Chapter 2

Physics Studies with Photons and Electrons

2.1 Benchmark Channel: $H \rightarrow \gamma\gamma$

The $H \rightarrow \gamma\gamma$ channel has been studied since the initial planning of the LHC and SSC as an important channel for the discovery of Higgs particles at masses beyond the upper reach of LEP and below about 150 GeV [3, 15, 16]. The signature sought in the inclusive analysis is two high E_T isolated photons. The challenge for discovery of a Higgs in this mode is the small branching fraction of about 0.002, since in this mass range the dominant decay mode of the Higgs is $b\bar{b}$. The $\gamma\gamma$ decay mode can be well identified experimentally but the signal rate is small compared to the backgrounds coming both from two prompt photons (irreducible), and from those in which one or more of the photons are due to decay products or mis-identified particles in jets (reducible). It has long been understood that $H \rightarrow \gamma\gamma$ can be detected as a narrow mass peak above a large background. The background magnitude can be determined from the region outside the peak. After event selection, for an integrated luminosity of 20 fb^{-1} and for a Higgs boson mass of 120 GeV/c^2, we expect approximately 350 signal events in a mass window of 2 GeV/c^2 over 7000 background events.

In this study we present two complementary inclusive analyses for the $H \rightarrow \gamma\gamma$ channel: a standard cut based analysis and a high performance, discovery-oriented analysis, based on the method described in [17, 18]. Both are carried out with our present knowledge of the expected background, estimated with full detector simulation. Further details can be found in [19]. The study concentrates on the first years of LHC operation and uses simulated events with pileup corresponding to a luminosity of 2×10^{33} $cm^{-2}sec^{-1}$.

The idea of measuring the rate of background by using the mass regions adjoining the Higgs peak is extended to also measure the characteristics of the background, and using this information to help separate background from signal. The $H \rightarrow \gamma\gamma$ channel is particularly well suited to this technique because the signal is relatively small and can be confined to a narrow mass region thanks to the excellent photon energy and position resolution of the CMS detector [7].

By using photon isolation and photon kinematic information, significant additional discrimination between signal and background can be achieved. The optimised analysis uses this information to discriminate between signal and background by comparing data in mass sidebands with signal Monte Carlo. Use is made of a neural network, but likelihood variables or other techniques may prove to be better in the future. The expected purity in terms of signal/background, corresponding to each event, can be estimated based on this information and each event then can be used optimally to evaluate the likelihood of a signal plus background hypothesis compared to a background-only hypothesis.

In the optimised analysis the expected signal to background ratio is calculated for each event. By dividing the cut-based analysis in various categories with different s/b ratios results improve toward those that are obtained with the optimised analysis. If the maximum s/b ratio in the optimised analysis is limited to the best category used in the cut-based analysis, the performances of the two analyses are nearly identical.

The optimised, discovery-oriented analysis is particularly appropriate to the $\mathrm{H} \rightarrow \gamma\gamma$ channel because the Higgs signal appears in a narrow mass peak allowing analysis of the large background in the mass side-bands. The analysis will not be limited by the poor simulation of the background once data will be available.

The study described requires a comprehensive understanding and simulation of the CMS detector. The electromagnetic calorimeter is used to make the primary measurements of photon energy and position. The tracker is used to measure the position of the interaction vertex. The tracker, ECAL and HCAL are used to determine if the photon candidate is well isolated. While background characteristics will be measured from data, the signal must be well simulated to perform the analysis described below. This requires a detailed understanding of the detector performance as well as its calibration.

2.1.1 Higgs boson production and decay

For this inclusive study the Higgs boson production mechanisms with the largest cross-sections in the Standard Model have been simulated: gluon fusion, qqH production through Weak Vector Boson Fusion (WBF), associated Higgs production with W or Z bosons, and Higgs production associated with a $t\bar{t}$ pair. The cross sections for the different production processes [20] and the $\mathrm{H} \rightarrow \gamma\gamma$ branching ratios [21] are summarised in Table 2.1. The analysis described in this chapter has been limited to careful measurement of the inclusive diphoton channel, to address the main detector issues, and no use has been made of tagging leptons or jets. In the future, channel identification, based on additional leptons and jets. will improve the sensitivity. For the moment these 'tagged' channels are investigated individually in other studies [22, 23]. Figure 2.1 shows an event display of a $\mathrm{H} \rightarrow \gamma\gamma$ event with $\mathrm{M_H}$=120 GeV/c^2.

Table 2.1: Next to Leading Order cross sections for the different Higgs boson production processes and $\mathrm{H} \rightarrow \gamma\gamma$ branching ratios.

$\mathrm{M_H}$	115 GeV/c^2	120 GeV/c^2	130 GeV/c^2	140 GeV/c^2	150 GeV/c^2
σ (gg fusion)	39.2 pb	36.4 pb	31.6 pb	27.7 pb	24.5
σ (WVB fusion)	4.7 pb	4.5 pb	4.1 pb	3.8 pb	3.6
σ (WH, ZH, $t\bar{t}$H)	3.8 pb	3.3 pb	2.6 pb	2.1 pb	1.7
Total σ	47.6 pb	44.2 pb	38.3 pb	33.6 pb	29.7
$\mathrm{H} \rightarrow \gamma\gamma$ Branching Ratio	0.00208	0.00220	0.00224	0.00195	0.00140
Inclusive $\sigma \times B.R.$	99.3 fb	97.5 fb	86.0 fb	65.5 fb	41.5 fb

2.1.2 Backgrounds

Backgrounds with two real prompt high E_{T} photons are called "irreducible", although they can be somewhat reduced due to kinematic differences from signal processes in which high

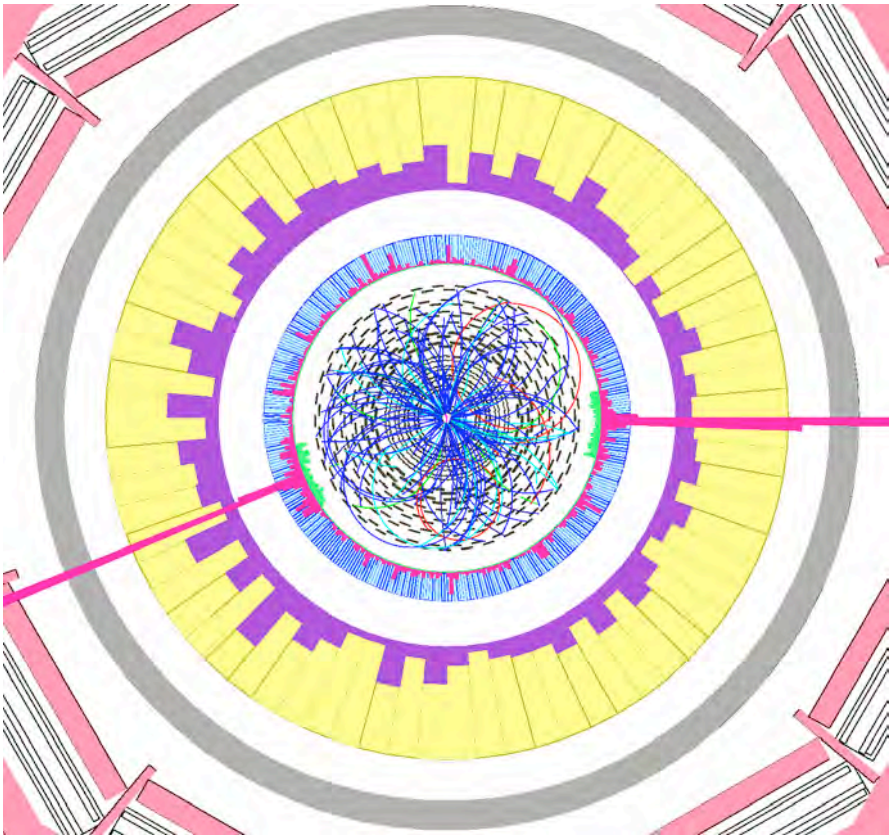

Figure 2.1: $\mathrm{H} \to \gamma\gamma$ event produced in gluon fusion with $\mathrm{M_H}$=120 GeV observed in the CMS detector.

mass particles are produced. Two photons can be produced from two gluons in the initial state through a "box diagram" or from initial quark and anti-quark annihilation.

Backgrounds in which at least one final state jet is interpreted as a photon are called "reducible" and are much harder to simulate since jets are copiously produced at the LHC and Monte Carlo samples that correspond to 10 fb^{-1} are much too large to fully simulate. Selections at generator level have been devised in order to be able to select multi-jet and γ plus jets events that contribute to the background of the $\mathrm{H} \to \gamma\gamma$ channel and reject events that have negligible chance of producing background to the final analysis.

The $\gamma + \mathrm{jet}$ sample can be viewed, from the selection point of view, as coming from two different sources: one where another photon is radiated during the fragmentation of the jet (two prompt photons), the other where there is only one prompt photon in the final state and the other photon candidate corresponds to a mis-identified jet or isolated π^0 (one prompt plus one fake photon). These two processes have been separated using generator level information, and are listed separately in the tables below. Also, different K-factors are applied.

The generator level pre-selection of $\gamma + \mathrm{jet}$ events that contribute to the $\mathrm{H} \to \gamma\gamma$ background is straightforward. For $\mathrm{pp} \to \mathrm{jets}$, a much tighter set of cuts at the particle generator level was carefully developed and studied. Groups of particles, protocandidates, which might form a photon candidate after event simulation are identified. Cuts are applied on the transverse energy of two protocandidates and on their invariant mass, and this involves an estimate on the lower and upper limits to the energy of the photon candidates that might be recon-

structed from the protocandidates after the simulation. An estimate is also made on likely level of isolation of the resulting photon candidate.

With such selection a rejection of a factor of about 41000 can be obtained, with an estimated inefficiency of 14% for $\mathrm{pp} \rightarrow \mathrm{jets}$ events generated with PYTHIA with $\hat{p}_\perp > 30$ GeV (transverse momentum of the products of the hard interaction). The inefficiency after the final analysis selection was estimated by using a looser pre-selection similar to that used for the $\mathrm{pp} \rightarrow \gamma + \mathrm{jet}$ simulation. Further details can be found in [19]. Events rejected by the pre-selection have rather low E_T photons and are not very important for the final analysis.

The Monte Carlo samples used are summarised in Table 2.2. All events were generated with PYTHIA [24], simulated with the GEANT-based [9] CMSIM [25] or OSCAR [8], and reconstructed with ORCA version 8.7.3 [10]. Pile-up events from minimum bias interactions were added to the hard interaction, assuming a luminosity of $\mathcal{L} = 2 \times 10^{33}\,\mathrm{cm}^{-2}\,\mathrm{s}^{-1}$.

K-factors are applied to take into account the expected differences between the lowest order cross sections given by PYTHIA and the NLO cross sections of the different background processes [26–30]. The K-factors used for each background are summarised in Table 2.3 and are estimated to have an uncertainty of 20-30%.

Table 2.2: Monte Carlo samples used in the $\mathrm{H} \rightarrow \gamma\gamma$ analysis with LO cross section from PYTHIA and total corresponding integrated luminosities of the analysed samples.

Process	$\hat{p}_\perp$ (GeV/c)	$\mathrm{M_H}$ ($\mathrm{GeV/c^2}$)	σ (pb)	Pre-sel. σ (pb)	Events Analysed	Int Lum. ($\mathrm{fb^{-1}}$)
$\mathrm{H} \rightarrow \gamma\gamma$ (gg fusion)	-	120	-	-	181K	-
$\mathrm{H} \rightarrow \gamma\gamma$ (WB fusion)	-	120	-	-	193K	-
$\mathrm{H} \rightarrow \gamma\gamma$ (gg fusion)	-	115–150	-	-	20K	-
$\mathrm{H} \rightarrow \gamma\gamma$ (WB fusion)	-	115–150	-	-	20K	-
$\mathrm{H} \rightarrow \gamma\gamma$ (WH,ZH,ttH)	-	115–150	-	-	20K	-
$\mathrm{pp} \rightarrow \gamma\gamma$ (born)	> 25	-	82	44	920K	30
$\mathrm{pp} \rightarrow \gamma\gamma$ (box)	> 25	-	82	31	668K	20
$\mathrm{pp} \rightarrow \gamma + \mathrm{jet}$	> 30	-	5×10^4	2.5×10^3	5.5M	2.2
$\mathrm{pp} \rightarrow \mathrm{jets}$	> 50	-	2.8×10^7	4.7×10^3	4.5M	1.0
Drell-Yan ee	-	-	4×10^3	4×10^3	460K	0.1

Table 2.3: Background K-factors applied to PYTHIA cross sections.

$\mathrm{pp} \rightarrow \gamma\gamma$ (born)	1.5
$\mathrm{pp} \rightarrow \gamma\gamma$ (box)	1.2
$\mathrm{pp} \rightarrow \gamma + \mathrm{jet}$ (2 prompt)	1.72
$\mathrm{pp} \rightarrow \gamma + \mathrm{jet}$ (1 prompt+ 1 fake)	1
$\mathrm{pp} \rightarrow \mathrm{jets}$	1

2.1.3 Reconstruction, selection, and signal significance calculation

2.1.3.1 Trigger

$\mathrm{H} \to \gamma\gamma$ events are selected with extremely high efficiency both by the Level-1 and High Level triggers that are described in details in reference [31]. Since in the analysis selection tighter E_{T} and isolation cuts are applied, the inefficiency due to the trigger is negligible.

2.1.3.2 Photon reconstruction

Photons are reconstructed with the standard ECAL algorithms [7, 32]. At this level the photon reconstruction efficiency is over 99.5% for photons in the region covered by the ECAL.

The energy resolution of reconstructed photons is excellent for photons that do not convert or that convert late in the tracker. Energy resolution deteriorates somewhat for photons that convert early in the tracker. Nevertheless, the photon energy resolution is substantially less affected by tracker material than is electron energy resolution and the Higgs reconstruction in the calorimeter is quite reliable even for converted photons.

For signal events, where this effect is relevant, the energy response of the individual crystals of the ECAL has been smeared using a miscalibration file randomly generated to correspond to the intercalibration precision expected after calibration with $\mathrm{W} \to \mathrm{e}\nu$ events obtained with an integrated luminosity of $10\,\mathrm{fb}^{-1}$, as described in [7]. The precision is 0.3% in the central part on the barrel, growing up to 1.0% at the edge of the barrel and in the endcaps.

The tools that have been developed to identify and reconstruct photon conversions in the tracker [33], and π^0 rejection tools developed for the endcap silicon preshower detector and the barrel crystals, have not yet been included in the analysis.

2.1.3.3 Primary vertex identification

The bunch length at LHC has an rms width of 75 mm resulting in a longitudinal spread of interaction vertices of 53 mm. If the mean longitudinal position is used (nominal vertex), the invariant mass of a two-photon state, such as the $\mathrm{H} \to \gamma\gamma$, is smeared by about $1.5\,\mathrm{GeV/c^2}$, due to the mis-measurement of the angle between the two photons related to the uncertainty of the photon directions.

The two high E_{T} photons coming from the Higgs boson decay are produced in association with other tracks that may come from the underlying event and initial state gluon radiation or from the other particles produced with the Higgs boson in the case of WBF fusion, WH or ZH production and $t\bar{t}\mathrm{H}$ production.

The charged tracks associated to the Higgs production vertex are typically harder than those coming from minimum bias interactions. Therefore the vertex can be identified by reconstructing the primary vertices in the event and selecting the one that most likely corresponds to the Higgs boson production, based on charged tracks.

At low luminosity (2×10^{33} $\mathrm{cm^{-2}sec^{-1}}$) we are able to identify the correct vertex, defined as being within 5 mm of the actual vertex, in about 81% of the signal events passing the selection described in section 2.1.4.1. Clearly these results will be affected by any significant variation of the characteristics of the pileup events from what is simulated in our pileup samples.

2.1.3.4 Photon isolation

Detailed studies have been made of photon isolation and its optimisation [34, 35]. Fake photon signals due to jets can be rejected by looking for additional energetic particles accompanying the photon candidate. Charged pions and kaons can be detected in the tracker or in the calorimeters. Neutral pions and other particles decaying to photons can be detected in the ECAL. The hadron calorimeter may be important for detecting charged particles not efficiently reconstructed in the tracker, particularly at high η, or other particles like neutrons or $\mathrm{K}^0_{\mathrm{long}}$.

2.1.3.5 Separation into categories based on lateral shower shape and pseudorapidity

The shower shape variable R_9, defined as the fraction of the super-cluster energy found inside the 3×3 array of crystals centred around the highest energy crystal, is effective in distinguishing photon conversions in the material of the tracker. Photon candidates with large values of R_9 either did not convert or converted late in the tracker and have good energy resolution. Photons converting early have lower values of R_9 and worse energy resolution.

The variable R_9 has been shown to be very useful also in discriminating between photons and jets. This occurs both because of the conversion discrimination – either of the photons from a π^0 can convert – and because, looking in a small 3×3 crystal area inside the super-cluster, the R_9 variable can provide very local isolation information about narrow jets.

In the multi-category analysis the events are separated into categories based on R_9 so as to take advantage of better mass resolution where it is expected (the unconverted photons), and yet still use all the events (since the mass resolution varies by at most a factor of 2). This separation also tends to put background events involving jets into categories with lower R_9.

We also find that photons detected in the endcaps have worse energy resolution and higher background than photons detected in the barrel so that it is useful to separate events with one or more photons in the endcaps from those with both photons in the barrel.

2.1.3.6 Calculation of confidence levels

Confidence levels are computed by using the Log Likelihood Ratio frequentist method, as described in [36]. Given the expected signal and background distributions in the final variable (the mass distribution for the cut-based analysis), we simulate many possible outcomes of the experiment by means of Monte Carlo. This is done both in the hypothesis that the signal exists and that it does not exist. To compute a confidence level, we order our trials according to an estimator. This is a single number that is useful to order random trials from most background-only-like to most signal-plus-background-like. The simplest and probably best estimator is the Log Likelihood Ratio (LLR) which compares the likelihood of the data to come from a background-only distribution to the likelihood to come from a signal-plus-background distribution. Each likelihood is the product of probabilities from all the bins. The median confidence level is computed both for discovery and for exclusion.

2.1.3.7 Effect of systematic errors

To include systematic errors the background and signal expectation are randomised by the systematic error during the generation of the random trials, while keeping their expectations at the nominal value. If necessary the correlations between the errors on the different analysis bins is included. It is observed that the signal systematic error has no effect on the median

LLR of signal-plus-background experiments, nor on that of background-only experiments. Of course the distribution corresponding to the signal-plus-background experiments is enlarged by the systematic error on the signal and this makes exclusion more difficult. On the other hand the effect of the systematic error on the background is very large, because of the small signal over background ratio. The mean of the distributions is still unchanged but the widths are enlarged both for background-only experiments and for signal-plus-background experiments. This decreases both the discovery and exclusion sensitivities.

2.1.4 Cut-based analysis

2.1.4.1 Selection

Two photon candidates are required with pseudo-rapidity $|\eta| < 2.5$, with transverse energies larger than 40 GeV and 35 GeV respectively, and satisfying the following isolation requirements:

- no tracks with p_{T} larger than 1.5 GeV/c must be present inside a cone with $\Delta\mathrm{R} < 0.3$ around the photon candidate. We only consider tracks with hits in at least two layers of the silicon pixel detector, therefore converted photons are likely to be rejected only if they convert before the second pixel layer;
- the total E_{T} of all ECAL island basic clusters with $0.06 < \Delta\mathrm{R} < 0.35$ around the direction of the photon candidate, regardless of whether they belong to the supercluster or not must be less than 6 GeV in the barrel and 3 GeV in the endcaps;
- the total transverse energies of HCAL towers within $\Delta\mathrm{R} < 0.3$ around the photon candidate must less than 6 GeV in the barrel and 5 GeV in the endcaps.

In order to further reduce the background that is higher when at least one of the photons is detected in the electromagnetic calorimeter endcaps and to increase the performance of the analysis in the forward region additional isolation requirements are applied for events where one, or more, of the candidates has $|\eta| > 1.4442$. For these events the candidate in the barrel is required to satisfy the tighter isolation selection that is applied to photons in the endcaps: ECAL isolation less than 3 GeV and HCAL isolation less than 5 GeV.

Table 2.4: Expected background after the selection for Higgs boson masses between 115 and 150 GeV/c^2, expressed in fb/GeV

Process	115 GeV/c^2	120 GeV/c^2	130 GeV/c^2	140 GeV/c^2	150 GeV/c^2
$\mathrm{pp} \to \gamma\gamma$ (born)	48	44	36	29	24
$\mathrm{pp} \to \gamma\gamma$ (box)	36	31	23	16	12
$\mathrm{pp} \to \gamma + \mathrm{jet}$ (2 prompt)	43	40	32	26	22
$\mathrm{pp} \to \gamma + \mathrm{jet}$ (prompt+fake)	40	34	22	19	14
$\mathrm{pp} \to \mathrm{jets}$	29	27	20	18	14
Drell-Yan ee	2	2	1	1	1
Total background	203	178	134	109	86

Figure 2.2 shows the mass distribution after the selection. The efficiency for a 120 GeV/c^2 Higgs boson is 30% and the total expected background is 178 fb/GeV. The number of expected background events for the different types of background is shown in Table 2.4 while the Higgs efficiency in different mass windows is shown in Table 2.5. The efficiency is com-

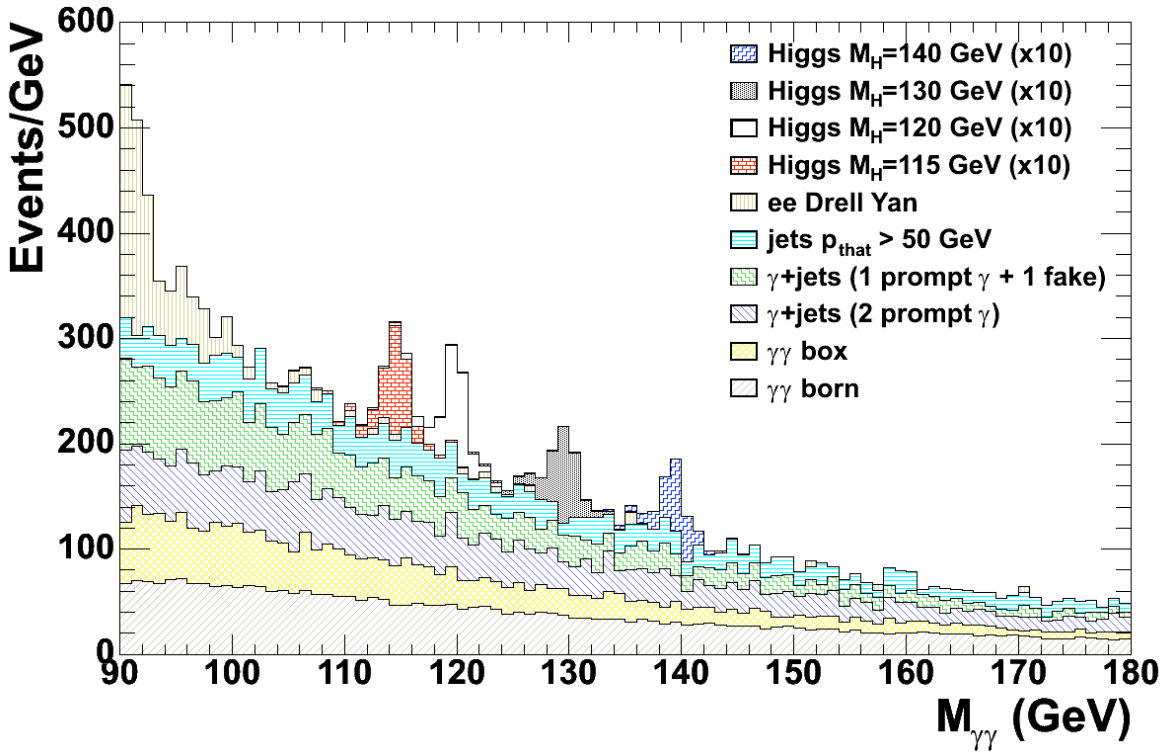

Figure 2.2: Diphoton invariant mass spectrum after the selection for the cut-based analysis. Events are normalised to an integrated luminosity of 1 fb^{-1} and the Higgs signal, shown for different masses, is scaled by a factor 10.

Table 2.5: Selection efficiency for the Higgs signal in different mass windows.

M_H(GeV/c^2)	Window ±1 GeV/c^2	Window ±1.5 GeV/c^2	Window ±2.5 GeV/c^2	Window ±5 GeV/c^2	Window Total
115	17%	21%	25%	28%	29%
120	18%	22%	26%	29%	30%
130	18%	22%	27%	31%	32%
140	18%	23%	28%	32%	34%
150	28%	24%	29%	33%	36%

puted using all generated signal events. The signal contribution to the total number of events is very small, particularly outside the mass region under study. The background can be estimated by a fit to the data mass distribution.

The error on the background estimation comes from two sources:

- the statistical precision which decreases with the size of the mass range that is used to perform the fit;
- the systematic error related to the shape of the function that is used to fit the distribution.

It is not possible to know the exact functional form of the background shape and the error must be estimated by assuming a function, simulating a distribution and then using a different function to fit the data. Clearly this error grows with the size of the mass range used. For a reasonable mass range of $\pm\ 10 - 20\,\mathrm{GeV/c^2}$ excluding +3 and -5 $\mathrm{GeV/c^2}$ from the Higgs boson mass under study and for an integrated luminosity of $20\,\mathrm{fb}^{-1}$ the statistical and systematic errors are estimated to be 0.4% and 0.5% respectively. The statistical error decreases with the integrated luminosity while the systematic error is constant.

2.1.4.2 Splitting into categories

Changing the cuts or adding new discriminating variables to this analysis does not give large improvements in the sensitivity. This can be seen, for example, from the fact that it is not possible to use the very powerful variable, $\mathrm{R_9}$, to reject events without loosing performance. This is because the increase in s/b ratio does not compensate the loss in efficiency.

The way to improve the sensitivity of the analysis is to keep all selected events but to split the sample into categories with different s/b ratios.

The following 3 possibilities are considered:

- 1 single category;
- 4 categories from 2 $\mathrm{R_9^{min}}$ ranges ($\mathrm{R_9^{min}}$ larger or smaller than 0.93) times 2 pseudo-rapidity regions $|\eta|^{max}$ in barrel or endcaps;
- 12 categories from 3 $\mathrm{R_9^{min}}$ ranges ($\mathrm{R_9^{min}} > 0.948$, $0.9 < \mathrm{R_9^{min}} < 0.948$ and $\mathrm{R_9^{min}} < 0.9$) times 4 pseudo-rapidity regions ($|\eta|^{\mathrm{max}} < 0.9$, $0.9 < |\eta|^{\mathrm{max}} < 1.4442$, $1.4442 < |\eta|^{\mathrm{max}} < 2.1$ and $|\eta|^{\mathrm{max}} > 2.1$).

Figure 2.3 shows the mass spectrum after splitting into four categories. The signal over background ratio is much larger in the best category and the composition of the background varies between the different samples: irreducible backgrounds dominate for large $\mathrm{R_9}$ and reducible backgrounds are larger for small $\mathrm{R_9}$.

Table 2.6 shows, for the 12 category analysis, the fraction of events along with the maximum s/b ratio in each category.

Table 2.6: Fractions of events in each of the 12 categories and maximum s/b in the mass region of $120\,\mathrm{GeV/c^2}$.

	$\lvert\eta\rvert^{\mathrm{max}} < 0.9$		$0.9 < \lvert\eta\rvert^{\mathrm{max}} < 1.4442$		$1.4442 < \lvert\eta\rvert^{\mathrm{max}} < 2.1$		$\lvert\eta\rvert^{\mathrm{max}} > 2.1$	
	frac.	s/b	frac.	s/b	frac.	s/b	frac.	s/b
$\mathrm{R_9^{min}} > 0.948$	15.5%	14.7%	13.1%	9.0%	10.8%	6.1%	8.5%	4.5%
$0.9 < \mathrm{R_9^{min}} < 0.948$	9.4%	12.2%	6.8%	6.8%	6.7%	4.8%	2.7%	2.8%
$\mathrm{R_9^{min}} < 0.9$	8.3%	7.6%	11.1%	4.3%	5.4%	3.2%	1.7%	2.2%

2.1.4.3 Systematic errors

The total error on the background is approximately 0.65% and is due to the uncertainty of the function fit to the side-bands of the mass distribution, estimated to be 0.5%, plus the statistical error on the fit that is approximately 0.4% for an integrated luminosity of $20\,\mathrm{fb}^{-1}$.

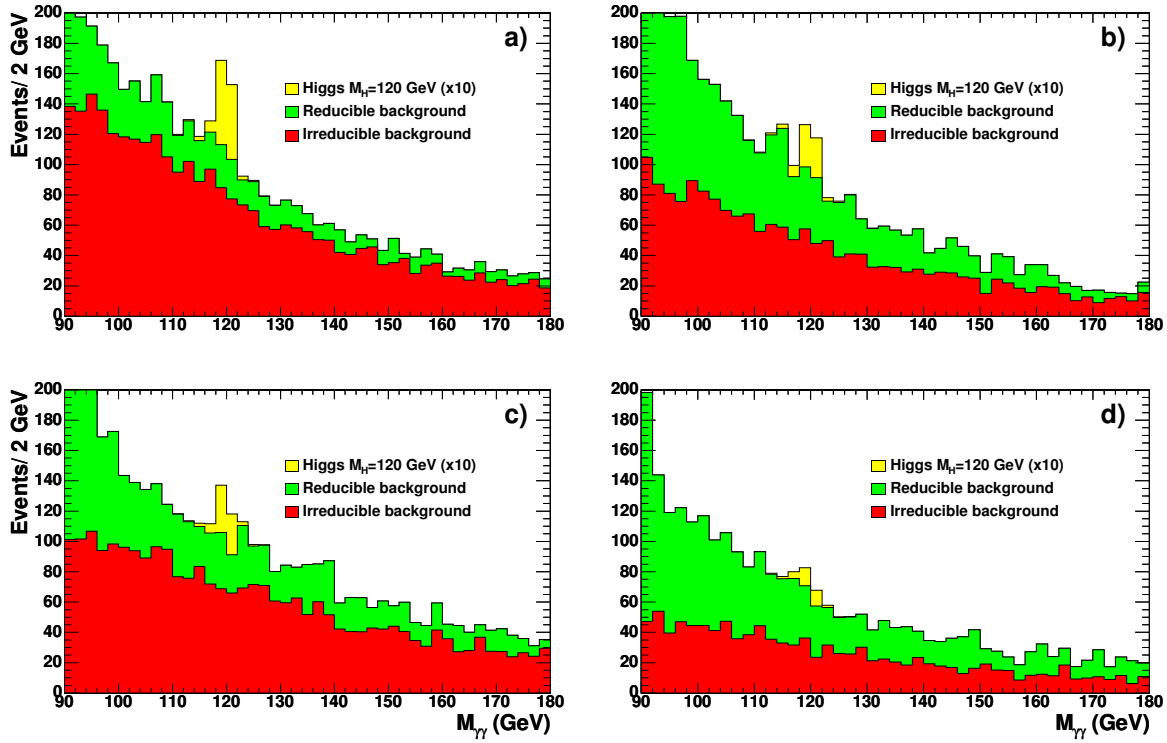

Figure 2.3: Invariant mass spectrum after the selection relative to the cut-based analysis with four categories defined in the text: barrel with large R_9 (a), barrel with small R_9 (b), endcaps with large R_9 (c) and endcaps with small R_9 (d), Events are normalised to an integrated luminosity of 1 fb^{-1} and the Higgs signal, shown for different masses, is scaled by a factor 10.

An error of 0.65% has a very large effect on the discovery CL when only one category is used. The reason is that a large fraction of signal events corresponds to a very low s/b, of the order of a percent. The effect can be reduced by applying a cut on the signal over background s/b. This corresponds to using events in a mass window around the analysed mass, until s/b becomes smaller than the chosen cut. The optimal cut for this analysis is 0.02.

When the events are split into categories the number of background events in each category is reduced on average by $1/\mathrm{N}_{\mathrm{cat}}$ and this increases the statistical error on the background estimation by approximately a factor $\sqrt{\mathrm{N}_{\mathrm{cat}}}$, but this error is completely uncorrelated between the different categories. The error related to the uncertainty of the fit function remains constant and it is also uncorrelated between the different categories because, due to the different cuts the background shapes are different and described by different functions. The total error is then less than the total error reduced by $1/\mathrm{N}_{\mathrm{cat}}$. This reduces the effect of the systematic error on the discovery.

The effect of the systematic error on the background estimation is also related to the signal over background of the analysis. A more sensitive analysis, for which a larger part of the signal has a higher s/b ratio, is less affected by the same relative uncertainty on the background.

Clearly the current understanding of the background is affected by larger uncertainties such as: cross section, diphoton kinematic distributions and efficiency of the selection (mainly affected by jet fragmentation, pile-up and by the structure of the underlying events).

The systematic error on the signal, that as has been mentioned has no effect on the discovery CL, has contributions from the theoretical uncertainty of the cross section (+15 -12% from the scale variation and +4 -5%), from the measurement of the integrated luminosity ($\sim$5%), from the trigger ($\sim$1%), from the analysis selection (that will be measured for example with Z$\to \mu\mu\gamma$) and from the uncertainties on the photon energy resolution. Other effects that could modify the ability to discover the Higgs boson are: uncertainties on the structure of the underlying events, that could change the efficiency of the primary vertex determination and the amount of material in the tracker before the electromagnetic calorimeter.

The effect on the performances of the analysis of an increase of 20% of the tracker material has been evaluated. The main effects on such change on the analysis would be:

- increase of the inefficiency of the track isolation requirements for early photon conversions, before or inside the second layer of the pixel detector.
- increase of the inefficiency of ECAL isolation cut;
- decrease of the value of $\mathrm{R_9}$ for all photons that would cause a migration of events from more sensitive categories to less sensitive categories.

It was estimated that such change would increase the luminosity needed to achieve a given discovery CL of approximately 6%. Given that the amount of tracker material will be known with a precision of $\sim$2% the related systematic error is less than 1%.

In what follows a conservative 20% systematic error on the signal is assumed. It affects exclusion of a signal, not discovery, since the signal rate is directly measured from data in case of discovery.

2.1.4.4 Results of the cut-based analysis

Table 2.7: Integrated luminosity needed to discover or exclude the Higgs boson with mass 120 GeV/c^2 with or without taking into account the systematic errors (fb^{-1}).

Analysis	5 σ discovery no syst	5 σ discovery syst	3 σ evidence no syst	3 σ evidence syst	95% exclusion no syst	95% exclusion syst
counting exp.	27.4	48.7	10.0	13.2	4.5	6.5
1 category	24.5	39.5	8.9	11.5	4.1	5.8
4 categories	21.3	26.0	7.5	9.1	3.5	4.8
12 categories	19.3	22.8	7.0	8.1	3.2	4.4

Table 2.7 shows the integrated luminosity needed to obtain 5σ discovery or 95% CL exclusion for a 120 GeV/c^2 mass Higgs boson with the different splittings. The effect of the systematic errors is also shown. We can observe how the performance increases and the effect of the error on the background estimation decreases with the number of categories. In the three cases (1, 4 and 12 categories) the event selection is the same and that the differences in performance come from the splitting of the total sample in different sub-samples with different sensitivities (s/b). In the split category analyses the computation of the log-likelihood ratio estimator is made separately for each 1 GeV/c^2 bin in mass, whereas in the "counting experiment" only a single (optimum) mass window is evaluated.

The integrated luminosity needed for discovery and exclusion, using the 12-category analy-

sis, for the mass range studied between 115 and 150 GeV/c^2 are shown in the plots at the end of the section (Figure 2.10). The Higgs boson can be discovered with mass between 115 and 140 GeV/c^2 with less than 30 fb^{-1} and excluded in the same mass range, at 95% CL, with less than 5 fb^{-1}.

As mentioned before, all these results have been obtained assuming an intercalibration of the ECAL, after having collected an integrated luminosity of 10 fb^{-1}. With the whole ECAL intercalibrated to a precision better than 0.5% over all the solid angle, the results improve such that approximately 10% less integrated luminosity is needed for discovery.

2.1.5 Optimised analysis estimating s/b for each event

In the optimised analysis 6 categories are used, 3 in which both photons are in the barrel and 3 in which at least 1 photon is in an endcap. The 3 categories are defined, as for the cut-based analysis, to have the lowest R_9 photon candidate with $\mathrm{R}_9 > 0.948$, $0.948 > \mathrm{R}_9 > 0.90$ and $\mathrm{R}_9 < 0.90$ respectively. The categories are labelled with numbers from 0 to 5: first the 3 barrel categories with decreasing values of R_9 then the 3 endcap categories again with decreasing values of R_9.

2.1.5.1 Mass distributions in categories

The diphoton mass distributions enable the separation of signal from background. Signal peaks sharply at the Higgs mass while the backgrounds are quite smooth. This allows good estimation of the magnitude of the background under the peak.

The best mass resolution and the best s/b ratio in the peak is found in category 0, with high R_9 in the barrel.

2.1.5.2 Loose selection of events for optimised analysis

Isolation requirements are applied to photon candidates prior to the computation of the neural network isolation variables $\mathrm{NN}_{\mathrm{isol}}$:

- the transverse E_{T} of the photon candidates must be larger than 40 GeV and the absolute value of their pseudo-rapidity less than 2.5;
- no tracks with p_{T} larger than 1.5 GeV/c must be present inside a cone with $\Delta\mathrm{R} < 0.1$ around the photon candidate;
- the total E_{T} of all ECAL island basic clusters with $\Delta\mathrm{R} < 0.3$ around the photon candidate, excluding those belonging to the super-cluster itself must be less than 5 GeV;
- the total transverse energies of HCAL towers within $\Delta\mathrm{R} < 0.35$ around the photon candidate must be less than 35 GeV;
- the sum of the transverse momenta of charged tracks within $\Delta\mathrm{R} < 0.2$ around the photon candidate must be less than 100 GeV/c.

Before optimising the final analysis, some additional cuts are applied. These both simplify the neural network training and slightly improve the performance. It is required that:

- the events pass the double photon High Level Trigger;
- the isolation neural net output is greater than 0.25 for both photons.

2.1.5.3 Optimised use of kinematic variables to separate signal and background

In addition to the mass, there are kinematic differences between signal and background. In particular the signal has a harder photon E_{T} distribution than the background – the background can have a high mass by having a large η difference between the photon candidates. Weak Boson Fusion and associated production of a Higgs with other massive particles enhance these differences between signal and background. The large, reducible backgrounds often have photon candidates that are not well isolated.

As with the Higgs searches performed at LEP, higher performance can be achieved if the expected signal over background, s/b, is estimated for each event. This is particularly effective if the s/b varies significantly from event to event. This is the case here due to wide variations in photon isolation and photon E_{T}. There is also significant dependence of the s/b on photon conversion and on location in the detector.

One photon isolation variable $\mathrm{NN_{isol}}$ for each photon, is combined with kinematic variables to help separate signal and background. A neural net is trained to distinguish background events, taken from the mass side-bands, from signal Monte Carlo events. There is no danger of over-training since background events from the signal mass region are not used and independent samples are used for the signal Monte Carlo. The input variables are devised to be insensitive to the diphoton mass so that the background rejection due to the kinematics and isolation is independent of the background rejection from the mass distribution.

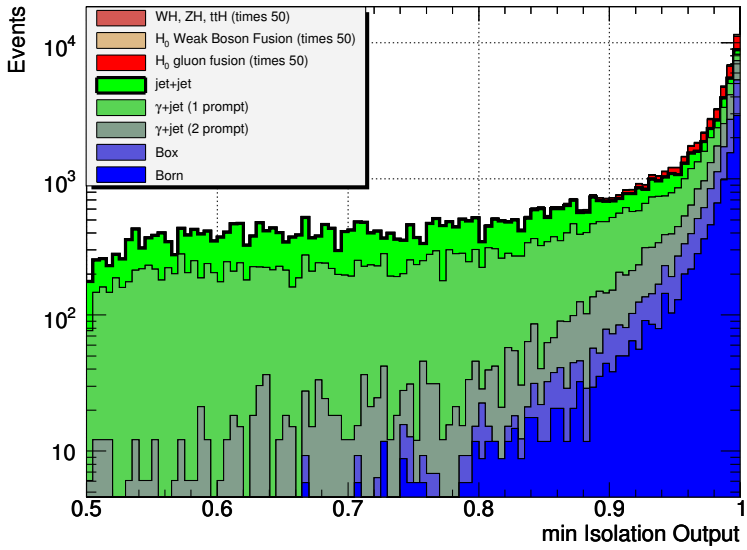

Figure 2.4: Distribution of the minimum value of the $\mathrm{NN_{isol}}$ variables of the two photon candidates. Events are normalised to an integrated luminosity of 7.7 fb^{-1} and the signal ($\mathrm{M_H}$=120 GeV/c^2) is scaled by a factor 50.

Six variables are used as inputs to a neural net. They are the isolation NN outputs $\mathrm{NN_{isol}}$ for the 2 photons, the transverse energies of the 2 photons, normalised to the diphoton mass, the absolute value of the rapidity difference between the 2 photons, and the longitudinal momentum of the photon pair.

The distributions of the input variables are shown for signal and background in Figures 2.4 and 2.5. Kinematic information that are likely to be highly sensitive to higher order correc-

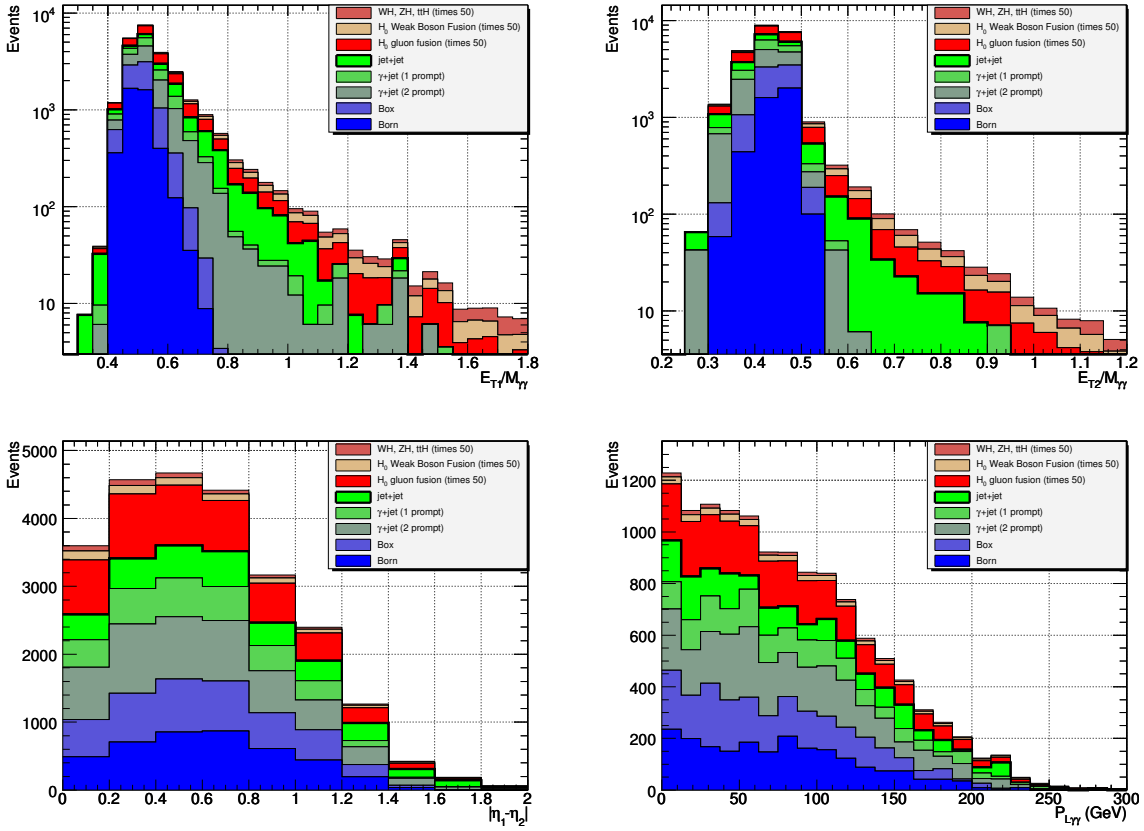

Figure 2.5: Distribution of the kinematic inputs to the neural network for signal and background sources. A value of the neural net output is required to be greater than 0.85. Events are normalised to an integrated luminosity of 7.7 fb^{-1} and the signal ($\mathrm{M_H}$=120 $\mathrm{GeV/c^2}$) is scaled by a factor 50.

tions to the background simulation has not been used. Such information, like the E_{T} of the Higgs boson candidate, the E_{T} transverse to the photon direction, and information about additional jets will ultimately be useful but may not be reliable until better simulations or actual data are available to train on.

The neural net is trained in each of the 6 categories independently. The net has 6 input nodes, 12 intermediate nodes in a single layer, and 1 output node. The error function has been modified from the standard to improve training toward a high signal over background region. A minimum neural net output cut is applied that eliminates 1% of the signal in each category and a function is fit to the distribution above that cut. These functions are used to bin the data and to smooth the background in a limited region.

It is useful to examine the neural net output distribution for events from different sources (Figure 2.6). Low NN outputs are dominated by photon candidates from jets which are not well isolated. The large peak at 0.85 represents both signal and background where the photon is relatively well isolated and the photon E_{T} is $\mathrm{M_H}/2$, corresponding to events with a large value of $\mathrm{NN_{isol}}$. Higher photon E_{T} events are found in the peak near 1. There is an enhancement of the signal, particularly for the WBF and associated production processes. The background there is dominated by events with at least one jet interpreted as a photon.

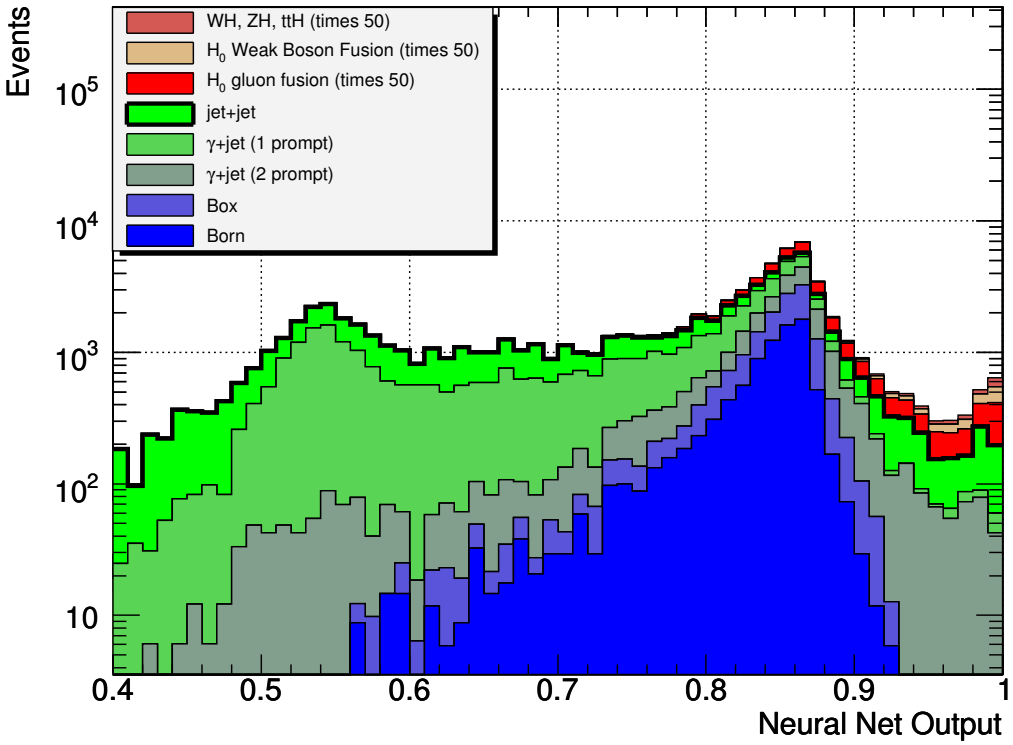

Figure 2.6: The neural net output for events in the barrel for each signal (M_H=120 GeV/c^2) and background source. Events are normalised to an integrated luminosity of 7.7 fb^{-1} and the Higgs signal is scaled by a factor 50.

2.1.5.4 Estimation of signal to background ratio for each event

In order to get the most information out of each event, the signal over background is estimated for each event. In the simplest analyses, cuts are applied to select only high signal over background events and those are counted. Such a simple analysis looses information because some of the events that are cut could contribute to the measurement and because some of the events that are accepted are not used optimally.

Events in the mass peak for the Higgs mass hypothesis under consideration have high signal over background expectation while events outside the peak have lower expected s/b. Similarly, events at high $\mathrm{NN_{kin}}$ output have higher s/b expectation. The kinematics and isolation information in $\mathrm{NN_{kin}}$ has been made independent of mass information so the two s/b ratios can be multiplied to get a good estimate of the s/b expectation for the event:

$$\left(\frac{s}{b}\right)_{est.} = \left(\frac{s}{b}\right)_{mass} \times \left(\frac{s}{b}\right)_{kin}$$

This is an estimate that is to bin signal and background events. If the estimate is bad, the performance of the analysis suffers because good s/b events are not well separated from bad ones. It is not possible for a bad estimate to make the analysis appear to perform too well. The s/b estimate need not be normalised correctly, since it is a relative number used to bin events.

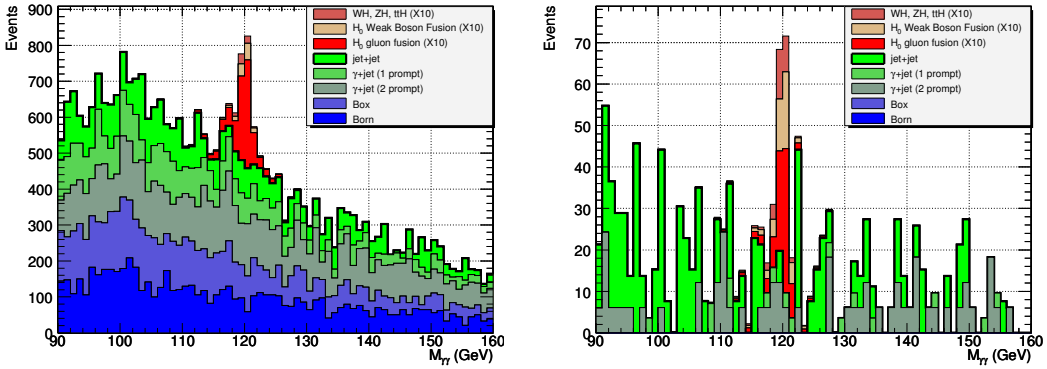

Figure 2.7: The diphoton mass distribution for each source for barrel events with kinematic neural net output greater than 0.85 (top) and 0.97 (bottom). Events are normalised to an integrated luminosity of 7.7 fb^{-1} and the Higgs signal (M_H=120 GeV/c^2) is scaled by a factor 10.

The events are binned according to the s/b estimate. Histograms are made in each of the six categories. The actual signal to background ratio is computed for the binned events and used to calculate confidence levels that data are consistent with a background-only hypothesis or with a signal-plus-background hypothesis.

2.1.5.5 Smoothing the background

The $H \rightarrow \gamma\gamma$ channel has the good feature that the mass is essentially independent of isolation and suitably chosen kinematic variables. With this factorisation assumption, background can be smoothed well even in regions with low statistics.

The background expectation in a bin must be reliably estimated in order to correctly calculate confidence levels. Downward fluctuations in the background estimation can have a significant impact on the CL. The number of simulated events for the irreducible (jet) backgrounds is about one seventh of the number that will be available in the data at the time it would be expected to discover the Higgs. Therefore problems with background estimation are even more difficult now than they will be when we have data.

The background distributions are very smooth in the mass variable, so the distribution in mass can be reliably smoothed. This is done by spreading each event over a ± 5 GeV/c^2 region according to the functions fit to the mass distribution. A wider mass region could be used but this would interfere with the training of the analysis on an independent sample in the mass side-bands.

The background distribution in the neural net output is also smoothed over a region of ± 0.05 using the fit functions. It is therefore quite important that the background fit functions accurately represent the neural net distribution. In the smoothing process, the normalisation of the background is carefully maintained to high accuracy.

With this two-dimensional smoothing accurate background expectations are obtained except in the regions with extremely small amounts of background. In such regions, bins must be combined until sufficient background events are available. If a s/b bin has too few MC background events contributing to it, it is combined with the nearest (lower s/b) bin. This is continued until there are sufficient events. This combination clearly reduces the sensitivity of the analysis but cannot be avoided without a more detailed understanding of the back-

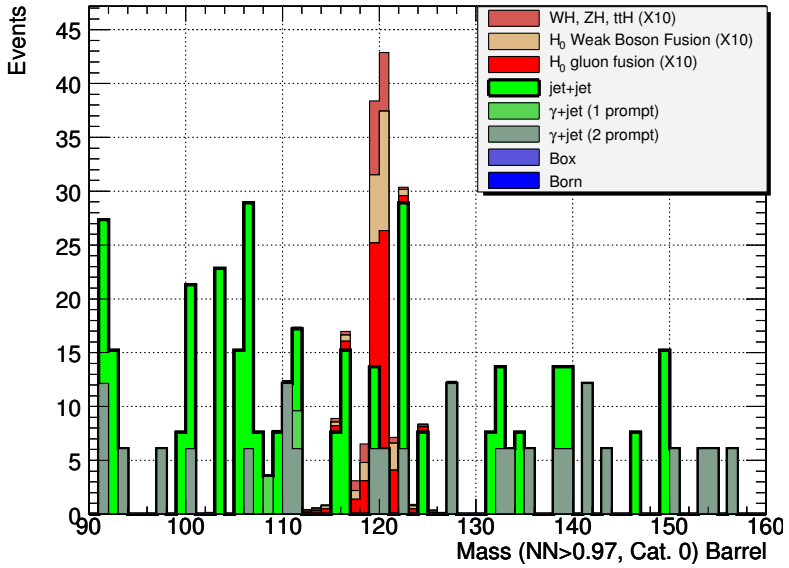

Figure 2.8: The diphoton mass distribution for each source for barrel events with kinematic neural net output greater than 0.97 and $R_9 > 0.948$. Events are normalised to an integrated luminosity of 7.7 fb^{-1} and the Higgs signal (M_H=120 GeV/c^2) is scaled by a factor 10.

ground, which is a goal for the future. At present, at least 20 Monte Carlo background events are required in a bin. Since the current MC samples contain about seven times less events than expected in the data, significant improvements are possible, allowing higher s/b bins to be used, resulting in better performance.

Figure 2.7 shows the mass distributions for barrel events with two different cuts on the neural net output. The looser cut simply excludes most of the obviously non-isolated candidates. It can be seen that all of the backgrounds are important at this level. The tighter cut highly enhances the s/b ratio and emphasises the importance of smoothing, which has not been applied to the background in this distribution.

Figure 2.8 shows the mass distribution for neural net output greater than 0.97 in category 0. Again it is clear that smoothing in two dimensions is needed to get a reasonable estimate of the background. It is useful to note that even in this very high s/b region, the largest contribution to the signal is from gluon fusion, although the relative contributions of the other production processes has increased.

2.1.5.6 Combination of categories into final s/b distribution

At this point the signal and background is binned in s/b in six categories. These could be used to calculate the confidence level, however, it seems most useful, in the light of future plans to analyse separate channels, to combine the categories into one s/b plot in a similar way as may be used to re-combine channels. The six histograms are combined into one which can be used calculate confidence levels. The combination is based on the actual signal to background in each bin. In principle, this is the same as combining results from different channels or even from different experiments in a way that makes optimal use of all channels and does not pollute high quality channels with data of lesser purity.

The final binning of data into s/b bins is shown in figure 2.9. The plot extends from very low

signal to background to a small number of events with $s/b > 1$.

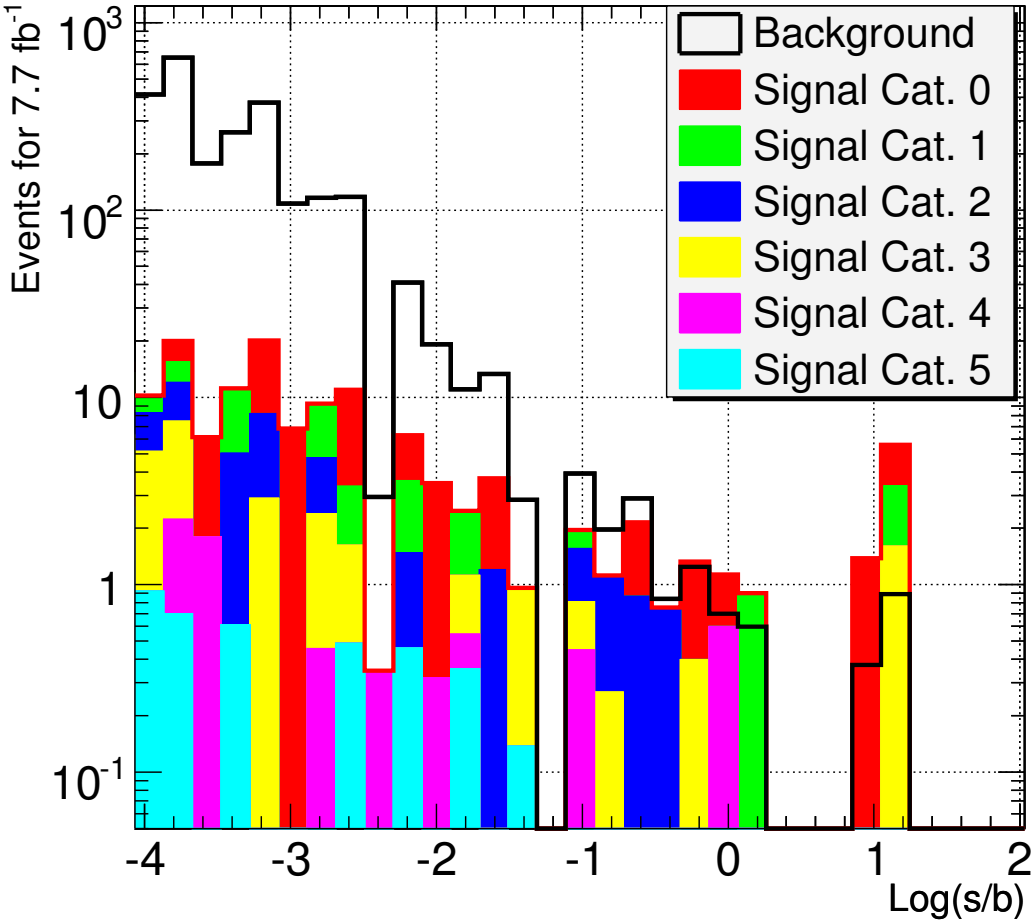

Figure 2.9: The final distribution of binned signal (M_H=120 GeV/c^2) and background in $\log(s/b)$ for an integrated luminosity of 7.7 fb^{-1}. Here the Higgs signal is normalised to the integrated luminosity and the statistics benefits of the smoothing of the background. Signal and background events are added independently.

The relative contribution of barrel and endcap categories can be estimated from the total LLR computed and LLRs computed excluding each category. The six categories have rather widely varying contributions to the Log Likelihood Ratio and hence to the performance of the analysis. Table 2.8 shows the fraction of signal and the fraction of the LLR for each category.

Some of the categories have a fairly small effect on the final result. This remains true after the application of systematic normalisation uncertainties described below. It is clear that photon conversions result in a significant deterioration of the performance. It is hoped to mitigate this somewhat by using the conversion track reconstruction in the future, but the poorer mass resolution cannot be recovered and a big effect is not expected.

Table 2.8: Performance in the six categories for $M_H = 120\,GeV/c^2$.

Category	Signal %	LLR %
0	27.8	48.0
1	16.1	24.8
2	21.7	11.9
3	16.6	9.7
4	9.0	4.1
5	8.8	1.5

2.1.5.7 Results of the optimised analysis

The same estimates of systematic error are used to obtain the results in the optimised analysis as are used in the cut-based analysis. Most of the development and studies have been made for a Higgs mass of $120\,GeV/c^2$. For this mass, a 5σ discovery can be made with about $7\,fb^{-1}$ luminosity. A 1% background normalisation uncertainty corresponds to an increase of the luminosity needed for a 5σ discovery from $7\,fb^{-1}$ to $7.7\,fb^{-1}$.

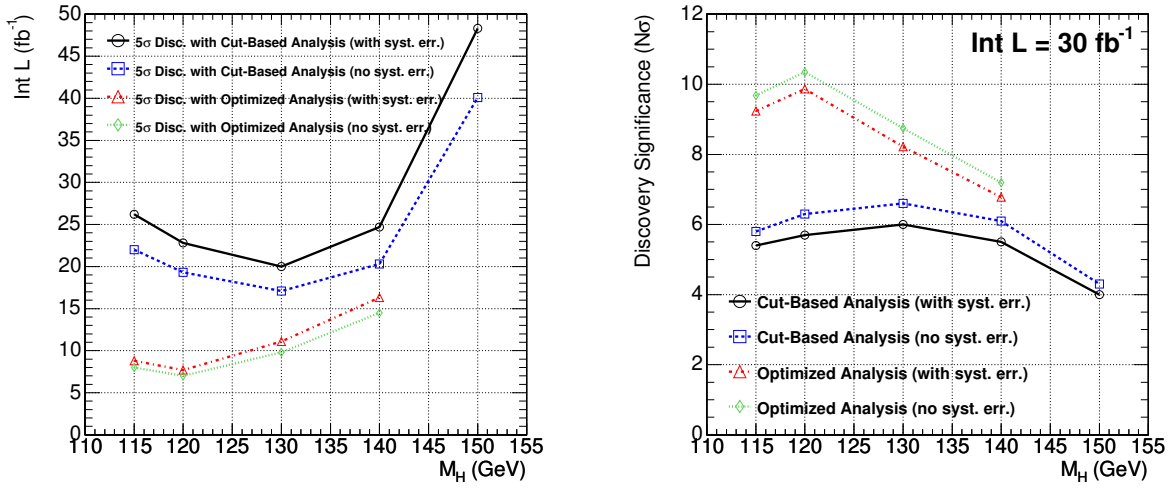

Figure 2.10: Integrated luminosity needed for a 5σ discovery (left) and discovery sensitivity with an integrated luminosity of $30\,fb^{-1}$ (right) with the optimised analysis. The results from the cut-based analysis in 12 categories are also shown for comparison.

There is a great deal of uncertainty in this benchmark estimate of luminosity due to our poor understanding of the backgrounds we will contend with when the LHC starts running, however, this is not considered here as a systematic error on a discovery since it is proposed to measure the background from the data. Figure 2.10 shows the luminosity needed for a 5σ discovery and the discovery sensitivity with an integrated luminosity of $30\,fb^{-1}$ for several Higgs masses, both for the fully optimised analysis and for the cut-based analysis using 12 categories described in Section 2.1.4.4. It seems possible to discover, or at least have strong evidence for a low mass Higgs in the first good year of running.

2.1.6 Measurement of the Higgs boson mass

If the Higgs boson will be discovered in the $H \rightarrow \gamma\gamma$ channel the we will be able to measure its mass. We have studied the mass measurements with the cut based analysis with two different methods:

- measurement from the ΔLog(likelihood) using all events;
- measurement from the ΔLog(likelihood) using the cut-based analysis split in 12 categories.

The expected statistical errors are shown in Table 2.9 for an integrated luminosity of 30 fb^{-1}. The statistical errors simply scale with $1/\sqrt{\mathrm{Int\,L}}$. The errors are slightly asymmetric, due to the tail of the reconstructed Higgs mass distribution at lower masses, the positive error being approximately 10% smaller than the negative. The Table shows the average between the two.

Table 2.9: Expected statistical errors on the Higgs boson mass measurement for 30 fb^{-1}.

M_H	115 GeV/c^2	120 GeV/c^2	130 GeV/c^2	140 GeV/c^2	150 GeV/c^2
All events	184 MeV/c^2 0.16%	184 MeV/c^2 0.15%	201 MeV/c^2 0.15%	222 MeV/c^2 0.16%	298 MeV/c^2 0.20%
12 categories	127 MeV/c^2 0.11%	139 MeV/c^2 0.12%	129 MeV/c^2 0.10%	156 MeV/c^2 0.11%	204 MeV/c^2 0.14%

As we can see the statistical error will be 0.1 to 0.2% already with 30 fb^{-1}, when the significance of the discovery would be 5 to 6 σ with the cut based analysis. Of course this measurement will be affected by the uncertainty of the absolute scale of the photon energy measurement that will be derived for example by the measurement of the Z mass in the radiative Z decays Z$\rightarrow \mu\mu\gamma$.

2.1.7 Summary

A standard cut-based analysis can discover the Higgs boson with 5σ significance between the LEP lower limit and 140 GeV/c^2 with less than 30 fb^{-1} of integrated luminosity. Approximately 5 fb^{-1} are needed to exclude its existence in the same mass range.

It has been shown that the $H \rightarrow \gamma\gamma$ channel can be used to discover a low mass Higgs with an integrated luminosity not too different from that needed for higher mass Higgs, 7.7 fb^{-1} at 120 GeV/c^2 with an analysis using an event by event estimation of the s/b ratio. Because of the excellent mass resolution expected in the diphoton channel, the background rate and characteristics from the data can be determined from diphoton events at masses away from the Higgs mass hypothesis.

An inclusive analysis has been presented. In future the various signal channels will be identified by looking for additional jets, leptons, or missing energy. This will clearly improve the sensitivity of the analysis.

2.2 Benchmark channel: $H \rightarrow ZZ^{(*)} \rightarrow$ 4 electrons

One of the most promising road towards a discovery at the LHC of the Higgs boson postulated in the SM is via single production followed by a cascade decay into charged leptons, $\mathrm{H} \rightarrow \mathrm{ZZ}^{(*)} \rightarrow l^+l^-l^+l^-$.

The single Higgs boson production benefits from a high cross-section, with values of about 40×10^3 fb at $m_{\mathrm{H}} = 130\,\mathrm{GeV/c^2}$ and decreasing monotonically to about 10×10^3 fb around $m_{\mathrm{H}} = 300\,\mathrm{GeV/c^2}$. The production cross-section is dominated ($\gtrsim 80\%$) over this mass range by gluon-gluon fusion processes via triangular loops involving heavy quark (mostly the top quark) flavours. The branching ratio for the $\mathrm{H} \rightarrow \mathrm{ZZ}^{(*)}$ decay in the SM is sizeable for any m_{H} value above $130\,\mathrm{GeV/c^2}$. It remains above 2% for $m_{\mathrm{H}} \leq 2 \times M_W$ with a peak above 8% around $m_{\mathrm{H}} \simeq 150\,\mathrm{GeV/c^2}$, and rises to values of 20 to 30% for $m_{\mathrm{H}} \geq 2 \times m_{\mathrm{Z}}$. The Z bosons have a 10% probability to yield a pair of charged leptons. Thus, the decay chain $\mathrm{H} \rightarrow \mathrm{ZZ}^{(*)} \rightarrow l^+l^-l^+l^-$ (in short $\mathrm{H} \rightarrow 4l$) offers a possibly significant and very clean and simple multi-lepton final state signature for the SM Higgs boson at the LHC.

Ultimately, the channel can provide a precision determination of the Higgs boson mass and production cross-section. The anti-correlation of the Z spin projections in the $\mathrm{H} \rightarrow \mathrm{ZZ}$ decay and the polarisation of each Z boson can be used to constrain, and eventually determine, the spin and CP quantum numbers of the Higgs resonance. Furthermore, the $\mathrm{ZZ}^{(*)}$ and $\mathrm{WW}^{(*)}$ decay modes are related via $SU(2)$ and the combination of channels could allow for cancellation of some systematic uncertainties in a determination of the Higgs coupling. But first and foremost is the necessity to be best prepared for a discovery at the LHC.

In this section, the discovery potential of the CMS experiment for the SM Higgs boson is discussed in the mass range of $120 \leq m_{\mathrm{H}} \leq 300\,\mathrm{GeV/c^2}$, focusing on the $4e$ channel. The analysis [37] relies on a detailed simulation of the detector response in the experimental conditions of the first years of low luminosity LHC running. The signal and background Monte Carlo datasets used for this prospective are described in Section 2.2.1. The detailed High Level Trigger (HLT) and reconstruction algorithms used at each step of this analysis have been presented in [7]. Basic, and in part compulsory, triggering and pre-selection steps for data reduction are described in Section 2.2.2. Simple observables from the electron reconstruction are used to characterise the event signature for this pre-selection step. The final event selection relies on more involved requirements for primary electrons coupled with basic event kinematics and is presented in Section 2.2.3. The selection is optimised to preserve a best signal detection efficiency and highest significance for a discovery. Emphasis is put on realistic strategies for the control of experimental errors and the estimation of systematic uncertainties on physics background rates. These are described in Section 2.2.4. Results on the expected discovery reach of the SM Higgs boson in CMS in the $\mathrm{H} \rightarrow 4e$ channel and for the measurement of its mass, width and cross-section are finally presented in Section 2.2.5.

2.2.1 Datasets for signal and background processes

Monte Carlo data samples for the signal from single SM Higgs boson production as well as for SM background from $\mathrm{ZZ}^{(*)}$ pair production, $\mathrm{t\bar{t}}$ pair production and $\mathrm{Zb\bar{b}}$ associated production are used. The signal and background processes are generated for pp collisions at the LHC at a centre-of-mass energy $\sqrt{\mathrm{pp}} = 14\,\mathrm{TeV}$, with pile-up conditions from multiple collisions as expected in a collider machine configuration providing an instantaneous luminosity of $2 \times 10^{33}\,\mathrm{cm^{-2}s^{-1}}$ (of $\mathcal{O}(10)\,\mathrm{fb^{-1}/year}$). All cross-sections are normalised within acceptance to Next-to-Leading-Order (NLO) calculations. The event generators are interfaced

with PHOTOS [38, 39] for the simulation of QED final state radiations. The non-perturbative parton density functions (PDFs) in the proton are taken to be the CTEQ6 distributions [12].

The Higgs boson is produced via either gluon fusion and weak boson fusion processes. The $4e$ signal samples are generated at various m_{H} with PYTHIA [24]. The Higgs boson is forced to decay into a Z boson pair. The Z bosons are subsequently forced to undergo a decay in electron-positron pair. The signal is normalised to the value of total cross-section at NLO calculated including all Higgs boson production processes via HIGLU [40], with branching ratios $BR(\mathrm{H} \rightarrow \mathrm{ZZ}^{(*)})$ calculated via HDECAY [41].

In the $4e$ channel (and similarly for the 4μ channel), an additional enhancement of the signal is considered which is due to the constructive final state interference between like-sign electrons originating from different $\mathrm{Z}^{(*)}$ bosons [42]. This enhancement has been re-evaluated with COMPHEP [43] and amounts to a factor 1.130 ± 0.006 at $m_{\mathrm{H}} = 115\,\mathrm{GeV/c^2}$, slowly decreasing to a negligible value when approaching $m_{\mathrm{H}} \approx 2m_{\mathrm{Z}}$.

The $\mathrm{ZZ}^{(*)}$ SM background continuum is generated using PYTHIA [24]. This includes only the t-channel contribution with $q\bar{q}$ in the initial state. The missing s-channel might contribute up to 10% for low Higgs boson masses and can be neglected for higher masses. The differential cross-section is re-weighted using m_{4e} dependent NLO K-factors obtained with MCFM 4.1, with an average K-factor of $< K_{NLO} >= 1.35$. Both Z bosons are constrained within the mass range 5-150 $\mathrm{GeV/c^2}$ and are forced to decay into charged lepton pairs, with the τ leptons subsequently forced to undergo leptonic decays via $\tau \rightarrow \mu\nu$ or $\tau \rightarrow e\nu$. The missing gg contribution is estimated to be of order 20% at LO [42], with $\pm 8\%$ uncertainties and with unknown NLO K-factors. Recent calculations with TOPREX [44] of the gluon fusion production process of two real Z confirm above assumptions, and this contribution has been shown to remain stable after kinematic cuts for a $\mathrm{H} \rightarrow 4l$ analysis. The cross-section here is simply increased by the mean expected contribution.

The $\mathrm{t\bar{t}}$ background sample is also generated with PYTHIA [24], with W bosons and τ leptons forced to leptonic decays, but with b quarks left to decay freely. Both gluon fusion and quark annihilation initial states are simulated and the cross-section is normalised to the NLO value of $840 \pm 5\%(\mathrm{scale}) \pm 3\%(\mathrm{PDF})$ pb [45].

The $\mathrm{Zb\bar{b}}$ background is generated using all lowest order $gg \rightarrow e^+e^-\mathrm{b\bar{b}}$ and $qq' \rightarrow e^+e^-\mathrm{b\bar{b}}$ diagrams (excluding diagrams involving the SM Higgs boson) calculated with COMPHEP [43] and interfaced with PYTHIA [24] for showering and hadronisation. All possible combinations of quarks are considered in the initial state. The total LO cross-section for $m_{ee} > 5\,\mathrm{GeV/c^2}$ is 115 pb of which about 89% originates from gg processes, 7.7% involve u-like quarks and 3.2% involve d-like quarks in the initial state. The hadronisation and decay of the b quarks are left free. A NLO K-factor of 2.4 ± 0.3 is applied. Signal and background events are filtered at generator level for further analysis if satisfying the following acceptance requirements: $\geq 2e^+$ and $\geq 2e^-$ with $p_T^e > 5\,\mathrm{GeV/c}$ in $|\eta| < 2.7$. In addition for the $\mathrm{Zb\bar{b}}$ background, at least two e^+e^- pairs with invariant mass in the range 5-400 $\mathrm{GeV/c^2}$ are required. In Table 2.10 cross-sections at NLO and after pre-selection, as well as number of events in data samples available for analysis after pre-selection are given.

Detailed simulation of the CMS detector is performed using the official CMS simulation OSCAR. Reconstruction of physics objects is performed in ORCA.

Table 2.10: Total cross-sections at NLO (pb), cross-section in the $4e$ channel within acceptance (fb), and number of accepted events in data samples available for analysis.

m_{H} (GeV/c^2)	σ_{NLO} (pb)	$\sigma_{NLO} \times BR \times Acc.$ (fb)	$N_{simul.}$
115	47.73	0.27	10000
120	44.30	0.48	10000
130	38.44	1.11	10000
140	33.69	1.78	10000
150	29.81	1.94	10000
160	26.56	0.92	10000
170	23.89	0.43	10000
180	21.59	0.98	10000
190	19.67	3.58	10000
200	17.96	3.94	10000
250	12.37	3.07	10000
300	9.58	2.60	10000
$\mathrm{ZZ}^{(*)}$	29.0	20.2	150 000
$\mathrm{Zb\bar{b}}$	276.3	120.4	87 000
$\mathrm{t\bar{t}}$	840	194.0	500 000

2.2.2 Data reduction

The events of interest for the Higgs boson search in the $\mathrm{H} \rightarrow 4e$ channel must satisfy a minimal set of requirements.

A first and compulsory condition for the events is to satisfy the CMS Level 1 (hardware) trigger conditions and the filtering of the (software) HLT. This triggering step is described in Section 2.2.2.1. The basic electron triggers are expected to be saturated by SM processes such as the single Z and W production. Further filtering is obtained with a minimal set of additional electron requirements as described in Section 2.2.2.2.

The pre-selection must preserve the signal acceptance, and especially the electron reconstruction efficiency, until later stages where the analysis can best profit from more involved algorithms applied to reduced event samples.

2.2.2.1 Triggering

The events must have satisfied the *single e*, *double e* or *double relaxed e* requirements at L1/HLT level. The *single e* trigger requires one isolated (charged) "electromagnetic" object with a threshold set at a reconstructed transverse energy in the electromagnetic calorimeter (ECAL) of $E_T = 26\,\mathrm{GeV}$. The *double e* trigger requires two isolated (charged) "electromagnetic" objects, each above a threshold of $E_T = 14.5\,\mathrm{GeV}$. In contrast, the *double relaxed e* trigger does not imposed isolation for the (charged) "electromagnetic" objects and the increased rate is compensated by a higher threshold of $E_T = 21.8\,\mathrm{GeV}$.

The trigger efficiency for the Higgs boson signal, normalised to the cross-section within acceptance as defined in Section 2.2.1, is above 95% for masses $m_{\mathrm{H}} > 130\,\mathrm{GeV/c^2}$.

2.2.2.2 Pre-selection of four electron candidates

Following the Level-1 and HLT filtering steps, the event candidates must further satisfy basic electron pre-selection requirements. These requirements are designed to reduce possible background sources involving "fake" electron contamination from QCD jets.

For Higgs bosons with a mass m_{H} below $300\,\mathrm{GeV/c^2}$, the $4e$ final state always involves at least one (or few) low p_T^e electron(s). In the range of m_{H} values below the Z pair production threshold, where the Z and Z* bosons themselves receive in general only small transverse momentum, the mean p_T^e of the softest electron falls in a range where a full combination of tracking and calorimetry information becomes important. The p_T^e spectra for signal events at $m_{\mathrm{H}} = 150\,\mathrm{GeV/c^2}$ is shown in Fig. 2.11a. The softest electron, which generally couples to the off-shell Z*, has a most probable p_T^e value below $10\,\mathrm{GeV/c}$ for masses $m_{\mathrm{H}} \lesssim 140\,\mathrm{GeV/c^2}$. Hence, an excellent electron reconstruction is essential down to very low p_T^e values, well below the range of $p_T^e \simeq 40-45\,\mathrm{GeV/c}$ for which the reconstruction will be best constrained in CMS via measurements with SM single Z and single W production. The control of systematic uncertainties from experimental data is a major issue for such low p_T^e electrons and this will discussed in detail in Section 2.2.4.

This analysis makes use of the elaborate reconstruction procedures which have been introduced very recently in CMS and have been described in detail in Ref. [46]. The electron identification and momentum measurements are somewhat distorted by the amount of tracker material which is distributed in front of the ECAL, and by the presence of a strong magnetic field aligned with the collider beam z axis. The procedures introduced in Ref. [46] provide new useful observables that allow to better deal with these detector effects, combining information from the pixel detector, the silicon strip tracker and the ECAL.

The pre-selection of the signal event candidates relies on the presence of at least 2 e^+ and 2 e^- candidates within the acceptance $|\eta| < 2.5$ and each with $p_T > 5\,\mathrm{GeV/c}$, verifying the following characteristics:

- $E_{\mathrm{sc}}/p_{\mathrm{in}} < 3$, where E_{sc} is the supercluster energy and p_{in} the track momentum at the interaction vertex,
- $|\Delta\phi_{\mathrm{in}}| = |\phi_{\mathrm{sc}} - \phi_{\mathrm{in}}^{\mathrm{extrap}}| < 0.1$, where ϕ_{sc} is the energy weighted ϕ position of the supercluster and $\phi_{\mathrm{in}}^{\mathrm{extrap}}$ is the ϕ of the track at vertex, extrapolated to the ECAL assuming a perfect helix,
- $|\Delta\eta_{\mathrm{in}}| = |\eta_{\mathrm{sc}} - \eta_{\mathrm{in}}^{\mathrm{extrap}}| < 0.02$, with notations as above,
- $H/E < 0.2$, where H is the energy deposited in the HCAL tower just behind the electromagnetic seed cluster and E the energy of the electromagnetic seed cluster,
- $\sum_{cone} p_T^{\mathrm{tracks}}/p_T^e < 0.5$, a loose track isolation requirement, whose calculation will be described in Section 2.2.3.1.

The electron pre-selection efficiency is shown in Fig. 2.11b and Fig. 2.11c as a function of p_T^e and η^e for the electrons from Higgs boson events at $m_{\mathrm{H}} = 150\,\mathrm{GeV/c^2}$. The efficiency steeply rises and reaches a plateau around 86% for $p_T^e \gtrsim 20\,\mathrm{GeV/c}$. The efficiency is above 90% for $|\eta| \lesssim 1.1$ and decreases towards the edge of the tracker acceptance when approaching $|\eta| \simeq 2.5$. The pre-selection efficiency for electrons from the same sample is represented in Fig. 2.11d as a two-dimensional map in the p_T versus η plane.

The absolute efficiencies for the Higgs boson signal at different m_{H} values and for the backgrounds are shown in Fig. 2.12a after triggering and the multi-electron pre-selection step. The acceptance for the Higgs boson signal is maintained above 50% in the full relevant mass range.

The signal and background events fulfilling the triggering and pre-selection steps are represented in the reconstructed invariant mass m_{4e} spectrum in Fig. 2.12b. The Higgs boson

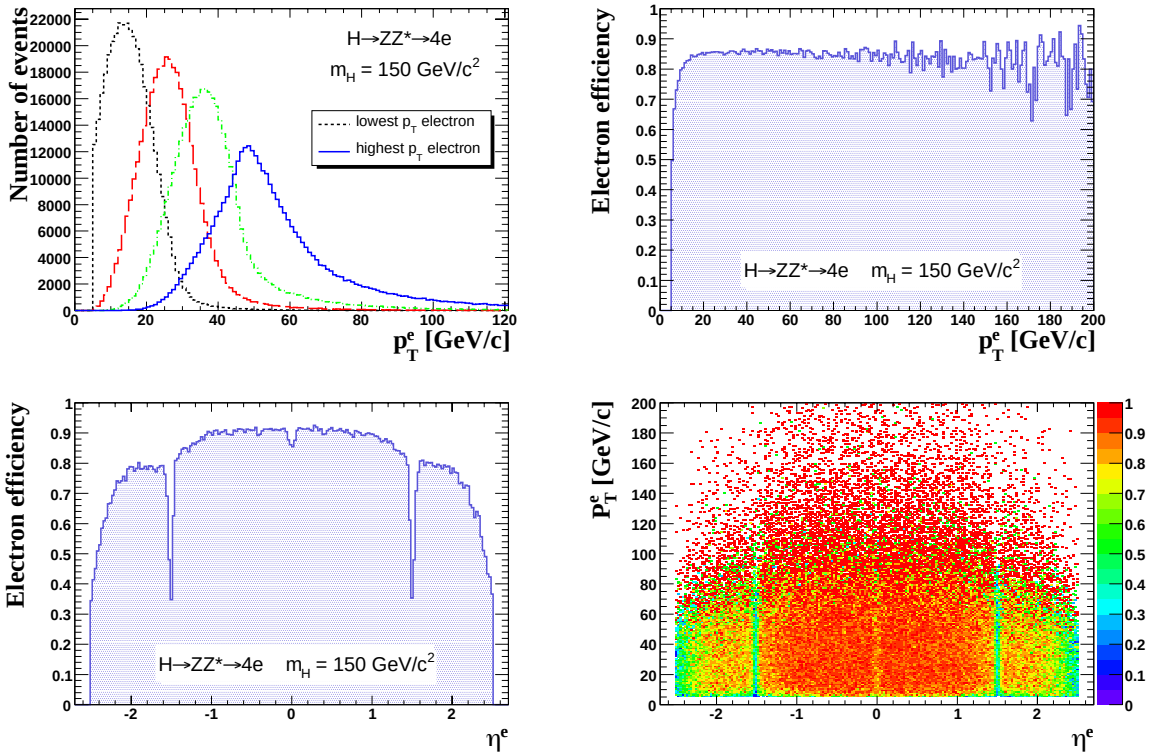

Figure 2.11: Electrons in SM Higgs boson $4e$ decay channel for $m_{\mathrm{H}} = 150\,\mathrm{GeV/c^2}$; a) transverse momentum of each of the four final state electrons; b) efficiency at pre-selection as a function of p_T^e; c) efficiency at pre-selection as a function of η_e; d) efficiency in the p_T^e versus η_e plane.

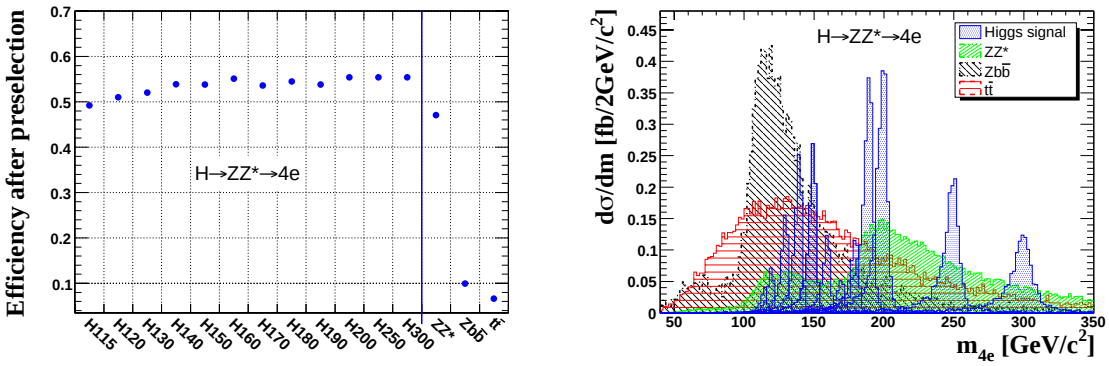

Figure 2.12: Higgs boson signal and dominant background sources after pre-selection step; a) overall pre-selection efficiency for m_{H} in the range from 115 to 300 $\mathrm{GeV/c^2}$ and the background from $\mathrm{ZZ}^{(*)}$ continuum, from $Zb\bar{b}$ and $t\bar{t}$; b) separate signal and background contributions to the spectrum of reconstructed invariant mass m_{4e}.

signal is seen to emerge above the background for masses around 150 GeV/c^2 and above $\simeq 2m_{\mathrm{Z}}$. More background suppression is required elsewhere.

2.2.3 Event selection and kinematic reconstruction

The further steps of the event selection rely on a more detailed characterisation of the electron candidates and simple kinematic expectations. The electrons from the Higgs boson, in contrast to at least one e^+e^- pair from the $\mathrm{t\bar{t}}$ and $\mathrm{Zb\bar{b}}$ backgrounds, are isolated and originate from a common primary vertex. The corresponding analysis requirements are discussed in Section 2.2.3.1. Profiting from the expectation of a narrow resonance in the m_{4e} spectrum, and of the likely presence of a real Z boson in the final state, the kinematics and its simple evolution with m_{H} can be further exploited. The electrons of the e^+e^- pair at lowest m_{ee} have on average a much harder p_T^e spectrum for the Higgs boson signal than for the $\mathrm{t\bar{t}}$ and $\mathrm{Zb\bar{b}}$ backgrounds. Moreover, the combination of the Z and Z$^{(*)}$ mass spectra distinguishes the Higgs boson signal from the ZZ$^{(*)}$ SM background continuum. These kinematic requirements are discussed in Section 2.2.3.2.

2.2.3.1 Isolated primary electrons

A loose vertex constraint is first imposed on the longitudinal impact parameter for the four electron candidates in each event. All electrons should verify $IP_L/\sigma_L < 13$, where σ_L is the error on the longitudinal impact parameter IP_L. The main vertex constraint is imposed on the transverse impact parameter of the electrons to suppress secondary vertices. Secondary electrons appear for instance in semi-leptonic decays in the hadronisation of the b quark jets in $\mathrm{Zb\bar{b}}$ and $t\bar{t}$ background events. The sum of the transverse impact parameter significance (IP_T/σ_T), i.e. the ratio of the transverse impact parameter IP_T over its error σ_T, is shown in Fig. 2.13a (Fig. 2.13b) for the e^+e^- pair with invariant mass m_{ee} closest (next-to-closest) to the nominal Z boson mass m_{Z}. For both of these background sources, the displaced vertices

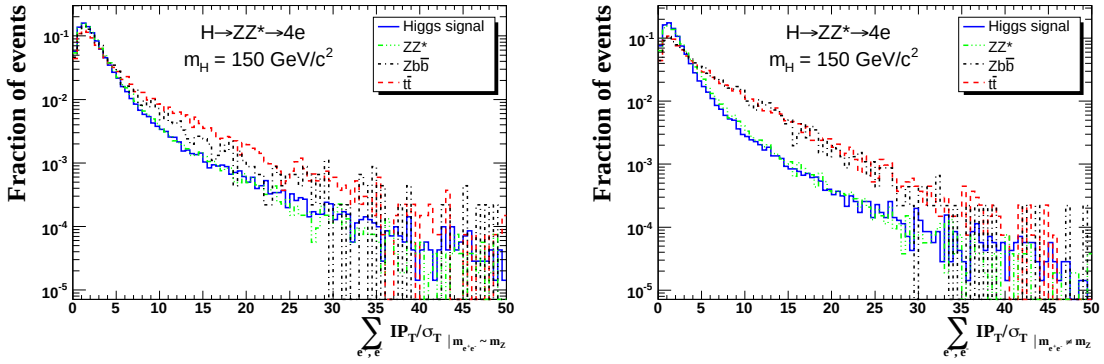

Figure 2.13: Sum of the transverse impact parameter significance (IP_T/σ_T) of e^+e^- pairs for a Higgs boson at $m_{\mathrm{H}} = 150\,\mathrm{GeV/c^2}$, for the ZZ$^{(*)}$ continuum, and for $\mathrm{Zb\bar{b}}$ and $\mathrm{t\bar{t}}$; a) $\sum IP_T/\sigma_T$ from the electrons of the e^+e^- pair with a reconstructed mass m_{ee} best matching the Z boson mass; b) $\sum IP_T/\sigma_T$ from the second e^+e^- pair.

are most likely to appear in the softest pair of reconstructed electrons. A best rejection power is obtained by imposing $\sum IP_T/\sigma_T < 30$ for the pair with $m_{ee} \simeq m_{\mathrm{Z}}$ and a more stringent cut of $\sum IP_T/\sigma_T < 15$ for the other pair.

Another powerful discriminant against secondary electrons in b jets or in general against fake electrons in QCD jets, is provided by isolation requirements. The electrons coupled to the Z or $Z^{(*)}$ in the $H \rightarrow 4e$ channel are expected to be on average well isolated from hadronic activity. Hadronic activity in single Higgs boson production appears in NLO processes, in the recoil against the Higgs boson. The Higgs boson itself generally receives a significant longitudinal boost in the laboratory reference frame but, as a scalar, decays uniformly in its centre-of-mass reference frame. In contrast, the electrons in the b jets from $t\bar{t}$ or $Zb\bar{b}$ are accompanied by significant hadronic activity.

Two partly complementary observables can be best used for the isolation of low p_T^e electrons. These rely either on measurements of primary tracks or on the energy flow in the hadronic calorimeter (HCAL). Both observables are insensitive to the eventual electron-induced electromagnetic showering in the tracker material. For the "track isolation", an isolation cone

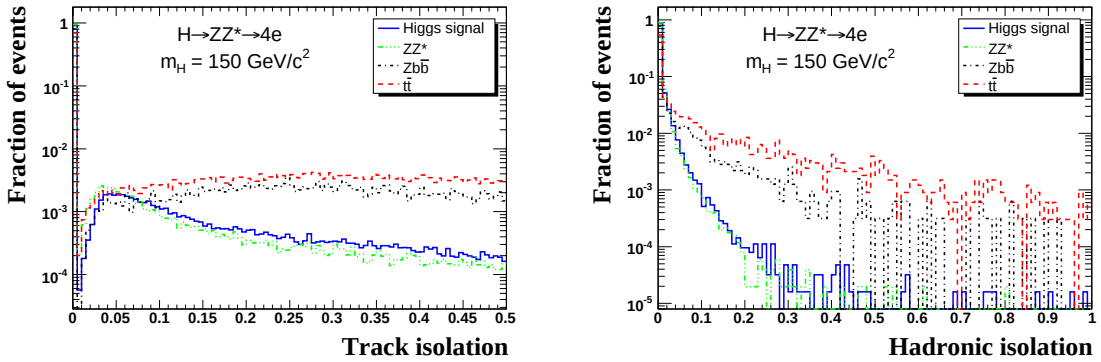

Figure 2.14: Electron isolation observables for the signal at $m_{\mathrm{H}} = 150\,\mathrm{GeV/c^2}$ and the SM backgrounds; a) track isolation, $\sum_{cone} p_T^{\mathrm{tracks}}/p_T^e$; b) hadronic isolation, $\sum_{cone} E_T^{HCAL}/p_T^e$, for the second least isolated electrons.

of size $\Delta R = \sqrt{\Delta\eta^2 + \Delta\phi^2} = 0.2$ is defined around the electron direction, and tracks with $p_T > 1.5\,\mathrm{GeV}/c$ originating from the same primary vertex within $|\Delta IP_L| < 0.1$ cm are considered. To avoid suppressing signal events, tracks attached to an electron candidate of opposite charge, and giving $m_{e^+e^-} > 10\,\mathrm{GeV/c^2}$, are discarded. All the 4 electrons from the Higgs boson candidate events must satisfy $\sum_{cone} p_T^{\mathrm{tracks}}/p_T^e < 0.1$. Distributions of this track isolation observable are shown in Fig. 2.14a. For the "hadronic isolation", all HCAL towers in an isolation cone size as above, and contributing with $E_T > 0.5\,\mathrm{GeV}$ are considered in the ratio $\sum_{cone} E_T^{HCAL}/p_T^e$. This ratio is required to be below 0.05 for at least three electrons. The cut is relaxed to 0.2 for the fourth electron. Distributions of this hadronic isolation observable are shown in Fig. 2.14b.

Further electron identification requirements must be imposed to suppress the possible background, involving "fake" electrons, from Drell-Yan processes at NLO where a Z^* recoils against jet(s). Different electron identification cuts are used depending on the distinct classes of track-supercluster electron patterns [46] in order to preserve the electron detection efficiency at all η^e. More details can be found in Ref. [37]. This tightening of the electron identification entails an absolute efficiency loss for the Higgs boson signal below 5%.

2.2.3.2 Kinematics

The cascade $H \rightarrow ZZ^{(*)} \rightarrow 4e$ for a Higgs boson, mostly produced at small transverse momentum, leads to very distinctly ordered p_T^e spectra for the four final state electrons. Moreover, the p_T^e spectra of the softest electrons for the Higgs boson signal is on average harder than the one expected from secondary electrons from the $Zb\bar{b}$ or $t\bar{t}$ backgrounds. Thus, it is advantageous to profit from the knowledge of the expected p_T^e distributions for the Higgs boson signal. A best set of p_T^e cuts as a function of m_H is given in Table 2.11.

Table 2.11: Electron p_T cuts, from the lowest to the highest p_T electron and reconstructed Z_1 and Z_2 invariant mass cuts.

m_H	p_T^1	p_T^2	p_T^3	p_T^4	$m_{Z_1}^{min}$	$m_{Z_1}^{max}$	$m_{Z_2}^{min}$	$m_{Z_2}^{max}$
(GeV/c^2)	(GeV/c)				(GeV/c^2)			
115	7	10	10	15	51	101	10	50
120	7	12	15	15	51	101	10	50
130	7	12	15	15	61	101	10	60
140	7	12	15	15	71	101	10	65
150	7	12	15	15	71	101	15	65
160	7	15	15	15	71	101	15	70
170	7	15	15	20	81	101	20	80
180	7	15	15	20	81	101	30	90
190	7	15	20	30	81	101	40	100
200	7	15	25	30	81	101	40	110
250	7	20	40	50	51	131	20	200
300	7	30	40	60	51	131	15	300

The cut on the softest electron is maintained to a lowest value for simplicity and to preserve the signal efficiency at low m_H. Otherwise the p_T^e cuts are seen to slowly evolve for as long as $m_H < 2m_Z$ and then rise faster above the Z pair production threshold. The p_T^e cuts lead for example [37] to a reduction by a factor of 5 to 10 of the $Zb\bar{b}$ background, and a factor of 3 to 5 of the $t\bar{t}$ background for $m_{4e} < 2m_Z$. Both backgrounds are also heavily suppressed above $2m_Z$.

Labelling Z_1 the boson reconstructed with an m_{ee} closest to the nominal Z mass and Z_2 the one reconstructed from the second e^+e^- pair, one expects for $m_{4e} < 2m_Z$ in the case of the Higgs boson signal that $m_{4e} \simeq m_{Z_1} + m_{Z_2}$ with most often the presence of a Z boson on its mass shell, $m_{Z_1} \simeq m_Z$. The Z boson masses saturate the phase space and are dominantly produced with small velocity in the Higgs boson rest frame. The requirement of one real Z boson suppresses further the $t\bar{t}$ backgrounds for low m_{4e}. The cut on Z_2 is powerful against the $ZZ^{(*)}$ continuum and further suppresses the $Zb\bar{b}$ and $t\bar{t}$ backgrounds. A set of optimal Z_1 and Z_2 cuts is given in Table 2.11 as a function of m_H. The cuts lead for example [37] for $m_{4e} \simeq 150\,\text{GeV/c}^2$ to a reduction of the $ZZ^{(*)}$ continuum by a factor of about 6.5 and a reduction of the $t\bar{t}$ background by a factor of about 2.5.

Figure 2.15a shows as an illustration the expected m_{4e} invariant mass distributions for the signal at $m_H = 150\,\text{GeV/c}^2$ and for backgrounds after triggering and pre-selection. The further background suppression from the isolated primary electron requirement, the p_T^e and Z mass cuts is seen by comparison in Fig. 2.15b. The global selection efficiency (normalised to the acceptance defined at the generation level) is given in Table 2.12 for the signal and backgrounds. Figures 2.15c and 2.15d show for illustration the possible outcome of two random Monte Carlo experiments corresponding to favourable and less favourable fluctuations of

Table 2.12: Summary of selection efficiencies normalised to the generation pre-selection efficiency.

m_{H}	115	120	130	140	150	160	170	180	190	200	250	300
						(GeV/c^2)						
Signal	24.3	26.0	31.2	35.2	36.0	37.4	38.0	39.9	40.9	42.5	41.2	38.6
$ZZ^{(*)}$	5.24	4.94	5.68	5.95	5.14	5.23	6.87	17.8	25.1	26.2	22.3	13.9
$Zb\bar{b}$	0.22	0.16	0.17	0.17	0.16	0.10	0.097	0.068	0.037	0.031	0.013	0.001
$t\bar{t}$	0.054	0.044	0.043	0.033	0.032	0.022	0.021	0.011	0.008	0.008	0.013	0.006

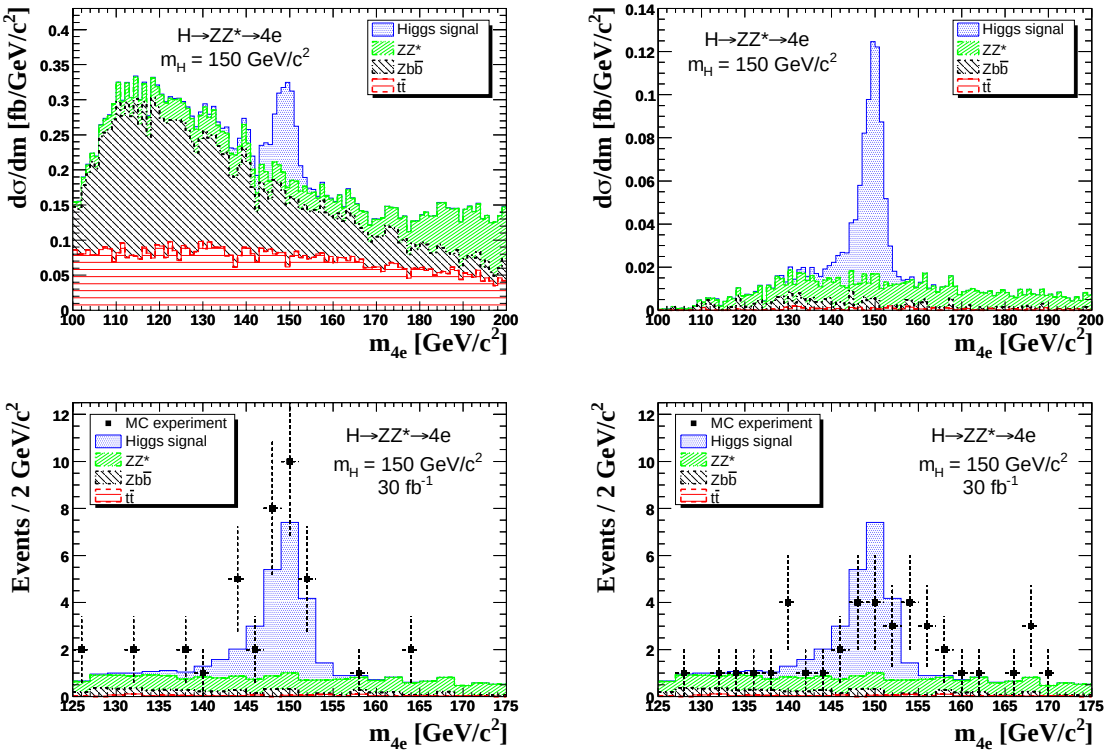

Figure 2.15: Distributions of the reconstructed invariant mass m_{4e} for the SM Higgs bosons signal at $m_{\mathrm{H}} = 150\,\mathrm{GeV/c^2}$ and for the SM backgrounds after (a) pre-selection step and (b) after all cuts. The number of events are normalised in cross-section. Single Monte Carlo experiments corresponding to an integrated luminosity of $30\,\mathrm{fb}^{-1}$ for (c) a favourable case and (d) a less favourable case.

the Higgs boson signal for an integrated luminosity of $30\,\mathrm{fb}^{-1}$. The Poissonian probability to have equal or more favourable (respectively equal or less favourable) fluctuations is of about 5% for the example cases shown.

2.2.4 Systematics

In this section the systematic errors are discussed in the context of a discovery via a simple event counting method. The "theoretical" and "experimental" sources of errors are distinguished. The theoretical uncertainties concern the estimation of the background rates within the cuts defining the acceptance of the Higgs boson signal and are discussed in Sec-

tion 2.2.4.1. The experimental uncertainties take into account the limited knowledge of the detector responses and efficiencies, and of the corresponding Monte Carlo modelling. These are discussed in Section 2.2.4.2. A comparison of different methods for the control of background systematics is presented in Section 2.2.4.3.

2.2.4.1 Theoretical errors

The theoretical uncertainty on the number of background events in the signal region from PDFs and QCD scales variations has been estimated by the MCFM program [47]. CTEQ6M PDF are used and 20 eigenvector parameters have been varied by $\pm 1\sigma$. Both QCD normalisation and factorisation scales have been varied independently up and down for a factor two from their nominal values of $2m_{\mathrm{Z}}$. The resulting uncertainties from PDF and QCD scale are of the order of 6% for direct estimation of ZZ background, from 2 to 8% for normalisation to single $\mathrm{Z} \to 2e$, and from 0.5 to 4% for the normalisation to sidebands (discussed further in Section 2.2.4.3). The gluon fusion cross-section uncertainties in the ZZ background of 8% is also considered as a part of theoretical uncertainties.

The uncertainty on the normalisation of the measurements to the pp luminosity of the LHC collider is estimated to be of the order of 3% for an integrated luminosity above $10\,\mathrm{fb}^{-1}$.

2.2.4.2 Experimental errors

The main remaining sources of experimental systematics expected in the CMS experiment after having collected of $\mathcal{O}(10)\,\mathrm{fb}^{-1}$, and relevant for the $\mathrm{H} \to 4e$ channel, originate from uncertainties on knowledge of the amount of tracker material in front of the ECAL, from the precision of the (pattern dependent) energy calibration of electron objects, and from the control of electron efficiencies. The strategy adopted consists of relying on experimental data, and in particular on single Z and W production, to minimise these systematic errors. The electrons from $\mathrm{W} \to e\nu$ and $\mathrm{Z} \to ee$ decays are used to control the energy measurements and reconstruction efficiencies.

A change of the integral amount of tracker material traversed by electrons before reaching the ECAL is susceptible of affecting the electron selection and identification efficiencies, as well as energy measurement scales and resolution. The uncertainty on the material budget will limit the precision of the acceptance calculations, when using the Monte Carlo model to extrapolate away from the kinematic domain best constrained via single Z and W measurements.

There are many observables that are directly or indirectly sensitive to the amount of tracker material, and that have been used in collider experiments. Examples are the distribution of converted photon vertices, or the shape of the E/p comparing tracker momentum measurement p to the energy E measured in the calorimeter in finite cluster volume, or a comparison of data and Monte Carlo for the Z mass resolution, etc. A new technique is used which is based on the electron GSF tracking introduced recently in Ref. [46]. The difference between the momentum magnitude at vertex and at the last hit, $p_{\mathrm{in}} - p_{\mathrm{out}}$, is a measure of the integral amount of bremsstrahlung. The mean fraction f_{brem} of the energy radiated along the complete trajectory is roughly proportional to the integral amount of material traversed. Hence, one can relate f_{brem} to the material thickness X/X_0 where X_0 is the characteristic radiation length via the formula $< X > /X_0 \simeq -\ln(1 - f_{\mathrm{brem}})$, where $f_{\mathrm{brem}} = (p_{\mathrm{in}} - p_{\mathrm{out}})/p_{\mathrm{in}}$.

The amount of tracker material measured in this way for single electron data is shown in

Fig. 2.16a. The results obtained in the configuration corresponding to the nominal tracker material coincide very well with the known material distribution as given in Ref. [7]. Figure 2.16b shows the ratio of the measured material thickness obtained in configurations where the amount of material was changed by $\pm 10\%$, normalised to the measurement results in the nominal case. The ratio is found to be remarkably stable as a function of η, despite the fact that the integral amount of material has a strong η dependence. Thus, single electrons can be used in CMS to tune the Monte Carlo model of the tracker material per η slice. Figure 2.16c shows that in a given η slice the measured material thickness is linearly correlated to a change (at least within a range of $\pm 10\%$) of the true material thickness. Similar results are obtained when considering various restricted range of p_T^e within a sample of uniformly distributed electrons in the p_T^e range from 5 to 100 GeV/c. With the electron statistics expected from single Z production for an integrated LHC luminosity of $\mathcal{O}(10)$ fb^{-1}, it should be possible to determine the tracker material thickness to a precision better than 2% over the full acceptance in η. Figure 2.16d shows that such a 2% uncertainty on the material budget will have almost no effect on electron reconstruction efficiency.

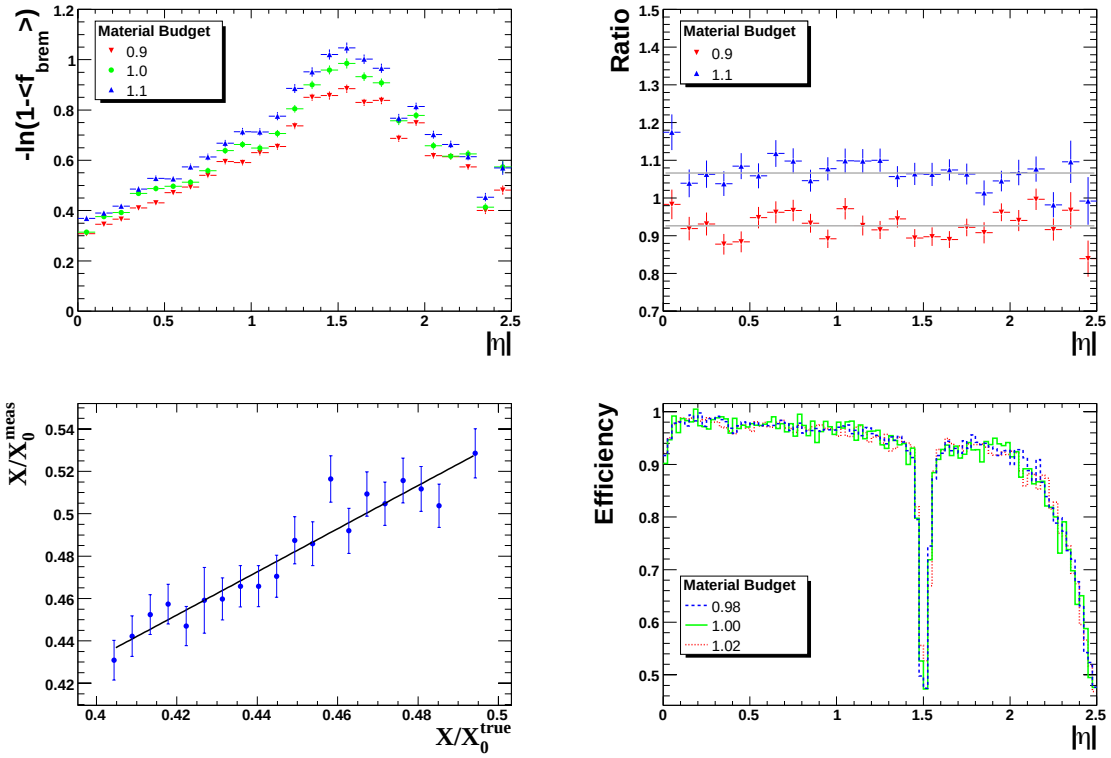

Figure 2.16: Sensitivity to variations of the tracker material budget from electron measurements based on GSF tracks; a) measured amount of material as a function of $|\eta|$ for the nominal tracker configuration and for an integral material budget changed by $\pm 10\%$; b) ratio of the measured mount of material as a function of $|\eta|$; c) measured versus true thickness in X_0 of the tracker material; d) effect of a change of 2% of the material budget on the electron reconstruction efficiency.

Electron reconstruction efficiencies and energy scales will be controlled by electrons from W $\rightarrow e\nu$ and Z $\rightarrow ee$ decay. Huge cross-sections of these two processes will allow for a significant reduction of reconstruction uncertainties already after few fb^{-1}. Electrons from Z $\rightarrow ee$

are produced centrally with a characteristic Jacobian p_T distributions around $45\,\mathrm{GeV}/c$. It is therefore expected that the best control of experimental systematics is obtained in the central part of the detector and for electrons around the Jacobian peak.

Electron reconstruction uncertainties as a function of η and p_T are given in Fig. 2.17a and Fig. 2.17b respectively, for an integrated luminosity of $0.15\,\mathrm{fb}^{-1}$. The expected behaviour of increased uncertainties when moving away from the Jacobian peak or from the central η region can be clearly seen. From the expected reconstruction errors evolution with the luminosity, all reconstruction efficiency uncertainties can be safely absorbed in a single factor of 1% per electron, for integrated luminosities larger than $10\,\mathrm{fb}^{-1}$.

The second important systematic effect is the uncertainty on the energy scale determination. Using single Z production, it has been shown in Ref. [48] that the absolute energy scale for electrons can in principle be controlled with great precision with average uncertainties reaching values below 0.1 %. The systematic uncertainty has to be studied as a function of p_T^e and η^e given the different electron spectrum in $\mathrm{H} \to \mathrm{ZZ}^{(*)} \to 4\mathrm{e}$ and $\mathrm{Z} \to ee$ decays. The reachable precision depends on the amount of integrated LHC luminosity. In this analysis, the second leg of a Z boson decay, tagged as an electron by imposing stringent electron identification requirements on the first leg combined with a kinematic constraint to the Z boson mass, is used as a probe to estimate systematics on the energy scale.

Uncertainties versus η and p_T for golden and showering electrons are shown in Fig 2.17c and Fig 2.17d, for the integrated luminosity of $0.15\,\mathrm{fb}^{-1}$. With expected evolution of these uncertainties with the luminosity, it is found that an uncertainty in energy scale of 0.5% in the barrel region, and 1% in the endcaps, for integrated luminosities larger than $10\,\mathrm{fb}^{-1}$, can be safely considered.

2.2.4.3 Control of background rates

Following the primary and isolated electron selection and the application of basic kinematic requirements, only the $\mathrm{ZZ}^{(*)}$ continuum remains as the dominant or sole background over the full mass range in consideration for the SM Higgs boson search. Thus, the determination of the mean expected number of SM $\mathrm{ZZ}^{(*)}$ background events in the signal region, defined e.g. by a simple sliding window in the m_{4e} spectrum, remains as a key issue.

The three main methods for the estimation of $\mathrm{ZZ}^{(*)}$ continuum contribution to the background in the signal region are:

- direct simulation of the $\mathrm{ZZ}^{(*)} \to 4e$ process,
- normalisation to the $\mathrm{Z} \to 2e$ data,
- normalisation to the sidebands.

The first method entirely relies on existing SM constraints and the theoretical knowledge, with uncertainties coming from the PDFs used to describe the colliding protons and from QCD scale variations. It furthermore is reliant on the LHC luminosity uncertainties, and on the Monte Carlo modelling of the acceptance and detector response for the uncertainties arising from electron reconstruction and selection. Otherwise, the method potentially benefits from the fact that the statistical precision on the mean background expectation is only limited by the Monte Carlo statistics, and can therefore be assumed negligible in the context of a prospective for an analysis to be performed in a future CMS experiment.

The second method aims at profiting from the fact that the SM single Z production cross-

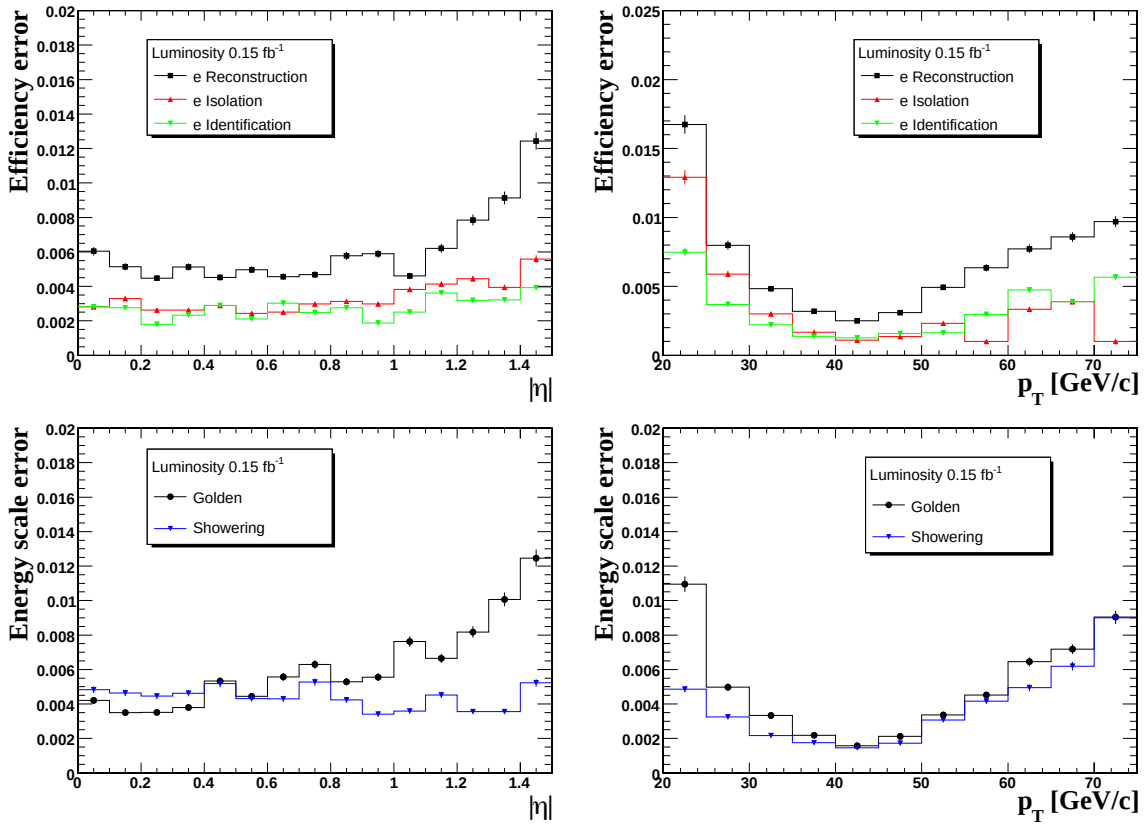

Figure 2.17: Control of experimental uncertainties using SM data; uncertainties on measurements of electron reconstruction, isolation and identification as a function of (a) η and (b) p_T; uncertainties on measuring the energy scale for golden and showering electrons as a function of (c) η and (d) p_T.

sections is measured with great precision in an experiment which will have integrated a luminosity of $\mathcal{O}(10)\,\text{fb}^{-1}$ at the LHC. Using the ratio of ZZ$\to$ $4e$ to Z $\to$ $2e$ rates allows to profit from a full cancellation of pp luminosity uncertainties, while providing a partial cancellation of PDF and QCD scale variations uncertainties (due to their correlations in a part of the initial state phase space) and a partial cancellation of experimental uncertainties.

In the method of the normalisation from sidebands, the number of background events *inside* the acceptance of the signal region is determined from the number of background events measured *outside* the signal region, by multiplying the latter with the ratio α_{MC} between *inside* and *outside* expectations as determined using Monte Carlo simulation. Using the sidebands one also expects to fully cancel luminosity uncertainties, to reduce PDF and QCD scale variation uncertainties and substantially reduce experimental uncertainties too. Statistical errors with sidebands normalisation come from the statistics of the background rate outside the signal region and can be a limiting factor for the method. By relaxing some of late analysis cuts, such as invariant Z mass, the background events rate outside the signal region increases, reducing therefore statistical errors for this method. The price to pay is an increased background rate in the signal region too and, therefore, some balancing is needed.

Using results from previous sections, both theoretical and experimental uncertainties are evaluated for two methods: normalisation to the Z $\to$ $2e$ measurements and normalisation

to the sidebands. For the normalisation to single Z $\rightarrow 2e$ measurements results are shown in Figure 2.18a. The overall systematic uncertainty with this method is of about 5%. Ex-

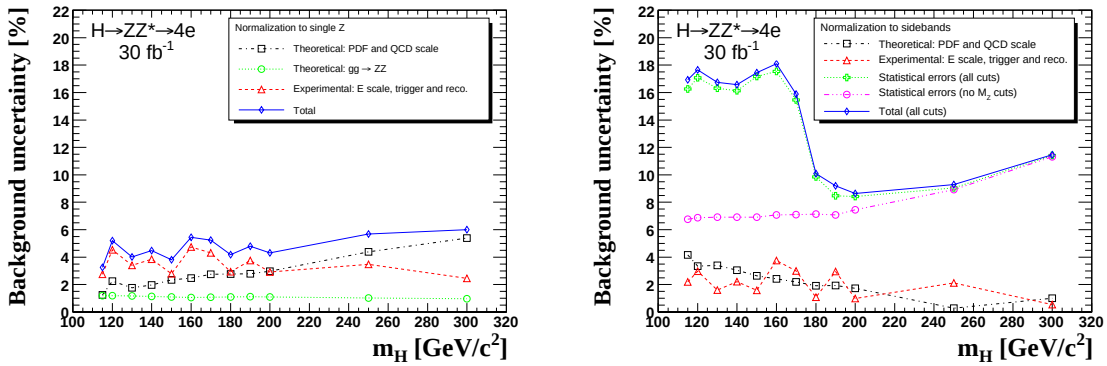

Figure 2.18: Theoretical and experimental uncertainty estimations for both methods for evaluation of background from data: (a) normalisation to the single Z $\rightarrow 2e$ measurements and (b) normalisation to the sidebands. Expected statistical errors for sidebands are also shown, for integrated luminosity of 30 fb^{-1}.

perimental uncertainties are seen to dominate for $m_{\mathrm{H}} \simeq 2m_{\mathrm{Z}}$ while theoretical errors take over above the pair production threshold. Uncertainties for the sidebands normalisation are shown in Figure 2.18b. Statistical uncertainties scale as the square root of the number of background events outside the signal region and are shown for an integrated luminosity of 30 fb^{-1} and for two analysis scenarios: after all analysis cuts and without cuts on the mass of both Z bosons. A trade-off in the second method is in a somewhat lower nominal significance (for about 8%) while statistical errors decrease by a factor of about 2.5. Full significance calculations with and without systematics and statistical uncertainties are presented in the following section.

2.2.5 $H \rightarrow 4e$ Observability, mass and cross-section measurements

2.2.5.1 Discovery reach

A simple counting experiment is used here to quantify the sensitivity of the experiment to the presence of a Higgs boson signal. The expected number of signal (N_S) and background (N_B) events are evaluated in a sliding window whose central position m_{4e} varies between 100 and 320 GeV/c^2. The size of the optimal window increases progressively from 6 GeV/c^2 at $m_{4e} = 115$ GeV/c^2 to 24 GeV/c^2 at $m_{4e} = 300$ GeV/c^2. The Table 2.13 presents for each Higgs boson mass hypothesis the mean expected number of signal and background events, and associated uncertainties.

The significance of the H $\rightarrow 4e$ signal observation is shown as a function of m_{H} in Fig. 2.19a as expected for an integrated luminosity of 30 fb^{-1}. The results are given for both the S_{cP} and the S_{cL} significance estimators. The S_{cP} is defined as the probability for a Poisson distribution with mean N_B to observe a number of events equal or greater than $N_S + N_B$, converted in the equivalent number of standard deviations of a Gaussian distribution. The S_{cL} corresponds to the widely used log-likelihood ratio significance [49] and is given for comparison. The effect of including experimental and theoretical systematics, described in section 2.2.4 and listed in Table 2.13, on the significance S_{cP} [50] is also shown, for two different methods

Table 2.13: Expected number of Higgs boson signal (N_S) and SM background (N_B) events for an integrated luminosity of 30 fb^{-1}, in the optimised window for the reconstructed invariant mass m_{4e}. The uncertainties (δN_B) are given for systematics from experimental (exp.) and theoretical (theo.) sources, for an analysis where the ZZ$^{(*)}$ continuum has been normalised to the measurement of single Z production.

m_{H}	115	120	130	140	150	160	170	180	190	200	250	300
						(GeV/c^2)						
N_S	1.52	2.97	8.18	15.80	17.19	8.38	3.76	9.95	34.05	38.20	27.68	21.69
N_B	2.26	1.94	3.71	4.31	3.68	3.10	3.37	6.42	14.62	17.29	13.40	7.63
δN_B												
exp. :	0.063	0.089	0.126	0.167	0.105	0.148	0.145	0.187	0.551	0.505	0.466	0.187
theo. :	0.039	0.049	0.079	0.098	0.095	0.084	0.100	0.191	0.440	0.549	0.602	0.417

of controlling the background uncertainties. A signal observation with a significance above 3 standard deviations is expected in the H $\rightarrow$ $4e$ channel alone for m_{H} in the range from 130 to 160 GeV/c^2, and above 180 GeV/c^2. The integrated luminosity needed for a 5 standard

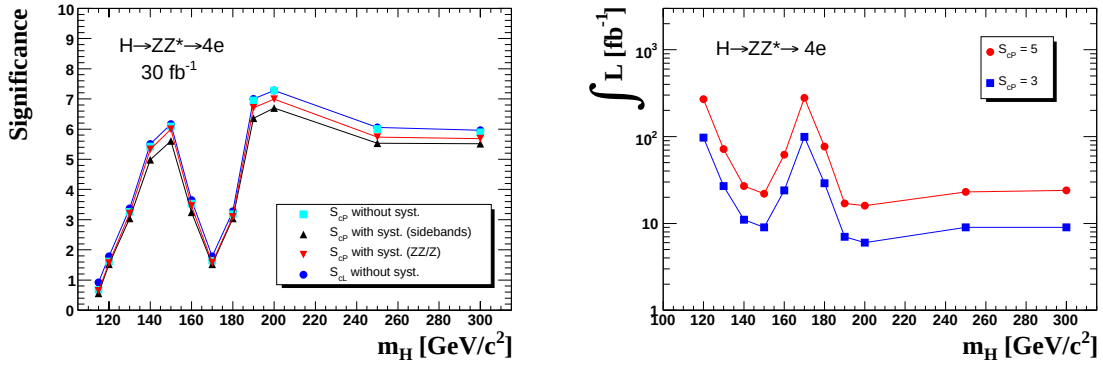

Figure 2.19: (a) Significance S_{cP} for an integrated luminosity of 30 fb^{-1} as a function of the Higgs boson mass without and with systematics included in both options of ZZ$^{(*)}$ normalisation to the measured sidebands or the measured single Z production cross-section. The significance S_{cL} is also shown. (b) luminosity needed for a 3σ observation and 5σ discovery with the systematics included using ZZ$^{(*)}$ normalisation to the Z cross-section.

deviations discovery of the SM Higgs boson in the H $\rightarrow$ $4e$ channel alone is also shown as a function of m_{H} in Fig. 2.19b. Systematic errors from normalisation to the Z cross-section have been included.

2.2.5.2 Mass, width and cross-section measurements

At an early stage of the Higgs boson search and discovery in the H $\rightarrow$ $4e$ channel, given very low statistics, a robust and simple estimation of m_{H} can be obtained by a simple mean (or weighted mean) of the m_{4e} values measured for individual events. The events falling in the pre-defined optimal mass window introduced in the above Section 2.2.5.1 and used to establish the signal significance, can be used for such purposes. For higher statistics, a fit of the m_{4e} mass distribution to a signal plus background shape can be used to extract simultaneously the mass and the cross-section $\times$ branching ratio of a Higgs boson signal. Detector

effects dominate the Higgs boson mass resolution below the Z pair production threshold and a sensitivity to the Higgs boson intrinsic width is expected only for masses well above $2m_{\mathrm{Z}}$.

The precision on the parameter measurements for the Higgs boson depend on the quality of the reconstructed electrons and can, in general, be improved using event-by-event errors on the electron momentum estimation [46]. Example cases for two different sub-samples of Higgs boson events differing by the pattern of the four reconstructed electrons are presented in Fig. 2.20. Clearly, event candidates built from four *non-showering* electrons in the barrel

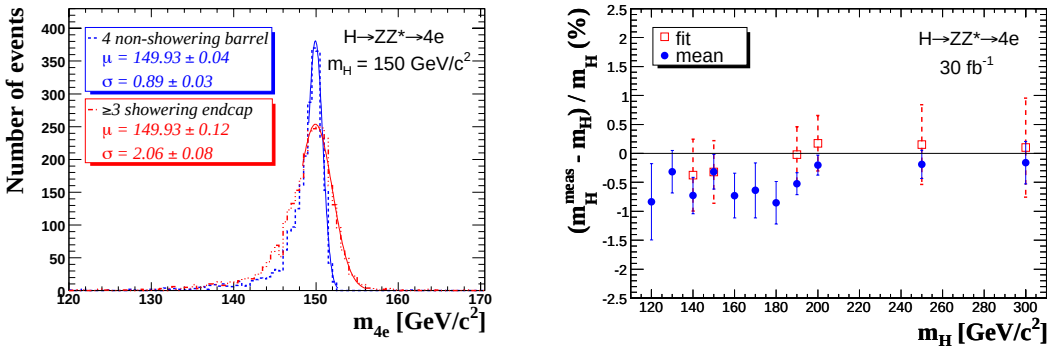

Figure 2.20: Mass measurements: (a) example case for two different event sub-samples differing by the pattern of the four reconstructed electrons; (b) relative errors as a function of the Higgs boson mass using the mean mass and the fitted mass as obtained for an integrated luminosity of 30 fb^{-1}.

part of the ECAL, a subset representing only about 1.76% of all signal events, allow for a much better m_{H} measurement (smallest errors on average and least dispersion of the mass measurement errors) than candidates built mainly from e.g. *showering* electrons in the endcaps part of the ECAL. About 36.7% of the signal event candidates contain three or more *showering* electrons. A weighted mean of the events of the m_{4e} distribution falling in the signal window has been considered for the estimation of the Higgs boson mass in Ref. [37]. A simple mean can be also used for simplicity.

The reconstructed Higgs boson mass and its error obtained from the mean value for events falling in the expected signal window is presented in Fig. 2.20b. The error is obtained from the dispersion of the mean values obtained from large number of Monte Carlo experiments at an integrated luminosity of 30 fb^{-1}. The results are shown as a function of the Higgs boson mass. The systematic bias on the mass estimate for the low m_{H} cases for this simple mean approach is due to the asymmetric shape of the reconstructed signal and can be modelled. In the mass ranges where the Higgs boson signal significance exceeds 3 standard deviations, the uncertainty on the mass determination is found to be everywhere below 0.4%. Its reaches values below 0.2% for $m_{\mathrm{H}} \simeq 200\,\mathrm{GeV/c^2}$. For comparison, results obtained by fitting the m_{4e} distribution are also shown. The fit method requires a significant number of events (typically $\gtrsim \mathcal{O}(10)$) to converge and provide reasonably stable results. The m_{4e} distribution is fitted by a signal plus background shape. The signal contribution is modelled with two Gaussians, describing respectively the core and the low m_{4e} tail of the signal distribution. The tail parameters (fraction, mean and dispersion) are fixed by fitting the "signal only" expectation. The background is modelled using a flat distribution up to about $m_{4e} \approx 2m_{\mathrm{Z}}$ and a linear function (non-zero slope) for higher Higgs boson masses. This has been found

to provide a sufficiently good model of the observation in a restricted mass range around the signal region. A likelihood fit is then performed on each Monte Carlo experiments and the reconstructed mass and precision are extracted from the distribution of the fitted values of the peak of the Gaussian core. Where the fit can be performed, Fig. 2.20b shows that an unbiased estimation of m_{H} is obtained within errors.

The fitted number of signal events is used to estimate the production cross-section by correcting for the global acceptance efficiency. The statistical precision on this measurement is here also obtained from the width of the distribution of the fitted parameters in Monte Carlo experiments. An unbiased measurement of the cross-section is obtained over the full mass range considered here, with a precision of the cross-section measurement between 20 and 30%. With such a precision, the influence of the detector systematics (about 5%) and of the uncertainty on the luminosity measurement (less than 3% for 30 fb^{-1}) is marginal. For an integrated luminosity of 60 fb^{-1}, the precision on the cross-section measurement improves to about 15%.

A measurement of the width is possible only for Higgs boson masses above $\gtrsim 2m_{\mathrm{Z}}$ where at the same time the Higgs natural width is becoming large and the detector resolution is improving. A Gaussian width with central values of about 2.3 GeV/c^2 for $m_{\mathrm{H}} = 200$ GeV/c^2 and 4.2 GeV/c^2 for $m_{\mathrm{H}} = 300$ GeV/c^2 is obtained from the fit, but with a rather large uncertainty of about 50%.