimming" is defined to select signal events with close to 100% efficiency, and reduce significantly backgrounds from QCD, W and Drell-Yan production together with other dominant backgrounds. The skimming is applied to the so-called "electron-" and "muon-primary data streams". These data streams are built as a logical "or" of, respectively, all electron and muon HLT paths, representing an expected event rate of about 30 to 40 Hz [28].

To further suppress backgrounds, a requirement of at least two leptons (e or μ) with $p_T^\ell > 10$ GeV/c and one additional lepton with $p_T^\ell > 5$ GeV/c is applied on the events to be retained. The Fig. 11 shows the generated distribution of the three highest p_T leptons in $\mathrm{H} \rightarrow \mathrm{ZZ}^{(*)} \rightarrow 4\ell$ signal events for the Higgs mass at 120 GeV/c^2 , for the example of a Higgs decay to 4 electrons.

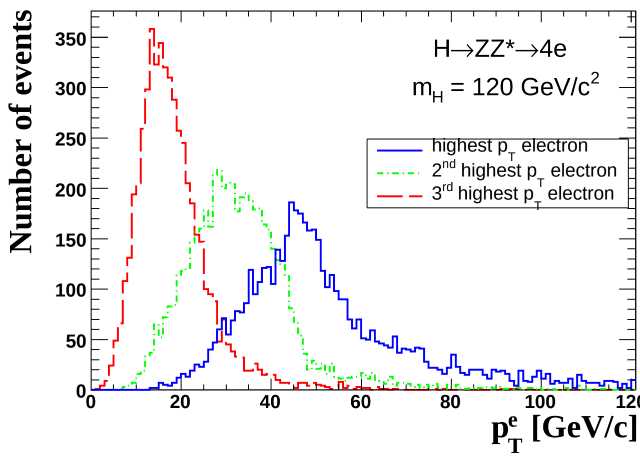


Figure 11: Generated distribution of the three highest p_T leptons in $\mathrm{H} \rightarrow \mathrm{ZZ}^{(*)} \rightarrow 4\ell$ signal events with m_{H} = 120 GeV/c^2 , for the example of the $\mathrm{H} \rightarrow \mathrm{ZZ}^{(*)} \rightarrow 4e$ channel.

The Figure 12 shows the skimming efficiencies, for all the studied Higgs boson masses and for the decay channels 4μ, $4e$, and $2e2\mu$ as well as for three main backgrounds. The efficiencies for the signal are evaluated for events where all four leptons were generated within the acceptance of the detector and that have been triggered by the HLT. For the $\mathrm{H} \rightarrow \mathrm{ZZ}^{(*)} \rightarrow 4\mu$ channel, the efficiency for the skimming step is $\gtrsim$ 99% for $m_{\mathrm{H}} \geq 135$ GeV/c^2 . For the $\mathrm{H} \rightarrow \mathrm{ZZ}^{(*)} \rightarrow 4e$ channel, the efficiency for the skimming step is $\gtrsim 92\%$ for $m_{\mathrm{H}} \geq 135$ GeV/c^2 . For the $\mathrm{H} \rightarrow \mathrm{ZZ}^{(*)} \rightarrow 2e2\mu$ channel, the efficiency for the skimming step is $\gtrsim$ 98% for $m_{\mathrm{H}} \geq 135$ GeV/c^2 . None of the events which are not surviving the skimming requirement would have passed the analysis selection.

5.2 Event pre-selection

A set of pre-selection cuts is then applied to suppress the contribution of "fake leptons". One main objective is to bring the QCD multijets and Z/W+jet(s) contributions at a level comparable or below the contribution of the three main backgrounds, $t\bar{t}$, $Zb\bar{b}$ and $\mathrm{ZZ}^{(*)}$. By reducing the number of extra (fake) leptons in signal events (coming e.g. from jets recoiling against the Higgs boson), the pre-selection also has the virtue of reducing the problem of the combinatorial ambiguities caused by the presence of more than 4 leptons. The choice of the 4 leptons truly belonging to the $\mathrm{H} \rightarrow 4\ell$ decay is thus post-poned beyond the pre-selection step.

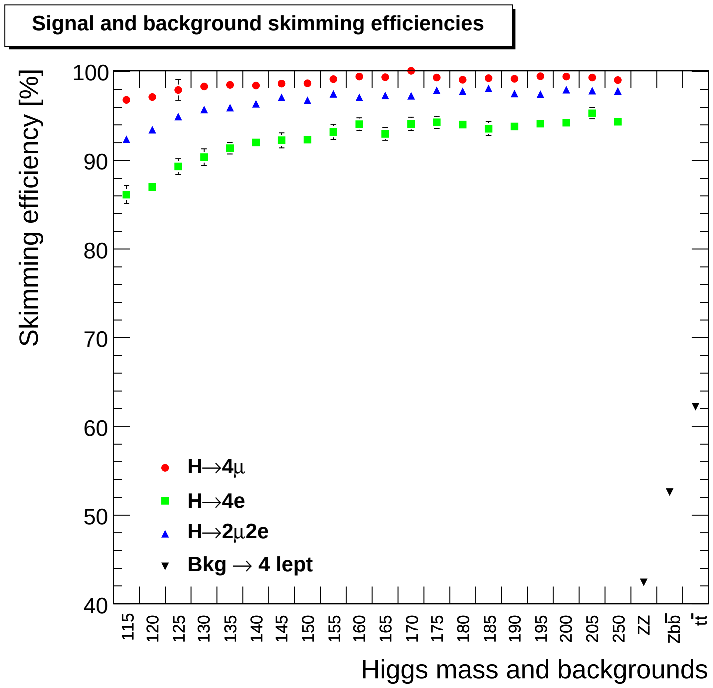


Figure 12: Skimming efficiencies for all considered Higgs boson masses, for all three 4-lepton channels and for three main backgrounds

The pre-selection "signal-like" events contains four steps:

- at least two l^+l^- pairs of identified leptons with opposite charge and matching flavors. The electrons are required to satisfy $p_T^e > 5$ GeV/c and $|\eta^e| < 2.5$, the muons are required to satisfy $p_T^\mu > 5$ GeV/c in the barrel, $p_T^\mu > 3$ GeV/c and $P^\mu > 9$ GeV/c in the endcaps, and $|\eta^\mu| < 2.4$
- at least two different matching pairs with invariant mass $m_{l^-l^+} > 12\ \mathrm{GeV/c^2}$,
- at least one combination of two matching pairs with an invariant mass greater than 100 $\mathrm{GeV/c^2}$,
- for $4e$/4μ and $2e2\mu$ channels, at least 4, respectively 2, loose track-based isolated electron/muon candidates. The loose track-based isolation is defined as $\sum p_T^{\mathrm{tracks}}/p_T^{\mathrm{e}} < 0.7$ for the electrons and $\mu ISO < 60$ for the least isolated muon (see section 6.2.1). This requirement preserves 98.5% of the signal events at $m_{\mathrm{H}} = 150$ $\mathrm{GeV/c^2}$ passing the previous step, for the example of the $\mathrm{H} \rightarrow \mathrm{ZZ}^{(*)} \rightarrow 4e$ decay channel.

The requirement of the two leptons pairs with opposite charge and matching flavours is a step beyond the trigger and skimming, corresponding to the principal characteristics of the signal events topology. The identification of leptons is chosen to be loose ($\sim$ 95% efficient on signal events) and brings additional rejection power against fake leptons. The cut on $m_{l^-l^+}$ is introduced to eliminate the contamination from low mass hadronic resonances. Requiring that at least one Higgs candidate fullfills an invariant mass of 100 $\mathrm{GeV/c^2}$ rejects more events with fake leptons and brings a safe additional step into the signal phase space. Imposing a loose track-based isolation is needed in the case of the electrons channels, to reject the fake QCD background events passing the previous requirements.

The efficiencies of the different steps of the pre-selection for the $\mathrm{H} \rightarrow \mathrm{ZZ}^{(*)} \rightarrow 4\ell$ signal events as a function of the Higgs mass are shown in Fig. 13, for the three decay channels. The total pre-selection efficiency on the signal events reaches a plateau value of 56%, 65% and 84 %, for $4e$, $2e2\mu$, 4μ channels respectively, around $m_H = 170\ \mathrm{GeV/c^2}$. The largest loss of signal events is due to leptons reconstruction inefiiciencies and charge mis-identification. [3)]

[3)] These aspects have been deeply studied recently, and largely improved (cf. section 4).

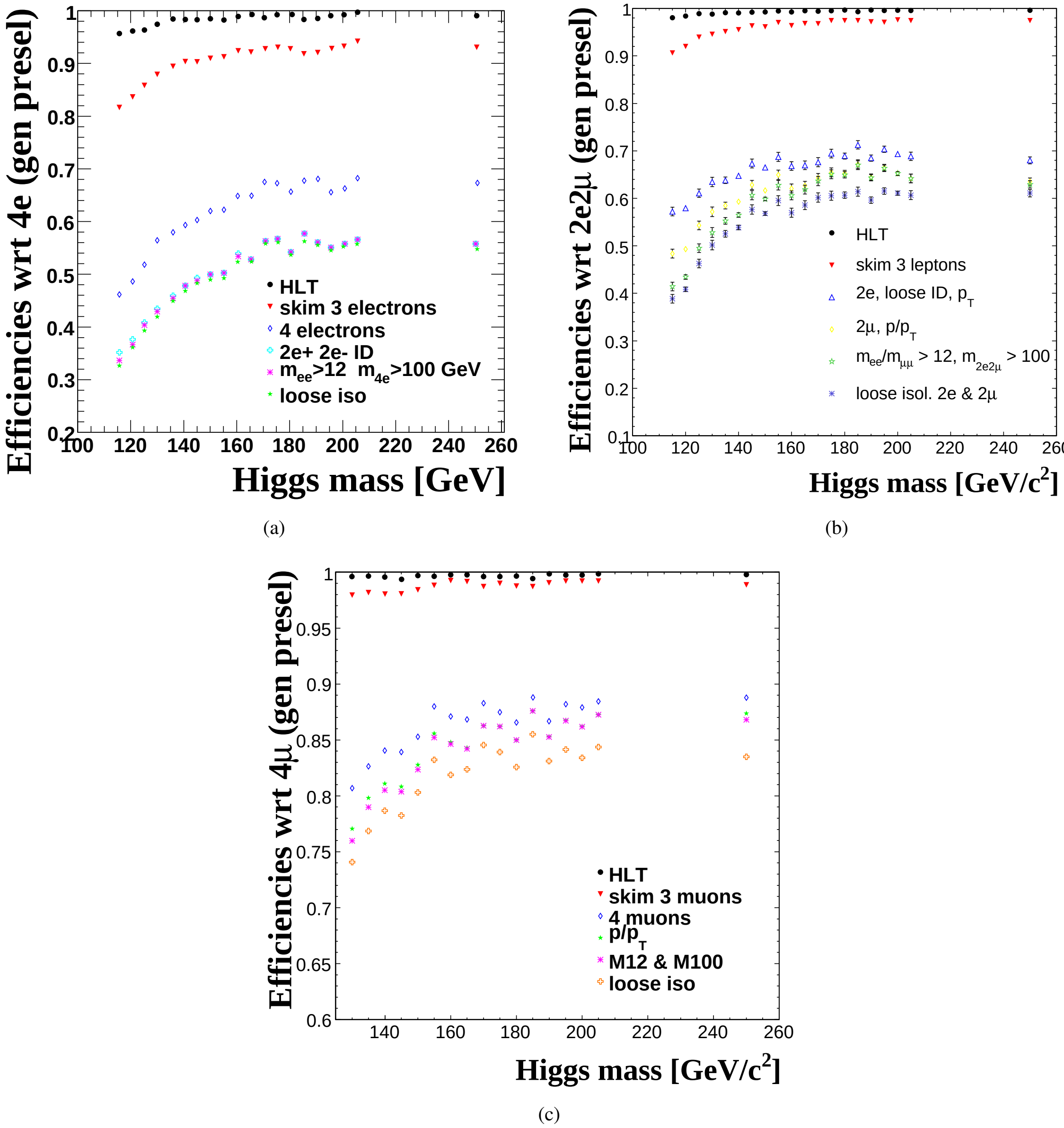


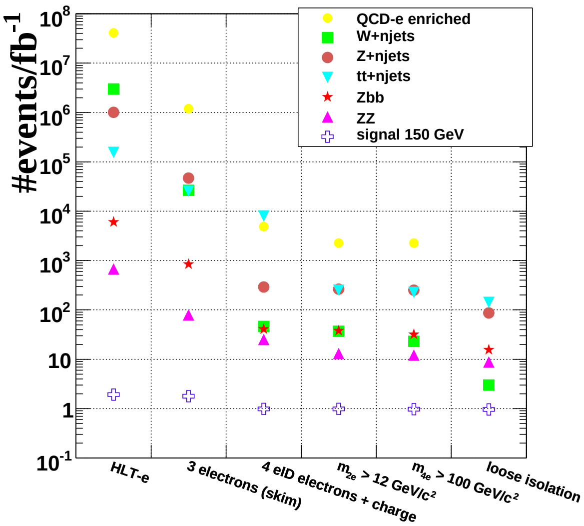
Figure 13: Efficiency of HLT, skimming and each pre-selection step for $\mathrm{H} \to \mathrm{ZZ}^{(*)} \to 4\ell$ signal events in the (a) $4e$, (b) $2e2\mu$ and (c) 4μ channel, as a function of the Higgs mass.

The expected reduction of event rates obtained from the pre-selection of signal-like events is shown in Fig 14. The background events, largely dominated at the first steps by the fake ones, are brought down until the same order of magnitude at the end of the pre-selection.

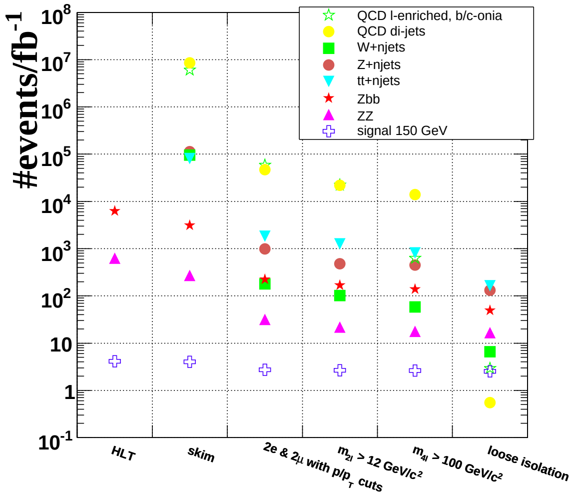
The invariant mass constructed with 4 leptons after pre-selection is shown in Fig. 15. For this plots particularly, and later in the analysis whenever lepton should be associated with vector bosons, if more that four leptons are reconstructed, the four leptons forming the invariant mass shown in the Fig. 15 are chosen in the following way:

- the pair formed with the same flavor and opposite charge leptons, with the invariant mass the closest to the Z mass,
- and the pair formed with the same flavor and opposite charge highest p_T remaining leptons.

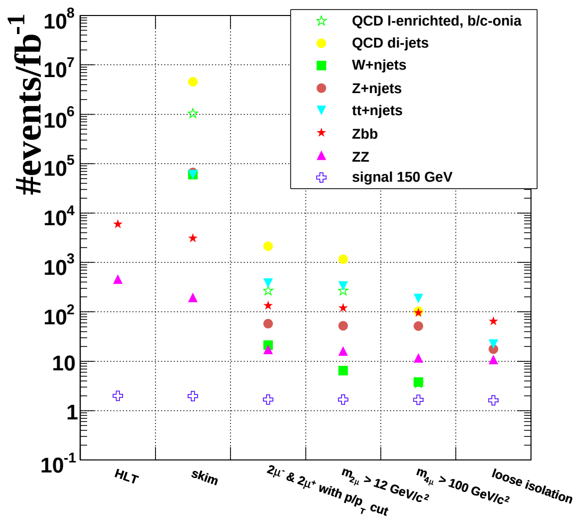
The Tab. 4 shows the numbers of expected events per fb^{-1}after the pre-selection for three Higgs masses and the backgrounds, for all three channels. The main background after pre-selection is Z+jets, followed by $t\bar{t}$ +jets and



(a)



(b)



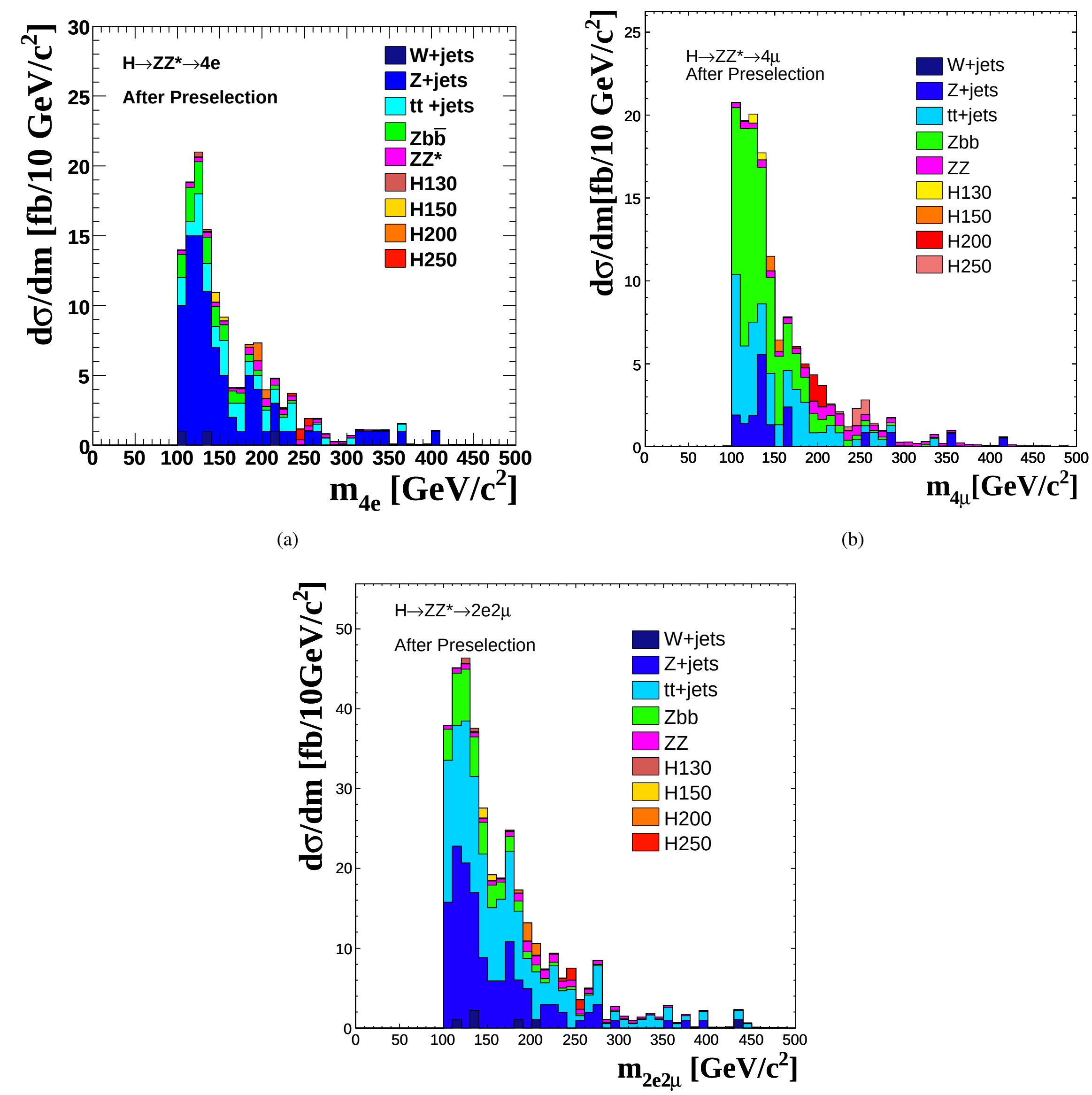
(c)

Figure 14: Reduction of the event rate for QCD, Z/W + jet(s), $t\bar{t}$ +jets, $(Z/\gamma^{*})\mathrm{b\bar{b}}$ and $\mathrm{Z}/\gamma^{*}\mathrm{Z}/\gamma^{*}$ backgrounds, and $\mathrm{H} \rightarrow \mathrm{ZZ}^{(*)} \rightarrow 4\ell$ signal at m_{H} =150 $\mathrm{GeV/c^2}$, after the skimming and each pre-selection step in the (a) $4e$, (b) $2e2\mu$ and (c) 4μ channel.

$(Z/\gamma^{*})\mathrm{b\bar{b}}$. At this stage, $\mathrm{Z}/\gamma^{*}\mathrm{Z}/\gamma^{*}$ is the least contributing background. All these remaining backgrounds, and especially Z+jets will be further reduced by tighter isolation, vertexing requirements and kinematic cuts.

	number of events per fb^{-1}		
	$4e$	4μ	$2e2\mu$
$\mathrm{H} \to 4l$ $m_H = 130$ GeV/c^2	0.53	0.90	1.20
$\mathrm{H} \to 4l$ $m_H = 150$ GeV/c^2	0.96	1.61	2.19
$\mathrm{H} \to 4l$ $m_H = 200$ GeV/c^2	2.03	3.11	4.53
$\mathrm{Z}/\gamma^*\mathrm{Z}/\gamma^*$	8.43	10.8	16.7
$(Z/\gamma^*)\mathrm{bb}$	15.5	64.7	38.1
$t\bar{t}$ +jets	27.5	22.3	163.7
Z +jets	87.0	17.6	132.1
W +jets	3.00	0	8.00

Table 4: Summary of the number of events expected per fb^{-1}after the pre-selection for the three $\mathrm{H} \to \mathrm{ZZ}^{(*)} \to 4\ell$ channels, for three different Higgs masses and the backgrounds.



(a) (b) (c)

Figure 15: Four leptons invariant mass after pre-selection in the (a) $4e$, (b) $2e2\mu$ and (c) 4μ channels for $\mathrm{H} \to \mathrm{ZZ}^{(*)} \to 4\ell$ signal events and the backgrounds.

6 Analysis Strategy and Discriminating Observables

6.1 General Considerations

The event sample obtained following the trigger, skimming and pre-selection steps is dominated by the $t\bar{t}$ and $\mathrm{Z}b\bar{b}$ backgrounds. The trigger and skimming allowed for a drastic reduction of the event rates while preserving the highest possible Higgs boson signal detection efficiency. At this early stage, the event sample was dominated by background contributions involving two or more "fake" leptons. The background events, largely dominated at the first steps by QCD and Z/W+jet(s), with high fake lepton rates coming from gluon or light quark jets, are considerably suppressed by the pre-selection steps. No QCD or W+jet(s) event survive the loose isolation step within the available Monte Carlo statistics. The Z+jet(s) rate is brought down at a level comparable to the rate from $t\bar{t}$.

Two major reducible backgrounds to the signal remain at this stage. These are the $t\bar{t} \rightarrow \mathrm{W}^{+}b\mathrm{W}^{-}\bar{b}$ and $\mathrm{Z}b\bar{b}$ with leptons coming from the decays of the b quarks. Such leptons are likely to be accompanied by hadronic products from the fragmentation and decay processes initiated in the b-quark jets. Moreover, because of the long lifetime of b-hadrons, they are likely to have a large impact parameter with respect to the primary vertex. Thus, lepton isolation and lepton impact parameter measurements allow for a powerful rejection of the reducible backgrounds. While these characteristics might be sufficient to eliminate the leptons from heavily boosted b-quark jets in $t\bar{t}$ events, the b-quark jets in $\mathrm{Z}b\bar{b}$ events are in general less collimated in the detector and lead to leptons with a softer p_T spectrum. In order to best preserve the signal detection efficiency while acting on low p_T^{ℓ} lepton candidates to suppress the $\mathrm{Z}b\bar{b}$ background, the isolation criteria for the leptons from the pair at lowest $m_{\ell^-\ell^+}$ can be made p_T^{ℓ} dependent.

In summary the main discriminating variables for the event selection are:

- isolation,
- impact parameter,
- p_T of all four leptons,
- two-leptons invariant masses,
- four-leptons invariant mass.

6.2 Discriminating Observables

6.2.1 Lepton isolation

Four leptons coming from the Higgs decay are expected to be isolated (i.e. not inside a jet) and this provides an excellent way to distinguish the signal from the *reducible* backgrounds, $Z\mathrm{b}\bar{\mathrm{b}} \rightarrow 4l + X$ and $\mathrm{t}\bar{\mathrm{t}} \rightarrow 4l + X$, where two leptons are produced inside the b-jets.

Muon Isolation. The isolation of a muon can be quantified by considering the energy or momentum of the particles in a cone around the muon track. Among the several investigated variables, the one that showed the highest discriminating power is μISO, defined as

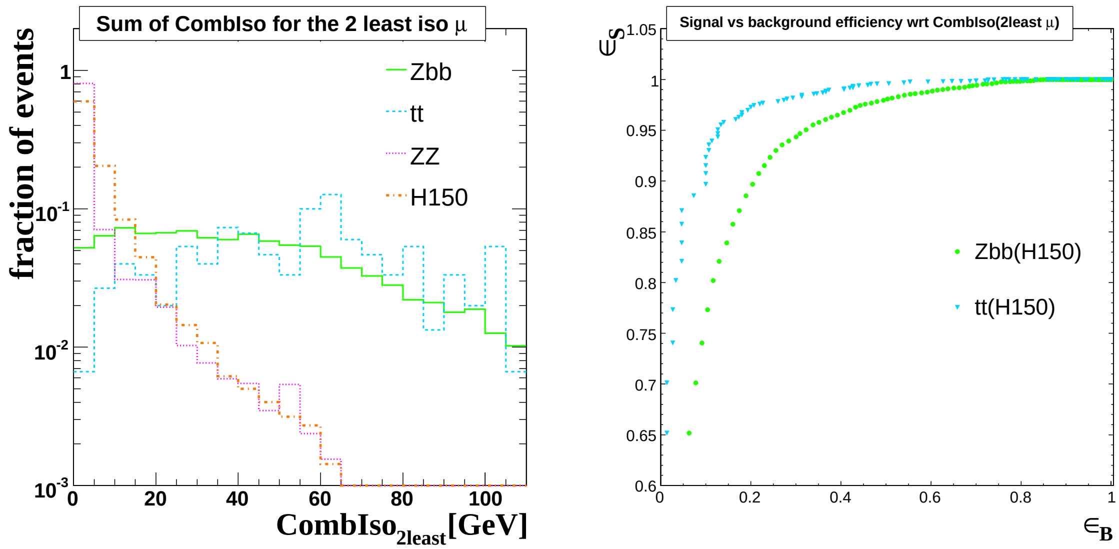
$$\mu ISO = 2 \cdot \mu ISO_{track} + 1.5 \cdot \mu ISO_{ECAL} + \mu ISO_{HCAL}$$

Track based and calorimeter based isolation variables are defined in the section 4. Distribution of combined variable for signal and main backgrounds is shown in Fig. 16 (left).

Studies have been carried out on the stability of the isolation variable, to check that the efficiencies do not strongly depend on the variable definition. Other μISO-like observables have been introduced, changing the parameters and bringing a linear, hyperbolical, circular or elliptical dependence on TrkIso and CalIso. The best discriminating observable has been chosen. The differences with respect to μISO signal and background efficiencies are anyway small and compatible with statistical errors.

After having determined the isolation variables for all muons, a decision about the whole event has to be taken. It has been found that the most effective way to define an event as isolated, is to consider either the least isolated muon out of the four ones coming from two Z(Z*)-candidates decays, or the two least isolated ones. In this latter case, the isolation variables of the two μ's are be summed. It turned out that it is better to consider the sum of the two least isolated muon isolation variables. The efficiency improvement is small and often compatible with statistical fluctuations, but it is common to most choices of isolation variables and background samples, due to the fact that both b and $\bar{b}$ quarks tend to decay semileptonically, yielding two muons that are likely to be non-isolated.

The signal efficiency for the sum of the two least isolated muons μISO observable are shown in Figure 16 (right) as a function of the background efficiency. For a cut $\mu ISO < 30$ is 95%, while background efficiencies are 29% and 9% for $Zb\bar{b}$ and $t\bar{t}$ respectively. Efficiencies are calculated with respect to preselection.



(a) (b)

Figure 16: (a) Muon combined isolation variable distribution for the sum of two least isolated muon, for signal and main backgrounds. (b) Combined isolation cuts power against $t\bar{t}$ and Z $b\bar{b}$ backgrounds.

Another promising selection criterion arises from observing the bidimensional distribution of the sum of μISO for two least isolated muons (henceforward $X_{2\ least}$) versus the p_T of the third and fourth muon (when sorting by decreasing p_T the four muons coming from the Higgs decay). This is a reasonable conclusion, as the muons originated in the b-jets have usually low p_T, while the muons from Z or W decay are more energetic. Therefore in $Z\mathrm{b\bar{b}}$ and $\mathrm{t\bar{t}}$ events, the third and fourth muons are usually characterized by low p_T and high $X_{2\ least}$ values, unlike signal events.

The $X_{2\ least}$ vs $p_{T,3}$ distributions for signal and irreducible backgrounds, shown in Fig 17, are well separated, so that the plane can be divided into two regions respectively dominated by the signal or the backgrounds. It's easy to see that such regions are best separated by a slanting line of the form $X_{2\ least} = A \cdot p_{T,3} - B$. Optimization procedures indicate $A = 1.5$ GeV^{-1} and $B = 15$ as the best values in terms of background rejection and signal for Higgs boson mass of 150 GeV/c^2 and it can be optimized further with different masses, as can be seen in Fig 17.

Electron Isolation. For electron isolation track based and calorimetric based isolation variables are combined in the following way:

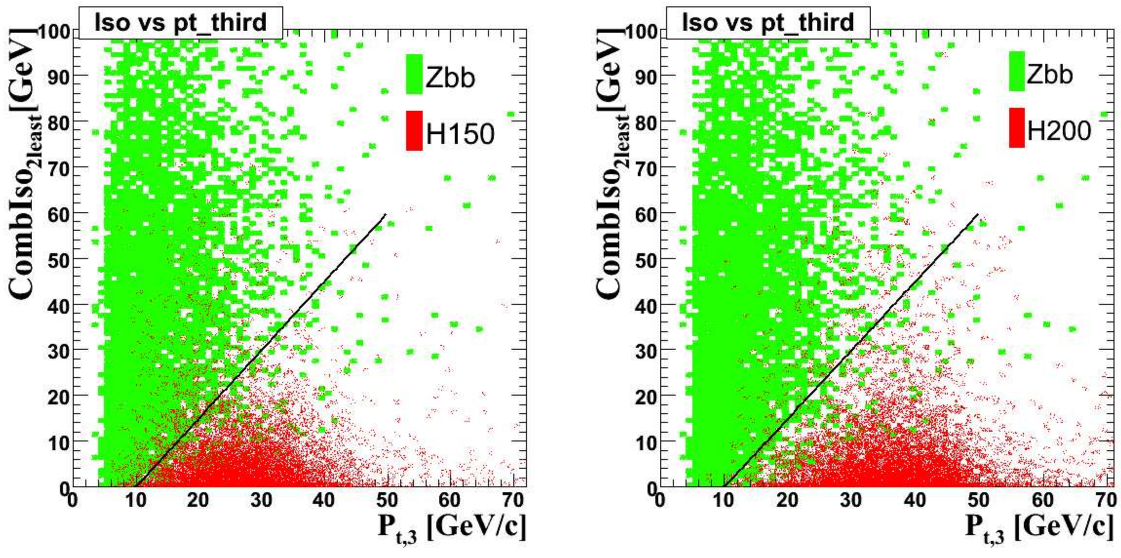
$$eISO = eISO_{track} + eISO_{HCAL}$$

Track based and calorimeter based isolation variables are defined in the section 4. As a final discriminating variable a sum of combined variable for two least isolated electrons is used. Distribution of this variable for signal and main backgrounds is shown in Fig. 18 (left), while the rejection power against $t\bar{t}$ and Z$b\bar{b}$ backgrounds is shown in Fig. 18 (right). The signal efficiency for $eISO_{2least}$ observable and for a cut $eISO_{2least} < 0.35$ is 96%, while background efficiencies are 39% and 22% for $Zb\bar{b}$ and $t\bar{t}$ respectively. Efficiencies are calculated with respect to preselection. Bidimensional cut in $eISO_{2least}$ versus third and fourth electron p_T is also used for electrons in $H \to 4e$ analysis.

6.2.2 Impact parameter requirements

The leptons from the Higgs boson signal originate from a common primary vertex (PV) in contrast to the leptons from at least one $\mathrm{l^+l^-}$ pair reconstructed in $\mathrm{t\bar{t}}$ and $\mathrm{Zb\bar{b}}$ background events. Therefore the impact paremeter (IP) is expected to be close to 0 for particle originated in the PV and increases with the distance of the production vertex from the PV: hence, due to the long b-quark lifetime, the particles in b-jets have usually a large IP. This information is exploited to further improve the separation of the signal and background events. Various methods have been explored: "3D" impact parameter of muons divided by its error, combination of longitudinal and transversal impact parameter of electrons divided by the corresponding errors, compatibility of the 2 leptons vertex or of the 4 leptons vertex with the primary vertex, distance between 2 leptons at the closest approach to the primary vertex. The first two methods result to give the best signal efficiency and the best background reduction and were thus used for this analysis. The comparison between the results obtained with these different methods will give us a good confidence on the stability fo the results.

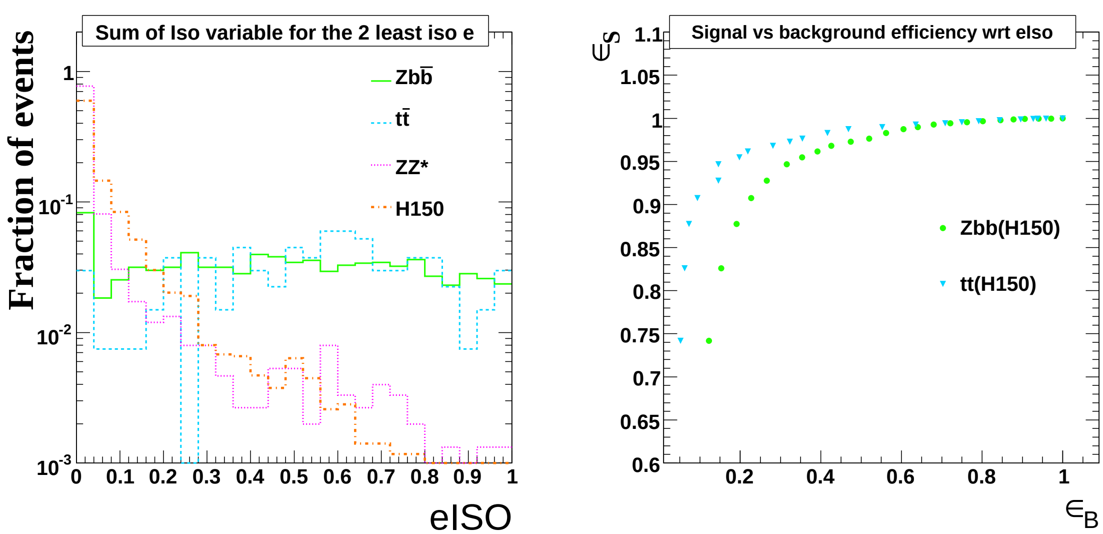
Muon Impact Parameter.



(a) (b)

Figure 17: Combined muon isolation variable versus pT of the third lepton for different Higgs boson masses of 150 GeV/c^2 (left) and 200 GeV/c^2 (right)

.



(a) (b)

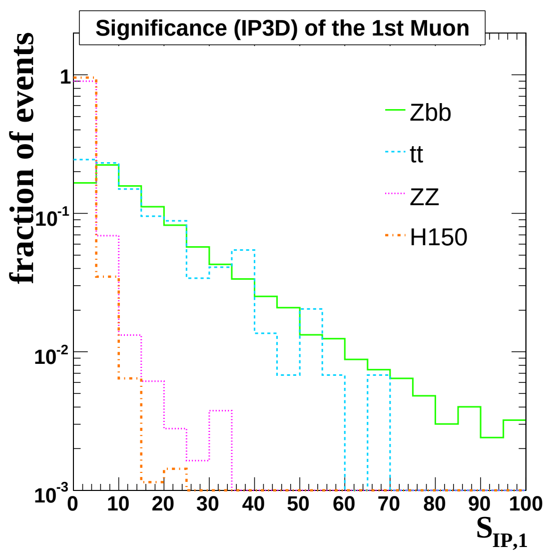
Figure 18: (a) Electron combined isolation variable distribution for the sum of two least isolated electrons, for signal and main backgrounds. (b) Combined isolation cuts power against $t\bar{t}$ and Z $b\bar{b}$ backgrounds.

To take into account the finite resolution of the detector, the variable to be used for the analysis is actually the *IP significance*, i.e. $s_{IP} = \frac{IP}{\sigma_{IP}}$, where IP is the 3D distance between the PV of the event and the point of closest approach of the track to the PV. After sorting the s_{IP} of the four muons in increasing order, the fourth, i.e. worst (distribution shown in Fig 19(a)) or the third (second worst) or both can be used to distinguish signal from background. The best criterion found is to require $s_{IP}(4^{st}\,\mu) < 12$ and $s_{IP}(3^{nd}\,\mu) < 4$. The power of the 3DIP eIP used against $t\bar{t}$ and $Zb\bar{b}$ backgrounds, after preselection, is illustrated in Fig 19(b).

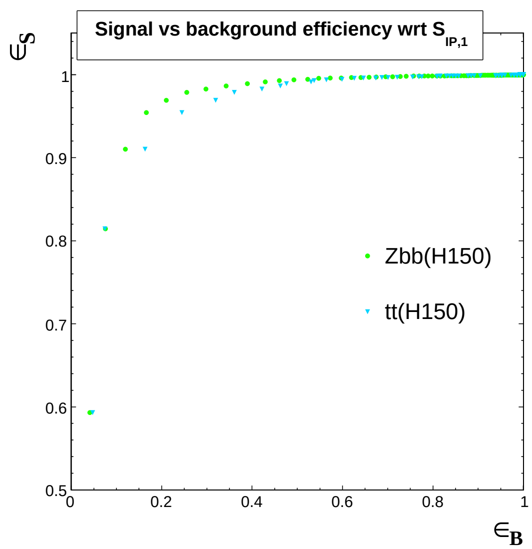
Electron Impact parameter. A similar method has been explored for electrons. First, the significance of the longitudinal impact parameter for each electron candidates is calculated: $SLIP = IP_L/\sigma_L$, where σ_L is the uncertainty (typically 20 μm for primary tracks) on longitudinal impact parameter. Then the individual transverse impact parameter significances, $STIP = IP_T/\sigma_T$ (where σ_T is the uncertainty on IP_T), are summed up within e^+e^- pairs associated the Z and Z* bosons. The final electron impact parameter is constructed from the combination of $SLIP$ and $STIP$ for pair associated to the Z* boson:

$$eIP = SLIP + 2 \cdot (STIP_{e^+} + STIP_{e^-})$$

Distribution of the combined variable eIP in $H \to 4e$ events, for signal and main backgrounds, is shown in Fig 20(a), while the power of the combined variable eIP used against $t\bar{t}$ and $Zb\bar{b}$ backgrounds, after preselection, is illustrated in Fig 20(b).



(a)



(b)

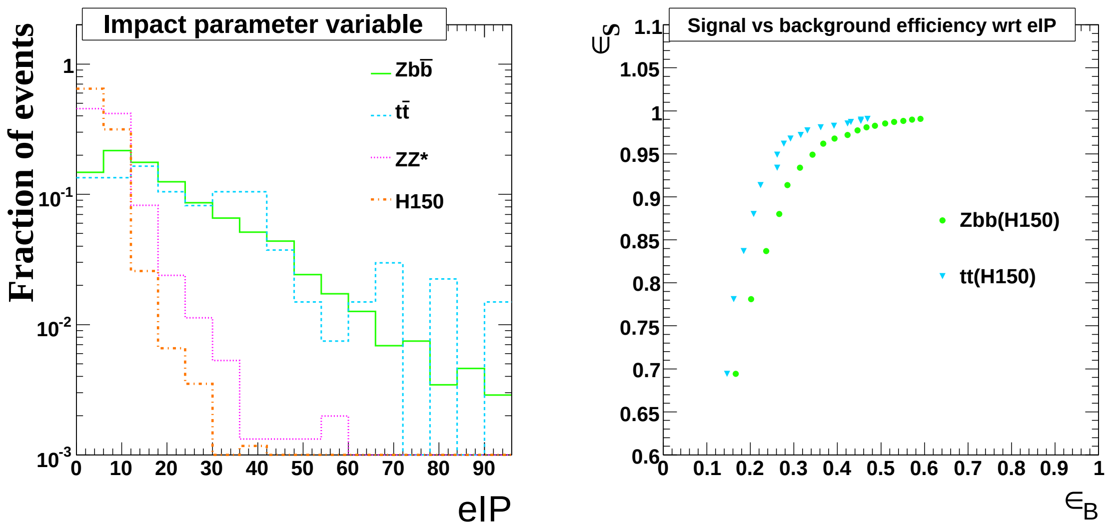
Figure 19: (a) Muon impact parameter significance distribution for the worst muon, for signal and main backgrounds. (b) Impact parameter significance cuts power against $t\bar{t}$ and Z $b\bar{b}$ backgrounds.

6.2.3 Kinematics

Taking advantage of the expectation of a narrow resonance in the m_{4l} spectrum, and of the likely presence of a real Z boson in the final state, the selection can be further improved using mass dependent kinematic requirements.

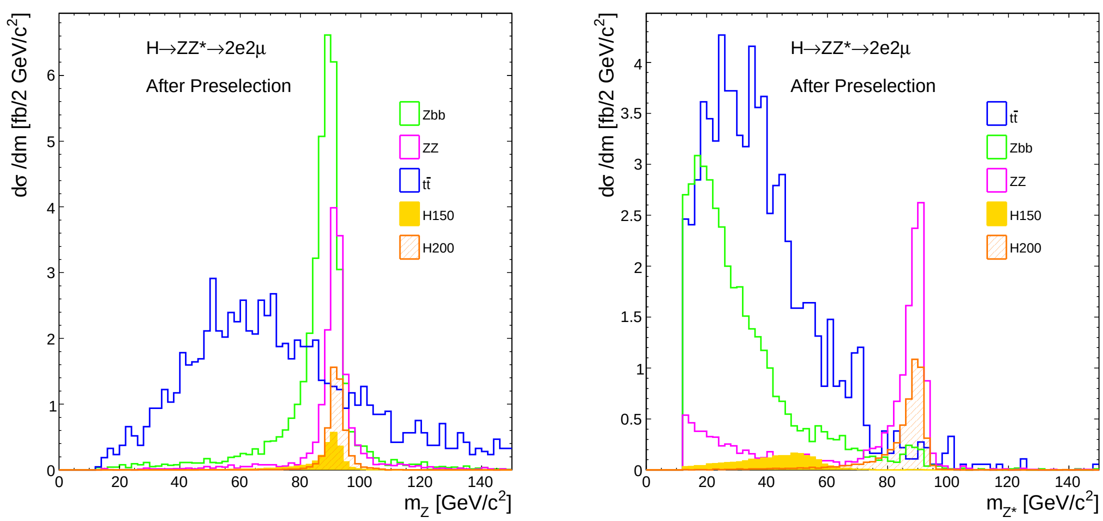
First, the electrons of the l^+l^- pair associated to the Z* have a much harder p_T^e spectrum for the Higgs boson signal than for the $t\bar{t}$ and $Zb\bar{b}$ backgrounds. Second, the mass spectrum of the Z bosons distinguishes the Higgs boson signal from the $ZZ^{(*)}$ background. These kinematic requirements are used in final event selection.

As an example of discrimination power of a cut on invariant mass of leptons pairs, distribution of invariant mass is given in Fig 21 for signal and main backgrounds.



(a) (b)

Figure 20: (a) Electron impact parameter variable distribution for signal and main backgrounds. (b) Impact parameter cuts power against $t\bar{t}$ and Z $b\bar{b}$ backgrounds.



(a) (b)

Figure 21: (a) Invariant mass of two leptons closest to the nominal Z mass for signal and main backgrounds, in the $2e2\mu$ channel. (b) Invariant mass built from two remaining leptons with the highest p_T, matching charge and flavor, in the $2e2\mu$ channel.

Source	$ZZ^{(*)}$ background (%)
Luminosity	-
Trigger, reconstruction and identification efficiency	6
Lepton isolation	2
Miscalibration and misalignement	2
Theory	3
MC statistics	3
Total	8

Table 5: The main source of systematics uncertainties and their value for the $ZZ^{(*)}$ background.

7 Systematic Uncertainties

For a very low integrated luminosity at the LHC where the discovery of a SM-like Higgs bosons in the $H \rightarrow ZZ^{(*)}$ channel is unlikely, the emphasis of the analysis will be put on the understanding of detector measurement uncertainties as well as on the control of background uncertainties from data.

At integrated luminosity of 1 fb^{-1}, for which analysis in this paper is presented, only handfull number of events is expected in the finally selected signal region. Therefore expected uncertainties will be largely dominated by statistical errors.

For completeness, main sources of systematics and their control are summarized in this section, with some novel ideas about measuring $Zb\bar{b}$ background from data and verifying if the remaining background is composed mainly of ZZ background. In table 5 the systematic errors for the main background are summarized.

7.1 Experimental Uncertainties

The main experimental sources of systematic uncertainties for four lepton channels are:

- the integrated luminosity,
- the trigger efficiency,
- the lepton reconstruction and identification efficiency,
- the lepton isolation cut efficiency,
- the mis-calibration and mis-alignment,
- the four-lepton mass M_{4l} absolute scale and resolution $\sigma(M_{4l})$.

The absolute luminosity normalization can be obtained from the measurement of inclusive Z or W vector boson production with the production cross-section taken from the theory. A limitation on the precision of the integrated pp luminosity comes from the limited knowledge on parton density functions in the calculation of the theoretical cross-section. Otherwise, uncertainties on the detector precision at early stages of operations at the LHC will affect the cross-section obtained e.g. from a fit to the measured Z peak. The uncertainty on the integrated luminosity is expected to be of $\mathcal{O}(10\%)$ for measured luminosities of up to 1 fb^{-1}.

The uncertainty on the trigger efficiency is expected to be negligible in the signal-like phase space of the four-lepton analysis where the absolute trigger efficiency approaches 100 %.

Methods have been developed to evaluate experimental systematics from data where possible, and to make use of extrapolations from Monte Carlo simulation of the detector response where needed. Data-driven techniques are used for example to evaluate systematics on lepton reconstruction efficiencies and lepton isolation (see sections 4.1.4 and 4.2.2). This is complemented by Monte Carlo extrapolation towards the very low p_T range. Monte Carlo simulations are used to estimate the propagate the effect of individual leptons of mis-alignment and mis-calibration on the measurement of M_{2l} and M_{4l} taking into account the residual errors expected from low luminosity data analysis. Thus, comparing for example the Monte Carlo expectation for M_{2l} with the measured Z mass resolution, will help in establishing the credibility of the experimental systematic errors which are noticeably difficult to monitor at the early stage of the detector operation at the LHC, when the changes in the system conditions are frequent.

The lepton momentum scale has been studied in details in ref.[40] and the systematics effects on the Z cross sections have been computed. In this analysis for the Z decaying into muons, only ”global muons” are considered, with very similar cut of the Z analysis and thus the systematics errors is negligible (i.e. 0.05%).

7.2 Background Uncertainties

After having applied selection cuts, the main remaining background is the $ZZ \to 4l$ production. The $Zb\bar{b} \to 4l$ and $t\bar{t} \to 4l$ backgrounds are very much reduced in all four-lepton mass range, due to mainly isolation cuts. Therefore it is of outmost importance to measure $ZZ \to 4l$ from data. At LHC start-up luminosities, when we expect small number of background events in the $4l$ channel, statistical errors will dominate over systematic errors. Their relative importance is changing with increasing statistics, so control of systematics becomes an important issue. In this section measurement and control of all three main background is presented, following an usual procedure of choosing the control region outside the signal phase space and then verify that event rage changes according to our expectations. A great care should be taken by choosing the control region, since two reducible backgrounds $Zb\bar{b} \to 4l$ and $t\bar{t} \to 4l$, after relaxing some cuts, quickly become dominant making this extrapolation more difficult.

An important question in the case of discoveries with a few events over the expected background very close to zero is whether we have missed any other possible source of backgrounds. The real question in this case is whether we have missed any other possible source of backgrounds. Once the first few events are observed, one will need to find a way to confirm that what we see is indeed what we think we see.

One way of doing this is to relax some of the cuts and observe whether the event rate changes according to our expectations. Unfortunately, such extrapolations do not always work. E.g., this does not apply when we have two backgrounds and one of them is dominant for tight cuts and the other quickly overwhelms the first one for loosened cuts. Then, extrapolations of rates from loose toward tighter cuts are quite useless.

For this particular analysis, there is one fortunate property that allows us to validate the nature of the first observed events without really changing cuts. The allowed combinations for four leptons originating from $(H \to)ZZ^{(*)} \to 4l$ in this case are: $\mu^+\mu^-\mu^+\mu^-$, $\mu^+\mu^-e^+e^-$, $e^+e^-e^+e^-$. Actually, the way the first 3-4 events are shared between the three channels is not very informative. Given the very low statistics, nearly all possibilities will be reasonably plausible. Also, many backgrounds would tend to populate these three channels in the same proportions as the signal would do.

However, if the events we observe are indeed originate from $(H \to)ZZ$ and the background is virtually zero, then if one tries to repeat the analysis for four lepton pairs, but not sorting them by a flavor or a charge, then no new additional events will be found. If the observed events, on the other hand, are due to unaccounted backgrounds (such as $Zjj \to l^+l^- + 2$ “fake leptons” or $QCD \to 4$ “fake leptons”) then, in a striking contrast, many more events will appear (e.g. $\mu^+\mu^-\mu^+e^-$ or $\mu^+\mu^-e^+e^+$ and many others).

For example, if unaccounted-background-induced electrons and muons have little correlation in flavor and charge (like fake electrons and muons in QCD jets would do), the ratio between rates of “wrong” and “right” combinations would be (256-36)/36=6.1, which comes from a count of plain combinatorial possibilities (Figure 22). Therefore, if the 3 observed events were due to such unaccounted background, we would see about 18 events of wrong flavor/charge combinations in the same mass window. However, if the three events are indeed due to $ZZstar$, then we should see no wrong combinations.

Figure 23 illustrates the same concept on the example of the $\mathrm{t\bar{t}}$ and $(Z/\gamma^*)\mathrm{b\bar{b}}$ backgrounds (which, of course, are well accounted for in this analysis). One can see that despite strong flavor/charge correlations in these backgrounds, the ratio remains fairly high. Moreover, it is fairly independent of the isolation cut, which indicates how robust this cross check is.

7.2.1 Measurement of the ZZ background from data

This particular background is the major “irreducible” source of background events. There are two ways to control its rate: using side bands or a reference process - Z inclusive production in this case. Both methods have pros and cons and they mainly concern statistical and systematic uncertainty of ZZ prediction (see also section on systematic uncertainties).

ZZ has an advantage of having a large cross section, relatively clean (from backgrounds) and easily identifiable events of a reference process, Z inclusive production, which can be used to calibrate or cross check number of ZZ events. As described in general introduction to analysis section, we can build using MC-truth calibration function of ZZ background (vs. m(4l)) using Z events. And then once we have real data we can correct the function to ratio of Z events in MC-truth to number of observed Z events, correcting this way to subtle unaccounted experimental effects. Such calibrations should be obviously done at different level of (pre-)selections and final cuts, which in turn may be corrected to ratio of efficiency of cuts measured from MC-truth to efficiency measured directly from data as have been explored in refs [PTDR Notes].

7.2.2 Measurement of the $Zb\bar{b}$ backgrounds from data

After the baseline selection cuts, the remaining background besides irreducible background is mostly composed by $Z\mathrm{b\bar{b}}$ events. An appropriate strategy is therefore needed to reject such process as well as to precisely measure it from data. Both tasks are fulfilled by finding a kinematic region with a high discriminating power between signal and $Z\mathrm{b\bar{b}}$ background.

Two very discriminating variables are the isolation variables and the transverse momentum of the third muons (sorted in decreasing order). The most effective way to suppress background while keeping the signal efficiency almost untouched is to combined these two variables. In figure 17 the Isolation (as defined in section 4.2.1) is plotted as a function of the p_T of the third muon ($p_{T,3}$), for $Z\mathrm{b\bar{b}}$ background (in green) and signal after the preselection. The $\mathrm{t\bar{t}}$ distribution is very similar to $Z\mathrm{b\bar{b}}$and as well the Z+jets and W+jets background. The suggested cut is the line superimposed in figure, i.e. the signal events are selected for $\mathrm{X2least} < 1.5p_T(3) - 15$. Such cut allows a very strong suppression of $Z\mathrm{b\bar{b}}$, $\mathrm{t\bar{t}}$, Z/W+jets backgrounds: the respective efficiencies are 2.8% and 1.0%, corresponding to 2.6 and 1.3 events. The efficiency loss is not negligible only for low masses and it corresponds to (as an example) 14% for 130 GeV Higgs mass, 6% for 150 GeV Higgs mass and 1.4% for 205 GeV Higgs mass and 0.7% for 250 GeV Higgs mass.

On the contrary, in the rejection region (i.e. $\mathrm{X2least} > 1.5p_{T,3} - 15$) $Z\mathrm{b\bar{b}}$ and $\mathrm{t\bar{t}}$ background events dominate while the signal is basically absent. Hence, provided a good control of the $\mathrm{t\bar{t}}$ and Z/W+jets background, $Z\mathrm{b\bar{b}}$ can be measured from data and compared to the predictions from MC. Finally the difference between data and MC can be propagated to the signal region. Being the number of remaining $Z\mathrm{b\bar{b}}$ events in the signal region very small (2.6 events for 1 fb^{-1} of integrated luminosity), the systematic error will not affect drammatically the measurement. In the background region, as already stated, $Z\mathrm{b\bar{b}}$ has to be separated by the $\mathrm{t\bar{t}}$ and W/Z+jets contribution. A very good variable is the invariant mass of the first Z candidate (i.e. the $\mu^+\mu^-$ pair which minimizes the difference $|m_{inv}(\mu^+\mu^-) \; - \; m_{\mathrm{Z}}|$). To better study the estimate uncertainty on the background from data, a "pseudo-experiment" with 1 fb^{-1} of luminosity is performed. In figure 24 the mass of the first Z candidate is shown for the events satisfying the condition $\mathrm{X2least} > 1.5p_{T,3} - 15$ at the chosen luminosity. The $Z\mathrm{b\bar{b}}$ events are

		m+ m+	m+ m-	m+ e+	m+ e-	m- m+	m- m-	m- e+	m- e-	e+ m+	e+ m-	e+ e+	e+ e-	e- m+	e- m-	e- e+	e- e-
m+	m+	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1
m+	m-	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1
m+	e+	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1
m+	e-	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1
m-	m+	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1
m-	m-	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1
m-	e+	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1
m-	e-	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1
e+	m+	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1
e+	m-	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1
e+	e+	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1
e+	e-	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1
e-	m+	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1
e-	m-	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1
e-	e+	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1
e-	e-	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1

Figure 22: Combinatorial possibilities for four-lepton combinations in assumption of no charge/flavor correlations. Gray cells indicate wrong combinations, colored cells indicate right combinations (red for 4μ, blue for $4e$, and green for $2\mu 2e$). Numbers inside cells indicates relative weight of events; in this case they are all taken equally probable.

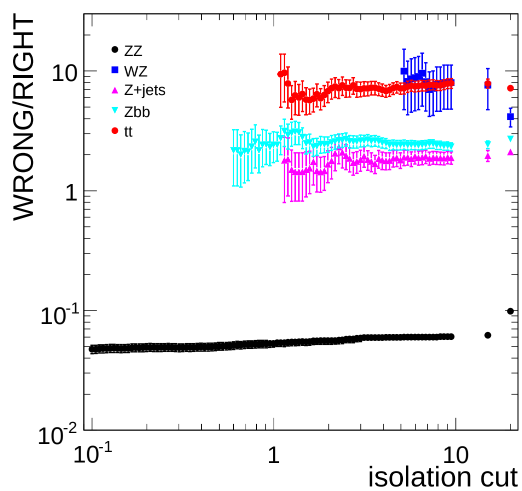


Figure 23: Ratio of "wrong" and "right" 4-lepton combinations for $\mathrm{t\bar{t}}$ and $(Z/\gamma^*)\mathrm{b\bar{b}}$ events as a function of the isolation cut

selected by requiring the Z mass to be between 80 and 100 GeV. Under the peak, the $t\bar{t}$ and W+jets events can be estimated fitting the side-bands with a linear function (shown in figure). By doing so we get the following numbers corresponding to 1 fb^{-1} of integrated luminosity: the total number of events under the peak is 136; the estimated number of $t\bar{t}$ and W+jets events under the peak (from the linear fit to the side bands) is 25.7, giving the remaining 110.3 events coming from $Zb\bar{b}$ and Z+jets. From the $Zb\bar{b}$ MC the number of events between 80 and 100 GeV are 80.7, while the number of events from Z+jets MC are 32.9. Thus in good agreement with the number found from the pseudo-experiment. As a systematic error we can quote or the difference between the numbers obtained from the pseudo-experiment and the MC ones (i.e. 3%, from 110.3/ (80.7+32.9)) or more conservatively, the difference between the events on the peak from the pseudo-experiment, i.e. 110.3, and the number of $Zb\bar{b}$ events, assigning an 100% error on the Z+jets contribution. In this case a 35% error.

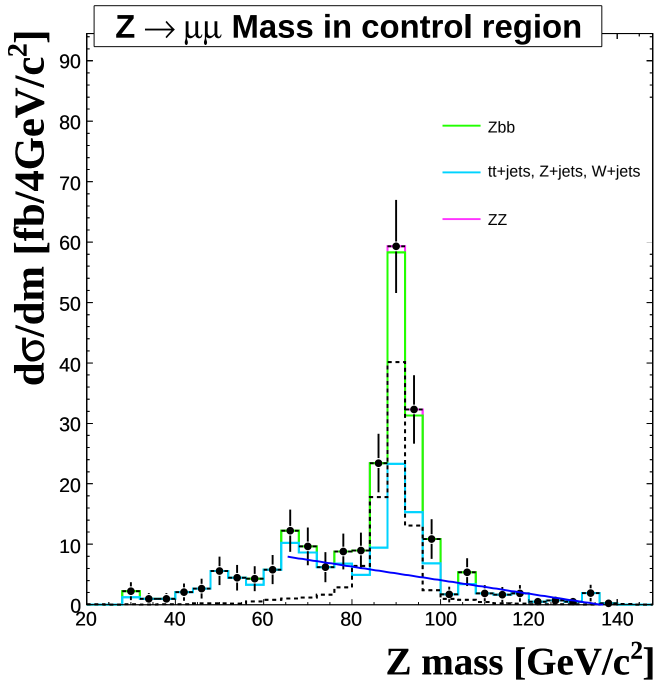


Figure 24: The invariant mass of the first Z candidate for the events in the region defined by $\mathrm{X2least} > 1.5 p_{T,3} - 15$ for a "pseudo-experiment" with a Luminosity of 1 fb^{-1} is shown by the black points. The blue line show the $t\bar{t}$+W/Z+jets, the green and purple line show $Zb\bar{b}$ and ZZ background respectively. The $Zb\bar{b}$ and Z+jets show a real Z resonance decaying into two muons. In $t\bar{t}$ and W+jets instead the two muons do not resonate. The dotted blue line is the linear fit to the side-bands around the Z peak. The black dashed line is the prediction from the $Zb\bar{b}$ MC. The uncertainty is derived by comparing the integral under the Z peak of the dashed black line, with the difference between the integral under the peak of the black points and the dotted blue line.

The systematics on the measurement of the isolation variable will be determined by the *tag and probe* method using high statistic $Z \to \ell\ell$ events and also $pp \to b\bar{b}$ events (e.q. see [**?**])

7.2.3 Control of QCD fake rates from data

If the events we try observe are indeed originate from ZZ (major irreducible background) and other background contributions are virtually zero, then if one tries to repeat the analysis for four lepton pairs, but not sorting them by a flavor or a charge, then no new additional events will be found. If the observed events (if any), on the other hand, are largely due to unaccounted backgrounds (such as $Zjj \to l^+l^- + 2$ "fake leptons" or $QCD \to 4$ "fake leptons") then, in a striking contrast, many more events will appear. For illustration purposes, Tab. 6 gives estimations for a few cases when "fake" leptons produced with different levels of flavor- and charge-correlations and assuming equal rates per flavor (e.g., probabilities per jet). One can easily extend this table by throwing in various levels of possible correlations.

Note the apparent danger of the $Z\gamma \to 2l + (e^+e^-)_{conv}$ and $Z\pi^0 \to 2l + (e^+e^-)_{conv}$ backgrounds that cannot be identified this way. However, the danger is only apparent: the probability of producing a conversion (e^+e^-)-pair with an invariant mass of 90 GeV is extremely small.

Table 6: Estimates for relative number of events with correct and incorrect $(4l)$-combinations for different $4l$-event sources. $(4l)_{OK}$ stands for $2\mu^+2\mu^-$, $\mu^+\mu^-e^+e^-$, $2e^+2e^-$. $(4l)_{BAD}$ stands for all other combinations of charges and flavors.

Event source	$(4l)_{OK}$-events	$(4l)_{BAD}$-events
$ZZ \to 4l$	N	0
$Zjj \to 4l$	N	$3N$
$QCD \to 4l$	N	$\frac{220}{36}N \sim 6.1\,N$
$t\bar{t} \to 4l$	N	$\frac{220}{36}N \sim 6.1\,N$
$bbbb \to 4l$	N	$\frac{220}{36}N \sim 6.1\,N$
$bb \to 4l$ (no B-oscillations)	N	$\frac{5}{3}N \sim 1.7\,N$
$Z\gamma \to 2l + (e^+e^-)_{conv}$	N	0

In addition, there are other handles for evaluating potential contributions of various backgrounds. For example, the Zjj cross section can be fairly well measured in the channel $Zjj \to 2l + jj$, $\sigma(Zjj)$. The probability of a jet faking a muon can be measured from QCD di-jet data stream, $\epsilon(j \to \mu)$. The two measurements can be further combined to prove that the estimated contribution of the Zjj background, $\sigma(Zjj)\epsilon(j \to \mu)^2$ is indeed negligible. The $Zb\bar{b}$ background can be treated similarly to check that its expected contribution is of the order $\sim$0.02 events.

8 Results

The lepton isolation observables, the lepton impact parameter observables, the p_T of each of the four leptons, the two-lepton invariant masses and the four-lepton invariant mass $m_{4\ell}$ can be combined to optimize the sensitivity to the Higgs boson as a function of the mass hypothesis m_{H}, and for a given integrated luminosity. Such mass dependent cut-based analyses have been discussed in previous studies [6, 7, 8] in the context of measurements at integrated luminosities of 30 fb^{-1} at the LHC. For the start-up integrated luminosity of 1 fb^{-1} considered in this analysis, and given an improved suppression of the distinguishable background sources, it is found sufficient to consider a baseline cut-based selection optimized around a central value of $m_{\mathrm{H}} \simeq 150\,\mathrm{GeV/c^2}$, leaving only a sliding window cut in the measured $m_{4\ell}$ spectrum to optimize the sensitivity for a Higgs boson of given mass m_{H}. This allows for a simple search procedure covering the mass range from $\gtrsim 130\,\mathrm{GeV/c^2}$ to $\lesssim 250\,\mathrm{GeV/c^2}$.

8.1 Baseline event selection

Looking at rejection power of discriminating variables the set of baseline cuts have been chosen for low mass Higgs searches for $4e$ and 4μ channels, with $2e2\mu$ channel mixing the best of two. Set of baseline selection cuts for all three channels is given in Table 7.

	Channel		
	$4e$	$2e2\mu$	4μ
Isolation	$eISO_{2least} < 0.35$ $eISO_{2least} < 0.06 \cdot p_T^3 - 0.9$ $eISO_{2least} < 0.035 \cdot p_T^4 - 0.2$	$eISO < 0.30$, both e $\mu ISO < 15$, both μ $\mu ISO < 15$, both μ	$\mu ISO_{2least} < 30$ $\mu ISO_{2least} < 1.5 \cdot p_T^3 - 15$ $\mu ISO_{2least} < 2 \cdot p_T^4 - 10$
IP	$eIP < 11$	$\mu IP_{max} < 12$ & $\mu IP_{2^{nd}max} < 4$	
Lepton p_T	$p_T^{min} > 7$ GeV/c	p_T^4 (p_T^3) > 7 (15) GeV/c	$p_T^{min} > 5$ GeV/c
M_Z	[50 GeV/c^2 , 100 GeV/c^2]		
$M_{Z^\star}$	[20 GeV/c^2 , 100 GeV/c^2]		

Table 7: Set of baseline selection cuts for all three channels.

Four lepton invariant mass after baseline selection is given in Fig. 25. The $t\bar{t}$+jets background is completely eliminated. The Z$b\bar{b}$ background is considerably reduced and now only survives towards low masses, with an event rate far below that of the ZZ$^{(*)}$ continuum. The signal from a SM Higgs boson is observed as a narrow peak with a mean expected number of events emerging above the continuum over the full mass range.

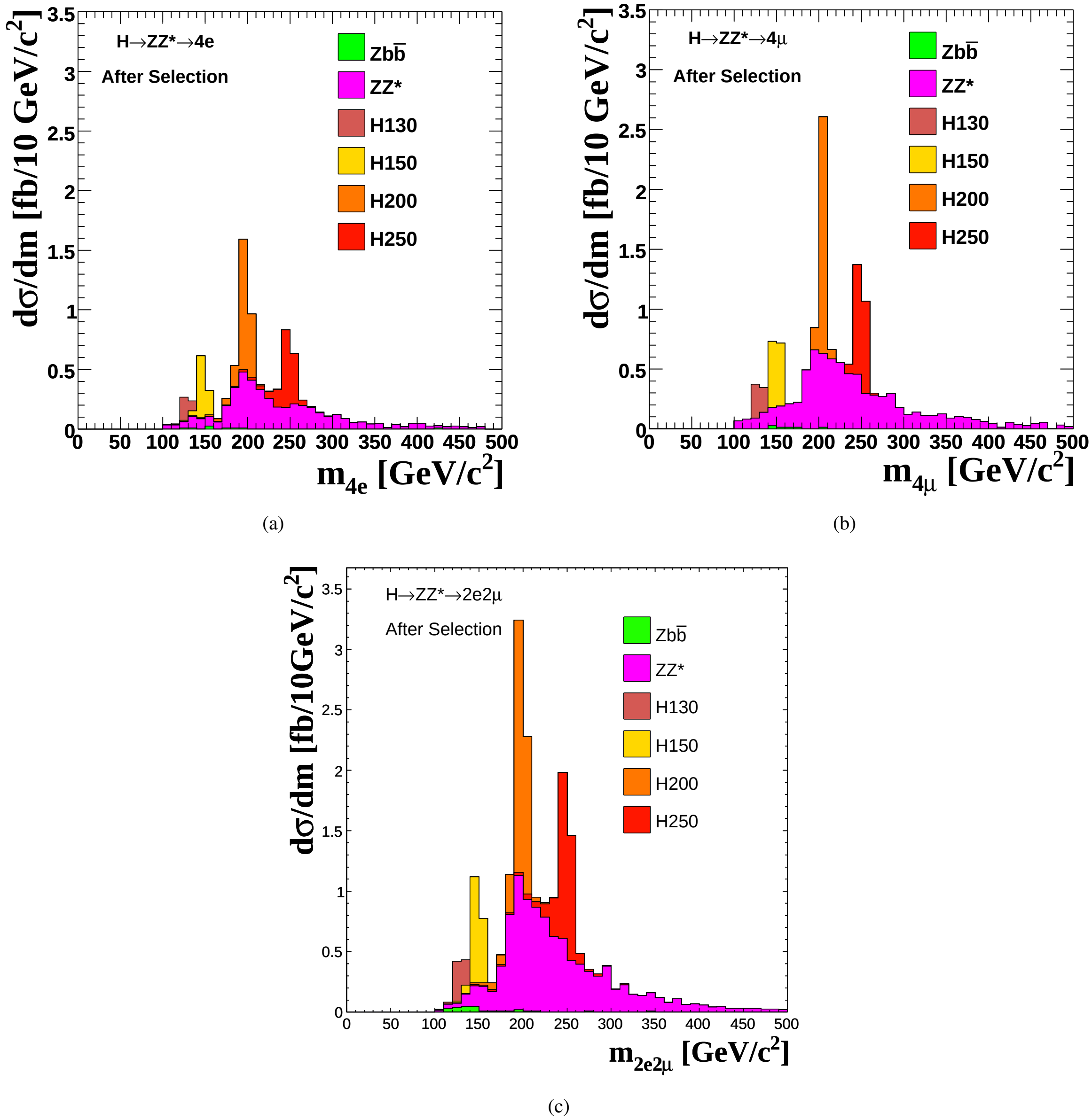


Figure 25: Four lepton invariant mass after baseline selection for (a) $4e$, (b) 4μ and (c) $2e2\mu$.

M_H (GeV/c^2)	Events at 1 fb^{-1}		Significance	95 % C.L. for R
	Signal	Bckgd		
130	1.17	0.23	1.31	3.02
140	2.24	0.43	2.05	1.69
150	2.87	0.46	2.51	1.34
160	1.44	0.38	1.40	2.59
170	0.69	0.54	0.48	5.64
180	1.71	1.30	1.07	2.76
190	6.22	2.59	2.86	0.91
200	6.76	3.09	2.89	0.89
250	5.14	2.56	2.43	1.10

Table 8: Expected significance and expected values of R at 95% confidence level for four selected Higgs boson masses.

8.2 Combined analysis results

In order to quantify the sensitivity of the experiment to the presence of a Higgs boson signal a simple counting experiment is used. The expected number of signal (N_s) and background (N_b) events are evaluated in a $m_{\mathrm{H}} \pm 2\sigma_m$ mass window around selected masses and are given in Table 8.

The counting experiment significance S_{cP} is defined as the probability from a Poisson distribution with mean N_b to observe a number of events equal or greater than $N_s + N_b$, converted in equivalent number of sigmas of a Gaussian distribution. The mean sensitivity expected for a combination of the three channels is given in Table 8 for various m_{H} hypothesis and considering $m_{\mathrm{H}} \pm 2\sigma_m$ mass windows. The expected mean significance for the signal observation reaches values above two standard deviations (2 σ) in a mass range from 140 to 155 GeV/c^2, and around 3 σ for mass hypotheses above 185 GeV/c^2.

The significances for the interpretation of the observation of an excess of signal-like events are smaller by about 1σ unit when considering the probability for a random fluctuation of the background anywhere in the mass range of this analysis. Thus, it is unlikely that an integrated luminosity of 1 fb^{-1} would yield an observation of a mass peak with an overall significance well above 2 σ. In absence of a significant deviation from SM expectations, an upper limit on the cross-section for the production of a SM-like Higgs boson can be derived. The results are given in Table 8 for various m_{H} hypothesis. and expressed in terms of the ratio of excluded and Standard Model cross sections $R_{95\%C.L.} = \sigma_{95\%C.L.}/\sigma_{SM}$. These results are represented in Fig. 28. A SM-like Higgs could be excluded at 95%C.L. for masses above 185 GeV/c^2.

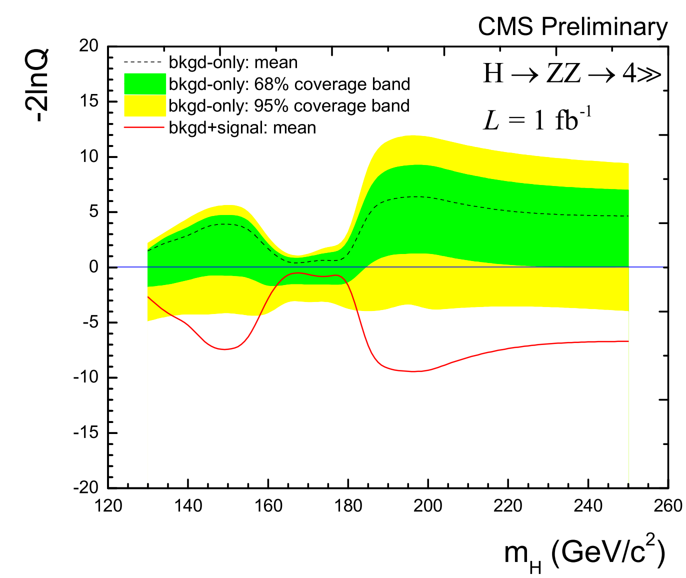


Figure 26: ...

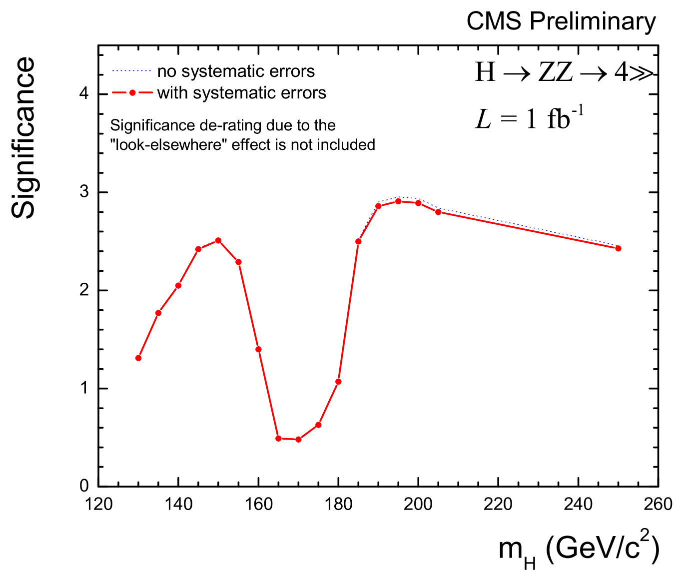


Figure 27: ...

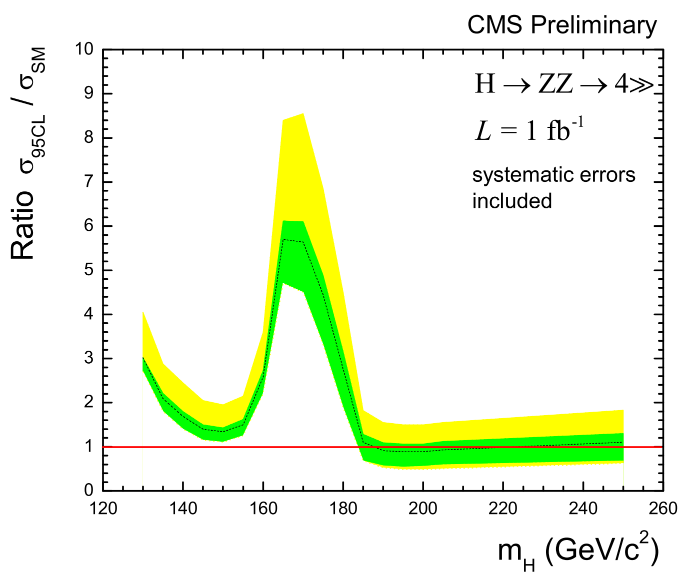


Figure 28: Excluded cross-sections for a SM-like Higgs boson, in absence of a significant deviation from background expectations for an integrated luminosity of 1 fb^{-1} at the LHC, shown as a function of the m_{H} hypothesis, and normalized to the mean expectation from the SM for that given mass. The dotted line is the mean expected 95%C.L exclusion curve obtainable in a single experiment. The green band correspond to $\pm 1\sigma$ uncertainties. The yellow band correspond to $\pm 2\sigma$ uncertainties.

9 Conclusions

A analysis prospective for the search of Standard Model-like Higgs boson decaying in ZZ* pairs was presented for the CMS experiment in the context of the startup luminosity of 10^{32} cm^{-2}s^{-1} at the CERN LHC pp collider. The analysis was performed for the leptonic decay channels $Z \to ll$ with $l = e, \mu$, assuming an integrated luminosity of 1 $\mathrm{fb}^{-1}$and the detector calibration and alignment knowledge of the first 100 pb^{-1}. A complete strategy has been established.

A combination of electron and muon trigger-paths and a loose data reduction skimming step are combined at early stages of the analysis to preserve efficiency for the selection of signal events. The requirement of the presence of at least two pairs of identified leptons, with opposite charges and matching flavours and satisfying minimal invariant mass constraints and loose isolation requirements, suppresses the fake rates from QCD multi-jet and Z/W+jet(s) at the level of, or below, $t\bar{t}$, Z$b\bar{b}$ and the $\mathrm{Z}/\gamma^*\mathrm{Z}/\gamma^*$ continuum.

The $\mathrm{Z}/\gamma^*\mathrm{Z}/\gamma^*$ continuum is indistinguishable from the Higgs boson signal on event-by-event. the next step of the analysis focuses on the suppression of the distinguishable $t\bar{t}$ and Z$b\bar{b}$ backgrounds. This exploits the presence of two b-quark jets which have provided secondary leptons (e.g. from heavy flavoured meson decays) misidentified as primary leptons. These fake primary leptons are suppressed by tight lepton isolation and vertex matching constraints.

The lepton isolation criteria is most powerful on average for hard b-quark jets where the fragmentation and decays products are more collimated. Loose isolation requirements considerably reduce the $t\bar{t}$ contamination. Tighter isolation cuts are necessary for soft b-quark jets. A best strategy was found to be to tighten (loosen) the lepton isolation requirement with increasing (decreasing) p_T of the third lepton, belonging to a matching pair, in the event. The loosening of the lepton isolation for harder b-quark jets is compensated by tight vertex matching constraints.

The Higgs boson search is performed with a sliding window in the hypothetical mass m_H, and using a simple sequential cut-based approach. In absence of a significant signal observed from the combination of the $4e$, 4μ and $2e2\mu$ analyses, 95% confidence level exclusions limits are obtained for Standard Model-like Higgs boson.

10 Acknowledgments

We wish to thank the CMS production team for providing us with the Monte Carlo samples used in this analysis. Thanks to our colleagues of the CMS Collaboration for the precious support.

References

[1] R. Barate *et al.* [LEP Working Group for Higgs boson searches], Phys. Lett. B **565** (2003) 61 [arXiv:hep-ex/0306033].

[2] G. Hesketh [CDF and D0 Collaborations], Nucl. Phys. Proc. Suppl. **177-178**, 219 (2008); The TEVNPH Working Group [CDF and D0 Collaborations], Fermilab-Pub-08-270-E, e-print arXiv:0808.0534v1 (August 2008) 14pp.

[3] S. Leone [CDF and D0 Collaboration], arXiv:0710.4983 [hep-ex].

[4] M. Aldaya et al., "Discovery Potential for the Standard Model Higgs boson in the $\mathrm{H} \to ZZ^{(*)} \to 4\mu$ decay channel", CMS Note 2006/106 (June 2007) 16pp.

[5] M. Aldaya et al., "A method for determining the mass of the Standard Model Higgs boson using the $\mathrm{H} \to ZZ^{(*)} \to 4\mu$ decay channel", CMS Note 2006/107 (June 2006) 13pp.

[6] S. Baffioni et al., "Discovery Potential for the Standard Model Higgs boson in the $\mathrm{H} \to ZZ^{(*)} \to e^+e^-e^+e^-$ decay channel", CMS Note 2006/115 (June 2006) 26pp.

[7] S. Abdullin et al., "Search Strategy for the Standard Model Higgs Boson in the $\mathrm{H} \to ZZ^{(*)} \to 4\mu$ decay channel" CMS Note 2006/122 (June 2006) 26pp.

[8] D. Futyan et al., "Search for the Standard Model Higgs Boson in the Two-Electron and Two-Muon Final State with CMS", CMS Note 2006/106 (July 2006) 32 pp.

[9] CMS Collaboration, "The Compact Muon Solenoid - Technical Proposal", CERN/LHCC 94-38 (December 1994) 290pp..

[10] The CMS Collaboration, "The CMS experiment at the CERN LHC", Journal of Instrumentation (August 2008) 361 pp.

[11] CMS Collaboration, "The Tracker Project - Technical Design Report", CERN/LHCC 98-6 (April 1998) 625pp.; *idem*, "Addendum to the Tracker TDR", CERN/LHCC 2000-016 (Feb 2000) 82pp.

[12] CMS Collaboration, "The Electromagnetic Calorimeter Project", Technical Design Report, CERN/LHCC 97-23 (15 December 1998).

[13] CMS Collaboration, "CMS Muon Technical Design Report", CERN/LHCC 97-32 (15 December 1997).

[14] CMSSW Full Simulation Offline Guide, https://twiki.cern.ch/twiki/bin/view/CMS/SWGuideSimulation .

[15] CMSSW Reconstruction Offline Guide, https://twiki.cern.ch/twiki/bin/view/CMS/SWGuideReco .

[16] T. Sjostrand, S. Mrenna and P. Skands,"PYTHIA physics and manual", JHEP **0605**, (2006) 026; *(PYTHIA version 6.225)*

[17] D. Acosta et al., "The Underlying Event at the LHC", CMS Note 2006/67 (June 2006) 22pp.

[18] M. Spira, Tables of Standard Model Higgs NLO cross-sections and Branching Ratios; https://twiki.cern.ch/bin/view/CMS/SMHiggsBoson .

[19] W. M. Yao *et al.* [Particle Data Group], J. Phys. G **33** (2006) 1.

[20] M. L. Mangano, M. Moretti, F. Piccinini, R. Pittau and A. D. Polosa, "ALPGEN, a generator for hard multiparton processes in hadronic collisions", JHEP **0307**, 001 (2003), hep-ph/0206293.

[21] F. Maltoni, "Theoretical Issues and Aims at the Tevatron and LHC", proceeding of Hadron Collider Physics Symposium, Les Diableretes, Switzerland, July 2005, printed version: Springer Proceedings in Physics, Vol. 108, 2006, ISBN: 3-540-32840-8.

[22] S. R. Slabospitsky and L. Sonnenschein, Comput. Phys. Commun. **148** (2002) 87 [arXiv:hep-ph/0201292].

[23] A Pukhov et al. , ”CompHEP”, e-print hep-ph/9908288 (August 1999) 126pp.; *(CompHEP version 4.2p1)*.

[24] J. Campbell and K. Ellis, ”MCFM - Monte Carlo for FeMtobarn processes”, http://mcfm.fnal.gov/ *(MCFM version 5.2)*

[25] S. Abdullin et al., “Relative Contributions of the t- and s-Channels to the $\mathrm{ZZ} \to 4\mu$ process”, CMS NOTE 2006/057 (May 2006) 7pp.

[26] C. Zecher, T. Matsuura, and J.J. van der Bij, ”Leptonic Signals from off-shell Z Boson Pairs at Hadron Colliders”, hep-ph/9404295.

[27] P. Bartalini et al., NLO vs. LO: kinematical differences for signal and background in the $H \to \mathrm{ZZ}^* \to 4\mu$”, CMS NOTE 2006/130 (June 2006) 29 pp.

[28] https://twiki.cern.ch/twiki/pub/CMS/TSG_24_IX_07/TriggerMenu160.pdf

[29] S. Baffioni et al., “Electron reconstruction in CMS”, CMS Note2006-40 (February 2006) 25pp.

[30] G. Daskalakis et al., “Measuring Electron Efficiencies at CMS with Early Data”, CMS AN 2007-019 (November 2007) 43pp.

[31] N. Adam et al., “Towards a Measurement of the Inclusive W–¿ev and Z–¿ee Cross Section in pp Collisions at sqrt(s) = 14 TeV”, CMS AN 2007-026 (July 2008) 67pp.

[32] P. Meridiani and R. Paramatti, “Use of $Z \to e^+e^-$ events for ECAL calibration”, CMS Note 2006-039 (January 2006) 18pp.

[33] R.Bellan, Nucl.Phys.Proc.Suppl.177-178 (2008) 253-254.

[34] N. de Filippis et al., “Update of Misalignment Scenarios for the CMS Tracker” CMS IN-2007/036 (July 2007) 15pp.

[35] N. de Filippis et al., “*Misalignment Scenarios for the Startup Conditions of the CMS Tracker*” CMS IN-2007/061 (October 2007) 11pp.

[36] N. Amapane et al., High Level Trigger Algorithms for Muon Isolation”, CMS Note 2002/040 (Nonvember 2002) 34pp.

[37] D. Acosta et al., Measuring Muon Reconstruction Efficiency From Data”, CMS Note 2006/60 (May 2006) 6pp.

[38] S. Abdullin et al., Sensitivity of the Muon Isolation Cut Efficiency to the Underlying Event Uncertainties”, CMS Note 2006/033 (February 2006) 11pp.

[39] CMS PTDR v.1

[40] CMS PAS EWK-07-002.

[41] CMS NOTE-2006/122, CMS CR-2007/042

11 Appendix

11.1 Bremsstrahlung Recovery

Inner or internal bremsstrahlung (IB) photons in the final state are defined as those radiated from one of the 4 leptons at the primary vertex. Figure shows the number of IB photons per signal event (m$_H$=195 GeV) at generator-level. At least one IB photon with $p_{T\gamma} > 1$GeV (value required to make an electromagnetic calorimeter supercluster at the reconstructed level) is present in nearly 25% of the events and approximately 10% of events have at least two such IB photons. For the PTDR, a maximum of 1 IB photon per event was recovered (the closest inside a cone of ΔR(lept,γ)<0.3 around the lepton). This technique was found to increase the efficiency of the Z mass window criteria used by 1-2%, depending on m$_H$.

A new "multi-brem" recovery technique has been developed, which can recover a maximum of 3 IB photons per event (and a maximum of 2 per lepton). The recovered photons are those with $p_{T\gamma} > 3$ GeV, in order of increasing ΔR(lept,γ), inside a cone of ΔR(lept,γ)<0.25. The choice of cone size is motivated by Figure 29, which shows the value of ΔR between generated internal bremsstrahlung photons with $p_T > 5$ GeV and their parent lepton, as well as that between other photons from the primary vertex with $p_T > 5$ GeV and the nearest lepton.

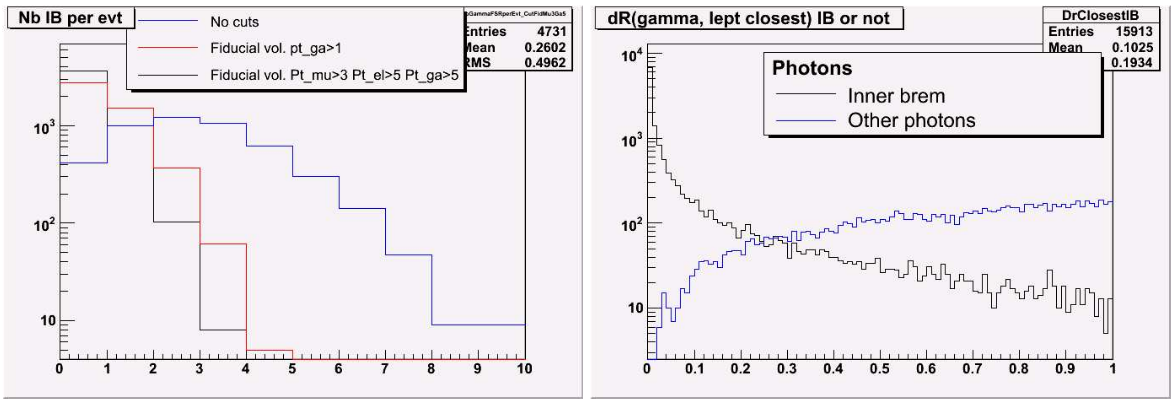
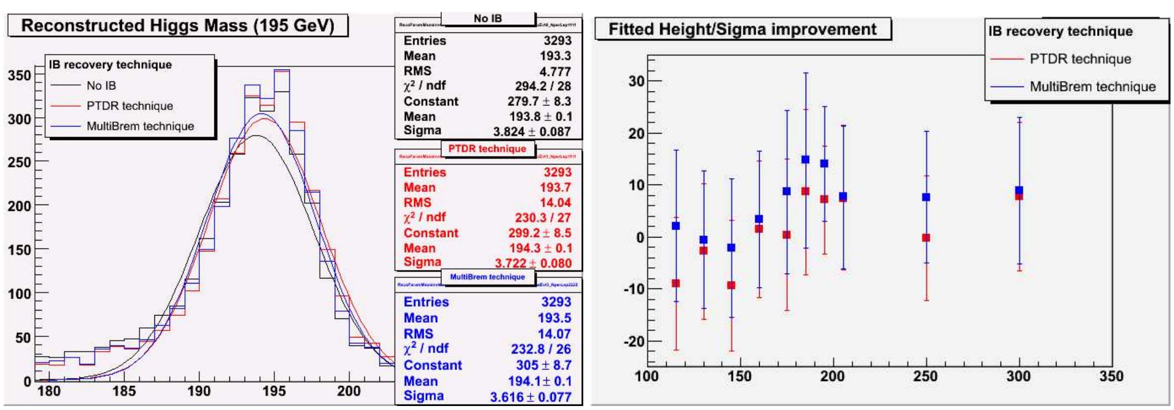


Figure 29: Number of inner bremsstrahlung per event for different sets of cuts (left).ΔR between generated internal bremsstrahlung photons with $p_T > 5$ GeV and their parent lepton (black line), and between other photons from the primary vertex with $p_T > 5$ GeV and the nearest lepton (blue line)for a H $\rightarrow$ ZZ* $\rightarrow$ 2e2μ skimmed signal sample (mH=195 GeV) (right).



(a) (b)

Figure 30: (a) Higgs reconstructed invariant masses in the 2e2μ (center) channels, for no IB recovery, the PTDR and multi-brem techniques. (b) Improvement in fitted height over sigma relative to the unrecovered case, for 10 Higgs boson masses. H $\rightarrow$ ZZ* $\rightarrow$ 2e2μ skimmed samples.

The performance of the multi-brem vs.the PTDR technique for 2e2μ channel is demonstrated in figures 30(a) for m$_H$=195 GeV. The reconstructed invariant 4-lepton mass is shown before and after IB recovery via the two techniques and the distributions are gaussian-fitted. The figure of merit chosen is the improvement of the ratio of the fitted peak height to the fitted sigma with respect to the unrecovered case. This improvement is equal to 24% (14.9%) for the multi-brem (PTDR) technique in the 4μ channel and 15,3% (9,9%) in the $2e2\mu$ channel. In the 4e channel both IB recovery techniques give slightly worse results than in the unrecovered case.

Figure 30 (b) shows the performance of the two IB recovery techniques on signal samples corresponding to Higgs boson masses between 115 GeV and 300 GeV, for the $2e2\mu$ channel. The multi-brem algorithm in all cases performs better than the PTDR algorithm; the improvement obtained is everywhere positive, with the most significant improvement corresponding to Higgs boson masses above 150 GeV.

In order to estimate the performance of IB recovery techniques in a low statistics, startup scenario, a series of toy Monte Carlo or 'gedanken-' experiments has been performed. Preselected signal samples (exclusive of lepton isolation requirements) in the $2e2\mu$ channel corresponding to mH=120 and 195 GeV have been divided into 150 (109) independant subsamples of 30 events each. This corresponds to an integrated luminosity for the signal of 16,7 fb^{-1} for mH=120 GeV and 2,5 fb^{-1} for mH=195 GeV. Results of the gaussian fits of the distributions of the reconstructed 4-lepton mass, the fitted sigma, fitted height and the ratio fitted sigma/height show that even in the low-statistics scenario, the effects are small (near 5% for mH=195 GeV and 1% for mH=120 GeV) but significant. The multi-brem technique gives better results than the PTDR technique for the two Higs boson masses tested.