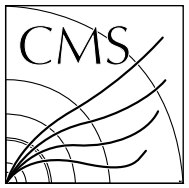
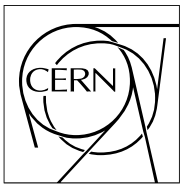
Available on CMS information server

CMS AN-2008/082



The Compact Muon Solenoid Experiment

Analysis Note



24 September 2008

A cut based method for electron identification in CMS

J. Branson, M. Gallinaro, P. Ribeiro, R. Salerno, M. Sani

Abstract

A cut based method for identification of electrons in CMS is presented. Different levels of electron identification with increasing purity are proposed. A Fixed Threshold level uses variables that are stable against a misaligned tracker and mis-calibrated ECAL, therefore expected to be well measured during early data taking. For tighter levels of identification, the electrons are categorized according to the estimate of energy fraction emitted by bremsstrahlung in the tracker material and the ratio of the energy collected in the electromagnetic calorimeter to momentum of the electron track. Simple sequential cuts are set according to the estimated ratio of signal to background events in one of each categories. The performance of the method is evaluated using Monte Carlo simulated samples of benchmark signal and background datasets for different misalignment and miscalibration scenarios.

Contents

1 Introduction

In physics analysis with multi-electron final states high identification efficiency is needed to enhance signal selection, in particular at low E_T where the background increases and the fake rate is much higher. The CMS detector has some unusual features that greatly impact on electron identification like the high magnetic field, a thick tracker and the lower ECAL response to pions with respect to electrons.

At the LHC startup period we would like a robust and simple identification until we have data to verify and tune the selection criteria. For those reasons the selection should rely on the most predictable and stable electron variables. For tighter identification a selection based on separate electron classes based on expected signal over background ratio should be used. More complicated and multivariate analysis could be used later when the detector will be better understood. Our goal is to design electron identification criteria aiming to 97% selection efficiency and a relative small fake rate.

Electrons can be categorized according to ECAL and tracker properties [1]. We propose a classification based only on the comparison between the momentum measured at the vertex (p_{in}) and at the ECAL (p_{out}).

This note describes the implementation of the cut based electron identification and it is organized as follows. We first describe the method focusing on two proposed techniques: fixed thresholds and categories based, each one with two level of identification (loose and tight). Then the performance of the cut based electron identification evaluated using Monte Carlo simulated samples of benchmark signal (Z→ee and W→eν events) and background (QCD events) is showed. We also present a comparison of the selections with different scenarii to check how the different miscalibration and misalignment condition would affect identification performance.

2 Description of the Samples

The following Standard Model datasets have been analyzed: Z→ee, W→eν, dj-jet in different pt hat bins. They all were produced with `PYTHIA`. Events were fully simulated using Geant4, digitized without pile-up and they were reconstructed using `CMSSW_1_6_7`. All the samples have been produced during the Computing, Software and Analysis Challenge (CAS07) CMS production.

In particular being interested in the start-up conditions we have compared the results using the same physical samples re-reconstructed according the foreseen knowledge of the calibration and alignment conditions of the tracker and the electromagnetic calorimeter. The four considered scenarii are: start-up conditions, 10 pb^{-1}, 100 pb^{-1} and ideal conditions. In Table 2 the details for the samples are collected.

Channel	Dataset	Scenario
$Z \to ee$	/Zee/CMSSW_1_6_7-ReReco0pb-1197808248/RECO	0 pb^{-1}
$Z \to ee$	/Zee/CMSSW_1_6_7-ReReco10pb-1197913408/RECO	10 pb^{-1}
$Z \to ee$	/Zee/CMSSW_1_6_7-CSA07-1192835490/RECO	100 pb^{-1}
$Z \to ee$	/Zee/CMSSW_1_6_7-ReRecoIdeal-1198082306/RECO	Ideal
$W \to e\nu$	/Wenu/CMSSW_1_6_7-ReReco0pb-1197808300/RECO	0 pb^{-1}
$W \to e\nu$	/Wenu/CMSSW_1_6_7-ReReco10pb-1197999395/RECO	10 pb^{-1}
$W \to e\nu$	/Wenu/CMSSW_1_6_7-CSA07-1197047869/RECO	100 pb^{-1}
$W \to e\nu$	/Wenu/CMSSW_1_6_7-ReRecoIdeal-1198082363/RECO	Ideal
GUMBO	/CSA07AllEvents/CMSSW_1_6_7-CSA07-Tier0-A1-Gumbo/RECO	10 pb^{-1}
GUMBO	/CSA07AllEvents/CMSSW_1_6_7-CSA07-Gumbo-B1-PDAllEvents-ReReco-100pb/RECO	100 pb^{-1}

Table 1: List of the samples used to evaluate electron identification performance. Signal samples have been simulated according to four different miscalibration and misalignment scenarii while for background only two scenarii were available.

3 Fixed Threshold Identification

This kind of selection is aimed for early data taking. It has been designed to be as simple, efficient and robust as possible. No electron classification is involved here.

Selection is performed with straight cuts on the following four quantities:

- H/E: hadronic to electromagnetic energy ratio,

- $\sigma_{\eta\eta}$: this shape variable measure the eta extent of the electron Super Cluster. It is defined as follows:

$$\sigma_{\eta\eta} = \frac{\sum_i^{5\times5} w_i(\eta_i - \bar{\eta}_{5\times5})^2}{\sum_i^{5\times5} w_i} \tag{1}$$

  where the index i runs over all the crystals in a 5×5 block of crystals centered on the seed crystal, η_i is the η position of the i^{th} crystal, $bar\eta_{5\times5}$ is the energy weighted mean η of the 5×5 block of crystals and w_i is the weight of the i^{th} crystal and is defined as

$$w_i = 4.2 + ln(E_i/E_{5\times5}) \tag{2}$$

  where E_i and $E_{5\times5}$ are the energy of the i^{th} and 5× 5 block of the crystal respectively.

  $\sigma_{\eta\eta}$ is corrected in the endcap to take into account the different crystal geometry. The applied linear correction is defined by:

$$\sqrt{\sigma_{\eta\eta}} - 0.02 * (|\eta| - 2.3) \tag{3}$$

  where η is the pseudo-rapidity of the reconstructed electron. Actually in the selection the square root of this variable is used.

- $\Delta\eta_{in}$: η difference between the electron Super Cluster and the electron track at vertex,
- $\Delta\phi_{in}$: ϕ difference between the electron Super Cluster and the electron track at vertex.

Fig. 2 and Fig. 3 show the distribution for the different scenarii for the four used variables in barrel and endcap. They are quite stable to the change of the different misalignment/miscalibration scenarios even if some differences is clearly visible. $\Delta\eta_{in}$ distributions show a shift of the peak of about 500 μm. This is consistent with the foreseen misalignment of the pixel detector in the start-up conditions. A similar shift can be seen also in $\Delta\phi_{in}$ comparing 0 pb^{-1} and 10 pb^{-1} scenarii. H/E and $\sigma_{\eta\eta}$ don't show any substantial variation in any scenario and anyway all the variables look more stable in 100 pb^{-1} configuration approaching the ideal scenario conditions. It is important to note that in the `CMSSW_1_6_X` release, which has been used for this study, no zero suppression threshold on HCAL RecHits was applied as it will be in next releases. Therefore the effect of zero suppression on H/E variable should be tested. Furthermore the gaps or "cracks" between ajacent ECAL modules and super-modules affects the measurement of H/E. Electrons incident in one of the crack will deposit their energy in the un-instrumented ECAL support structures and into HCAL leading to a mismeasurement of H/E variable. Figure 1 shows the geometrical distribution of the electrons with H/E > 0.05 and how the electrons with higher H/E are those in the ECAL cracks. Anyway the effect is not so strong due to the soft p_T spectrum of the leptons coming from Z decay. In this analysis we do not take special care of the electrons in crack.

The Fixed Threshold method has two levels of identification: a loose set of cuts to deal with the identification of electrons coming from Z decays and a second one for W decay where a tighter identification is needed.

In first approximation we can assume that loose isolation will commute with electron identification so in the first part of this note we concentrate only on identification.

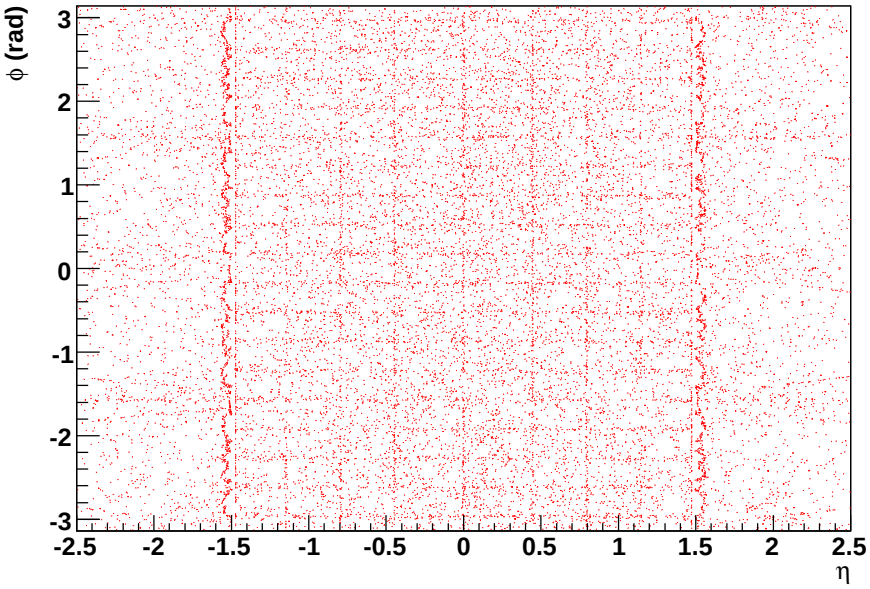


Figure 1: Distribution of the electrons with high H/E (>0.05) as a function of η and ϕ. The ECAL cracks are clearly visible in this picture.

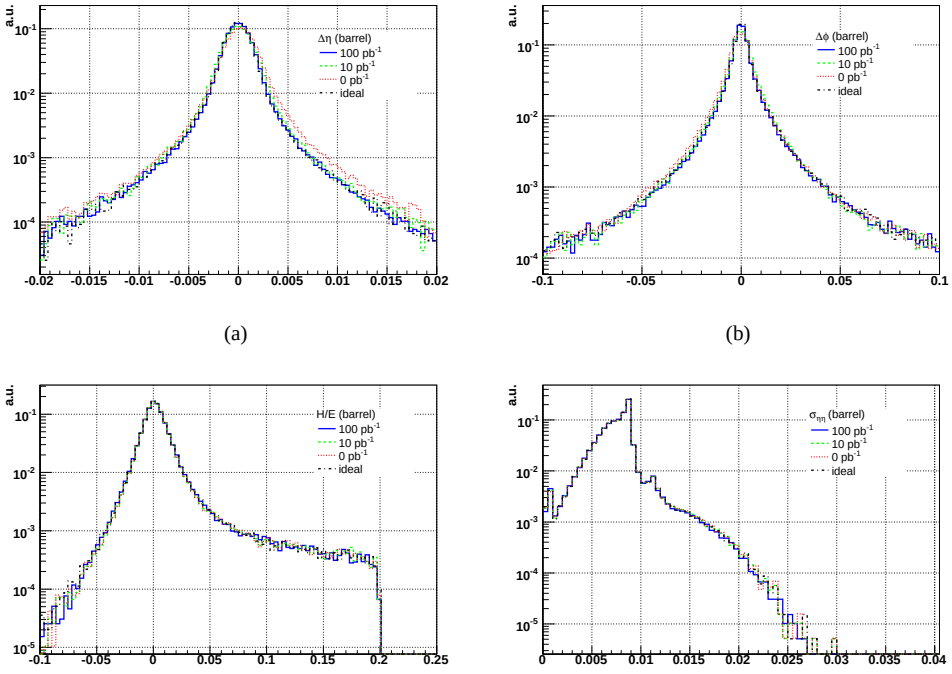
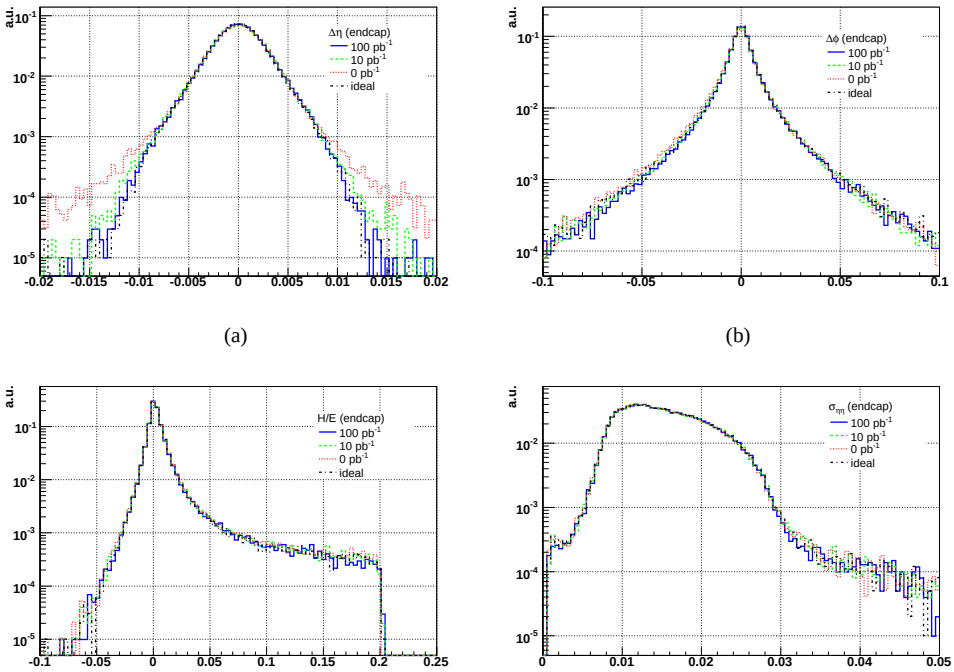


Figure 2: Distributions of the four variables used in the Fixed Threshold Identification for the four studied scenarii. In this pictures the barrel only distribution are shown: a) $\Delta\eta_{in}$ b) $\Delta\phi_{in}$ c) H/E d) $\sigma_{\eta\eta}$.



(a) (b)

(c) (d)

Figure 3: Distributions of the four variables used in the Fixed Threshold Identification for the four studied scenarii. In this pictures the endcap only distribution are shown: a) $\Delta\eta_{in}$ b) $\Delta\phi_{in}$ c) H/E d) $\sigma_{\eta\eta}$.

3.1 Loose identification

Two distinct sets of thresholds are used in the loose identification for barrel and endcap and Table 3.1 reports their values.

	barrel	endcap
H/E	0.115	0.150
$\sigma_{\eta\eta}$	0.0140	0.0275
$\Delta\eta_{in}$	0.0090	0.0105
$\Delta\phi_{in}$	0.090	0.092

Table 2: List of the thresholds used in the Loose Fixed Threshold identification. Two distinct set of cuts are used in barrel and endcap regions.

As it is clear from the Tab. 3.1, Fig. 2 and 3 these cuts affect only the tail of the distributions making this selection particularly independent of measured bremsstrahlung fraction and insensitive to tracker misalignment. For this reason this selection sometimes is referred to as "Robust".

Using simply these straight four cuts a 98% identification efficiency per electron in Z→ee channel can be achieved.

3.2 Tight Identification

In W→eν analysis some of the signal can be sacrificed to increase the purity of the selected electron sample thus reducing the uncertainty on the background estimation below a harmful level. This reduction is achieved by a tighter electron identification in conjunction with an appropriate isolation cut. For the moment we assume that isolation and identification commute so we concentrate only on the identification.

Thus a tighter set of cuts for the Fixed Threshold selection suitable for the W analysis has been proposed. As in the loose selection two distinct sets of thresholds are used for endcap and barrel and Table 3.2 reports their values. The Table shows clearly how now not only the tails of the distributions presented in Fig. 2 and Fig. 3 are involved in the cuts. So we expect a greater dependence of this selection to the different misalignment and miscalibration condition of the CMS detector. More studies on the Fixed Threshold Tight Identification have been carried on in the context of the W/Z Analysis [3].

	barrel	endcap
H/E	0.0150	0.0180
$\sigma_{\eta\eta}$	0.0092	0.0250
$\Delta\eta_{in}$	0.0025	0.0040
$\Delta\phi_{in}$	0.0200	0.0200

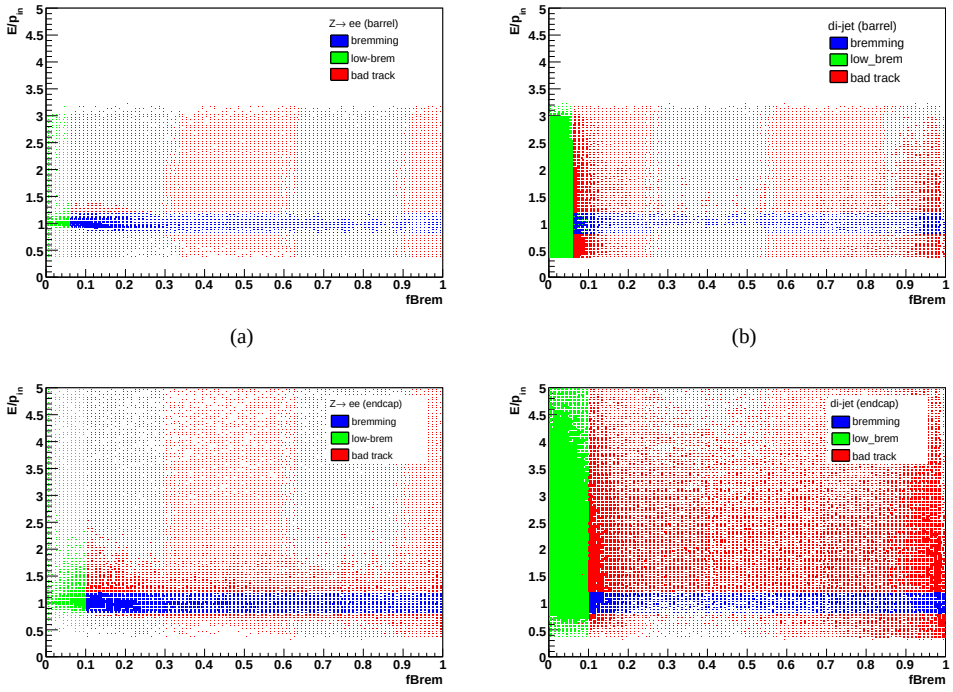
Table 3: List of the threshold used in the Tight Fixed Thresholds identification. Two different set of cuts are used in barrel and endcap regions.

4 Category Based Identification

The starting point for a more efficient electron selection is to introduce a classification of the electrons to achieve a good separation between real electrons and fakes. Then different set of cuts tuned on expected signal over background ratio can be used in each category.

A classification based on the E/p_{in} and fBrem ((p_{in}-p_{out})/p_{in}) properties of the electrons has been proposed for the Category Based Identification. Referring to Fig. 4 three classes are defined:

1. low-brem electrons: ($0.8 <$ E/p_{in} < 1.2, fBrem< 0.06 (barrel), fBrem < 0.1 (endcap)), fake-like region with high population from both real and fake electrons,
2. bremming electrons: ($0.8 <$ E/p_{in} < 1.2, fBrem> 0.06 (barrel), fBrem > 0.1 (endcap)), electrons-like region with little contamination from fakes,
3. bad track: (E/p_{in} > 1.2 or E/p_{in} < 0.8), region with not many real electrons.



(a) (b)

(c) (d)

Figure 4: Scatter plots which show fBrem as a function of E/p_{in} for signal barrel (a), background barrel (b), signal endcap (c) and background endcap (d). This plot has been used to determine electron classification.

The use of electron categories allows to take into account for the large differences in signal to background ratio in the regions of E/p_{in} vs fBrem plane and to put events of similar characteristics together (well measured electrons, electrons with track problems, fakes...). In Tab. 4 the relative electron population of the three categories is reported. It is clear that most of the electrons from signal go in first (good electrons) and last category (bad measured electrons) while di-jet events mostly populate low-brem category. So the robust electron/fake separation achieved with this classification (since pion tracks should have fBrem almost always 0) allows to use looser cuts for the best electrons while tighter cuts in the overlap region maintain robust rejection of fakes.

The physics motivations behind this kind of categorization rely on the fact that E/p is often well measured for electrons and it is not often measured to be less than 1 moreover electrons usually radiate a good dial of energy in the tracker. Contrary fakes from jets usually have fBrem around 0 because they are usually just charged pion tracks and many fakes from jets have E/p_{in} < 1 partly because of the low response of ECAL to charged pions.

Different miscalibration and misalignment conditions can determine a migration of electrons from one region to another. A re-tuning of the definition of the regions could be needed according to the current scenario, at the moment no detailed study has carried on this issue.

	barrel		endcap	
	signal	di-jet	signal	di-jet
bremming	55%	7%	37%	8%
low-brem	12%	74%	13%	66%
bad track	33%	19%	50%	26%

Table 4: Electron population of the three categories for signal and background. The numbers are calculated separately for barrel and endcap.

A set of five cuts on the following variables have been studied. Thresholds are different (where beneficial) for the 3 different classes:

- H/E, $\sigma_{\eta\eta}$, $\Delta\eta_{in}$, $\Delta\phi_{in}$,
- E_{seed}/p_{in}: the ratio between the energy of the SuperCluster seed and the electron track momentum measured at the vertex. This cut replaces the cut on E_{seed}/p_{out}, indeed we prefer not to use the p_{out} measurement which could be not too much reliable for electrons which brems a lot due to the increase of tracker hits in the final part of the track. The cuts are anyway set such that this only does much in the overlap category in which p_{out} and p_{in} are similar.

In Fig. 5 and 6 the discriminating variables for both signal and background are shown. Signal distribution have been split according to the three categories

As for the Fixed Threshold method two levels of identification are proposed. A looser one for electron coming from Z's and a tighter one specifically designed for W analysis.

4.1 Loose Identification

The loose identification has been primarily designed to have enough fake rejection to clearly extract Z→ee signal from the QCD background.

Two distinct sets of thresholds are used in Loose Category Based identification for barrel and endcap and Table 4.1 reports their values.

		bremming	low brem	bad track	$E/p_{in} > 1.5$
H/E	(barrel)	0.115	0.10	0.055	
	(endcap)	0.145	0.12	0.15	
$\sigma_{\eta\eta}$	(barrel)	0.014	0.012	0.0115	
	(endcap)	0.0275	0.0265	0.0265	
$\Delta\eta_{in}$	(barrel)	0.009	0.0045	0.0085	
	(endcap)	0.0105	0.0068	0.010	
$\Delta\phi_{in}$	(barrel)	0.05	0.025	0.053	0.09
	(endcap)	0.07	0.03	0.092	0.092
E_{seed}/p_{in}	(barrel)	0.11	0.91	0.11	
	(endcap)	0.	0.85	0.	

Table 5: Thresholds for the Loose Category Based identification. The cut values are different for each category and are further split for barrel and endcap electrons.

A looser $\Delta\phi_{in}$ cut is used for electrons with high E/p_{in}. As shown later the Loose Category Based selection allows a bigger fake rejection with a slight loose in the efficiency with respect to the Loose Fixed Thresholds identification.

4.2 Tight Identification

As in the case of the Fixed Threshold selection a tighter identification is needed in W analysis. Thus a tighter level of identification has been designed. It basically uses the same set of cuts as the loose identification. In addition the bottom left triangle of the E/p_{in} vs fBrem plot is cut out by requiring: $E/p_{in} > 0.9 * (1-\text{fBrem})$

Two distinct sets of thresholds are used in the selection for endcap and barrel and Table 4.2 reports their values.

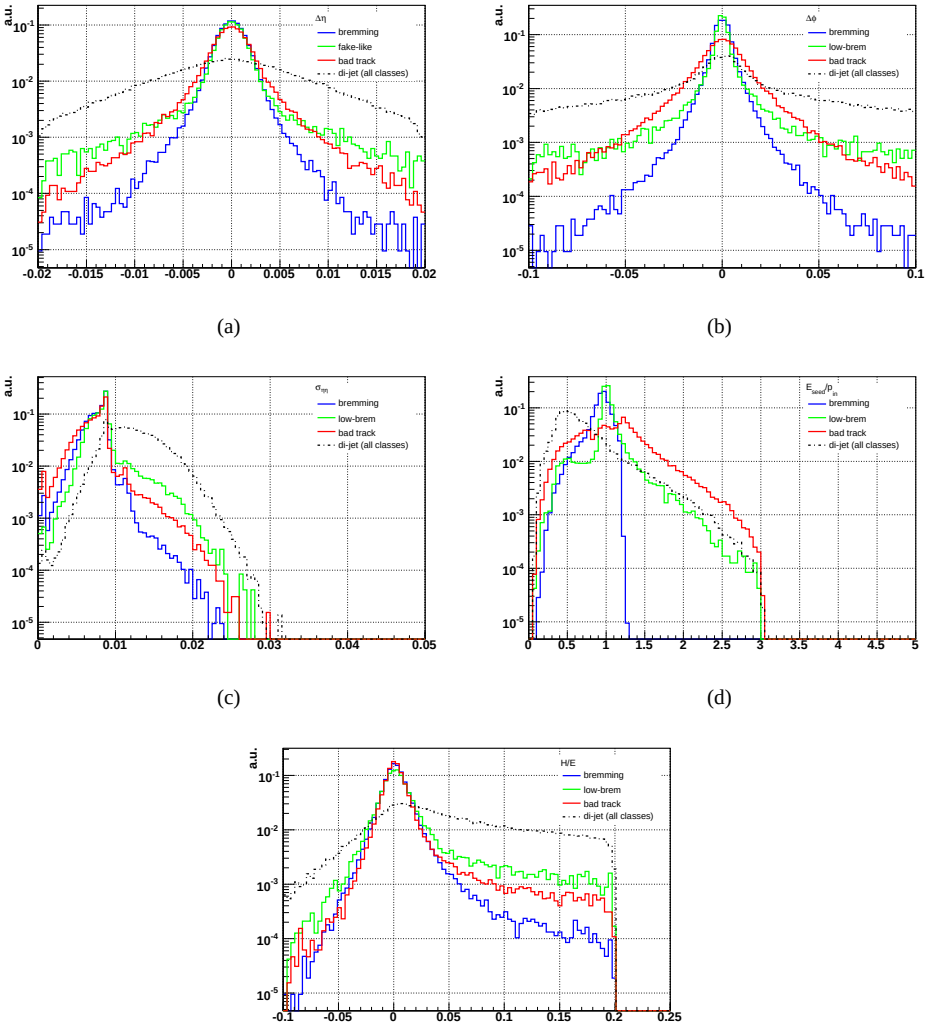


Figure 5: Discriminating variables used in the tight and loose Category based electron identification for signal and background. The distributions are normalized to unity. Signal distributions are split according to the electron classification. From top to bottom: $\Delta\eta$ (a), $\Delta\phi$ (b), $\sigma_{\eta\eta}$ (c), E_{seed}/p_{in} (d), H/E (e). The figures refer to barrel electrons.

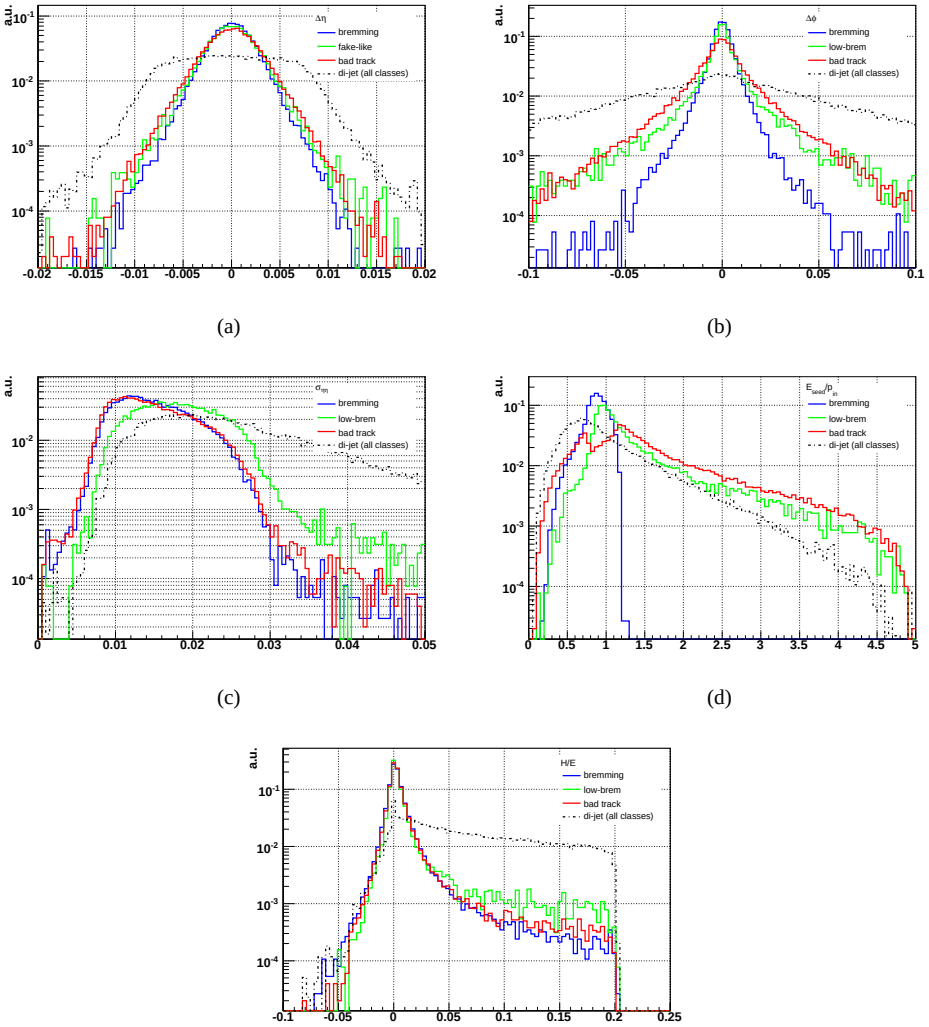


Figure 6: Discriminating variables used in the tight and loose Category based electron identification for signal and background. The distributions are normalized to unity. Signal distributions are split according to the electron classification. From top to bottom: $\Delta\eta$ (a), $\Delta\phi$ (b), $\sigma_{\eta\eta}$ (c), E_{seed}/p_{in} (d), H/E (e). The figures refer to endcap electrons.

		bremming	low brem	bad track	E/p$_{in}$ > 1.5
H/E	(barrel)	0.022	0.02	0.015	
	(endcap)	0.04	0.02	0.02	
$\sigma_{\eta\eta}$	(barrel)	0.01	0.0095	0.009	
	(endcap)	0.0275	0.025	0.024	
$\Delta\eta_{in}$	(barrel)	0.00525	0.0025	0.004	
	(endcap)	0.005	0.005	0.0055	
$\Delta\phi_{in}$	(barrel)	0.015	0.01	0.015	0.09
	(endcap)	0.025	0.016	0.025	0.092
E$_{seed}$/p$_{in}$	(barrel)	0.3	0.93	0.60	
	(endcap)	0.3	0.85	0.65	

Table 6: Discriminating variables used in the tight identification. Two different set of cuts are used in barrel and endcap regions, cuts are also applied according to electron class.

The thresholds for the tight level of selection of the electron identification variables have been set investigating signal and background efficiency plots versus variable under study. The cut values has been selected in order to ensure relative high efficiency and strong fake rejection. Fig. 7 shows the efficiency curves for the signal and the di-jet background for electrons in the barrel.

Tight selection reduce further the fakes (about a factor 10) with the respect to the Loose level with nearly 85% efficiency on the signal.

5 Performance of the methods

The performance of electron identification was estimated using simulated data by measuring the efficiency and purity for the selection of prompt electrons. The performance is reported in two ways. First, as the average fraction of correctly identified electrons in a proton-proton scattering where prompt electrons are produced in the hard process, such as electroweak boson production and subsequent decay to electrons. Second, as the average rate of identified electrons in di-jet QCD events.

The electron identification was applied on the electron sample reconstructed with the `PixelMatchGsfElectron` algorithm, as of `CMSSW_1_6_7`. The reconstructed electrons were required to have $p_T > 5\mathrm{GeV}$ and $|\eta| < 2.5$. The `PixelMatchGsfElectron` algorithm is based on the matching between the electromagnetic energy collected in an ECAL SuperCluster and the track of a charged particle. In the `CMSSW_1_6_7` release of the algorithm it is possible that one prompt electron will be associated to two reconstructed electrons, either sharing the same SuperCluster or sharing the same track, therefore raising the total number of reconstructed electrons. The prompt electron was considered to be correctly reconstructed if the distance $\Delta R = \sqrt{(\Delta\eta)^2 + (\Delta\phi)^2}$ between its momentum vector and the momentum vector of a reconstructed electron with the same charge was less than 0.05. An electron associated through this criteria to an electron of a different set is considered to be "matched". If two or more reconstructed electrons where within ΔR<0.05 of the prompt electron, the one with smallest E/p_{in} was considered to be matched. If two or more electrons, either prompt or non-prompt, were associated to the same reconstructed electron, only the one with p_T closest to the reconstructed electron p_T was considered to be matched.

5.1 Identification efficiency of prompt electrons

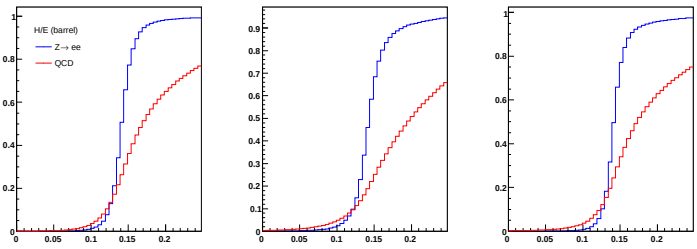
The selection efficiency for prompt electrons was evaluated using a sample of Z→ee and W→eν events. To test the different scenarii the samples was simulated and reconstructed with the CMS misalignment and miscalibration conditions expected at the start-up, after 10 pb^{-1}, 100 pb^{-1} of data taking and with the ideal conditions.

The efficiency is defined as the ratio between the number of electrons selected by the electron identification and the number of reconstructed electrons, given that the reconstructed electrons matched prompt electrons.

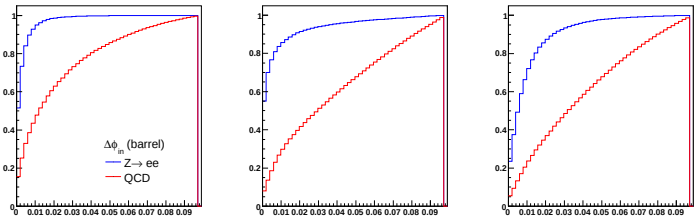
In Figure 8 it is shown the efficiency of electron identification in Z→ee events for either Fixed Thresholds and Category Based selections as a function of the prompt electron p_T, of the SuperCluster η, ϕ and z position of the primary vertex. The used samples were simulated according to the 10 pb^{-1} conditions.

The Loose Fixed Threshold selection provides an efficiency of 98% for electrons with $p_T > 10\mathrm{GeV}$, reaching the plateau of 99% efficiency for $p_T \sim 25\mathrm{GeV}$.

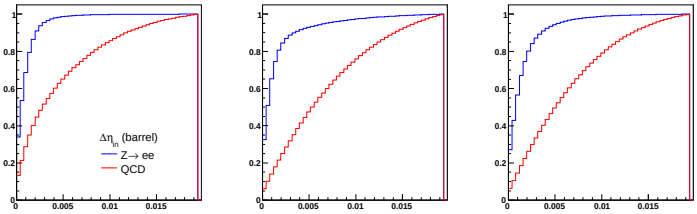
The tight selection either for Fixed Threshold and Category Based has considerably lower efficiency than robust



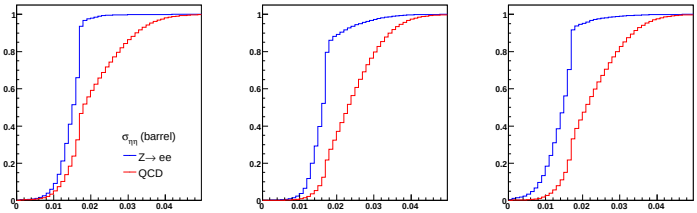
(a)



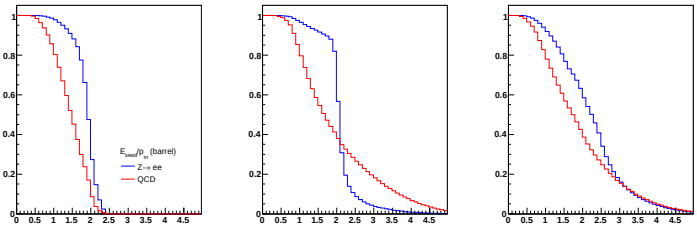
(b)



(c)



(d)



(e)

Figure 7: Efficiency distributions for the variables used in the selection for electrons in the barrel. Electrons are divided according the three categories from left to right: bremming, low-brem, bad track. Blue line for the signal (Z→ee), red line for the background (di-jets).a) H/E b) $\Delta\phi$ c) $\Delta\eta$, d) E_{seed}/p_{in} e) $\sigma_{\eta\eta}$.

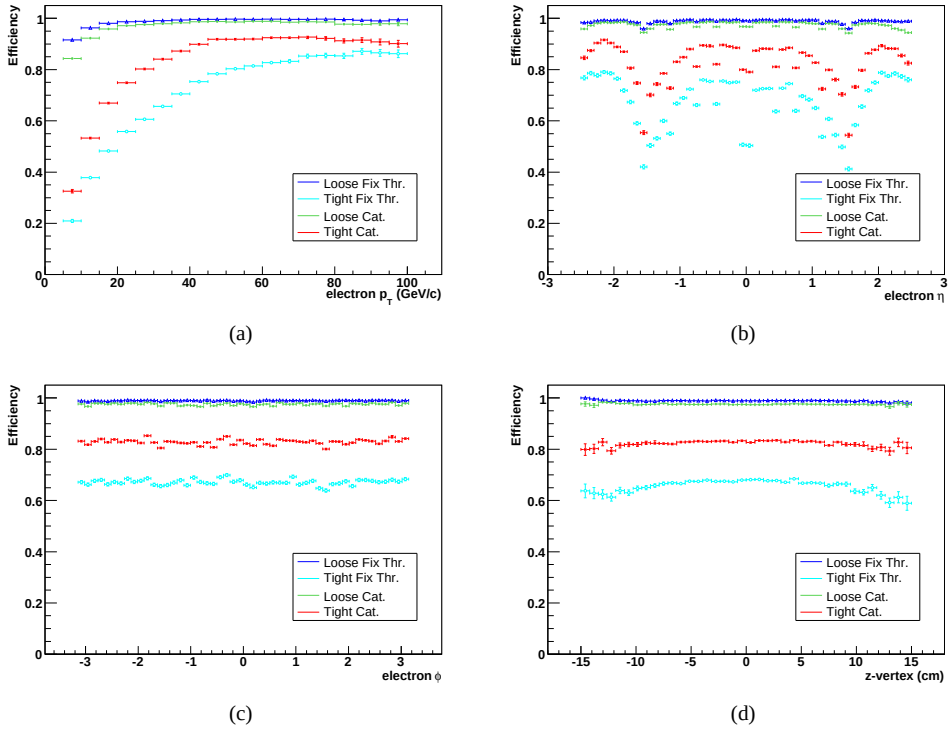


Figure 8: Selection efficiency for different identification selection levels from electrons in Z→ee events as a function a) p_T, b) η, c) ϕ d) z vertex. Samples were simulated with according to 10 pb^{-1} scenario.

for $p_T < 25\mathrm{GeV}$, reaching the highest efficiency for $p_T \sim 40\mathrm{GeV}$. The efficiency is approximately uniform in $\eta$for the loose identifications, whereas the tight selection efficiency seems affected by the ECAL intermodule cracks and the barrel-endcap gap. The strong η dependence for the tight selections seems to be correlated to the use of H/E variables but a more detailed study is needed. Furthermore the distributions show no dependence in ϕ and z position of vertex is as expected.

Figure 9 shows the same plots for W→eν events. The used samples were simulated according to the 10 pb^{-1} conditions. The results are almost the same, the small differences in the performance of the identification are mainly due to the different p_T spectrum of electrons in the two channels.

The effect of misalignment and miscalibration on electron identification was studied by comparing the efficiency evaluated from samples simulated and reconstructed in four scenarii: start-up, 10 pb^{-1} of data, 100 pb^{-1} of data and in an ideal alignment and calibration situation. In Figure 10 it is shown the efficiency of Loose Fixed Threshold selection for electrons in Z→ee events. The efficiency of this selection does not vary with the scenario, since the discrepancies are within statistical uncertainties. In Figure 11 it is shown the efficiency of Loose Fixed Threshold selection for electrons in W→eν_e events. Again the efficiency of the selection is insensitive to the scenario.

Table 7 reports the measured identification efficiencies for different scenarii and identification levels.

Table 7: Identification efficiency for offline candidates in Z→ee and W→eν. The Table reports the results for different scenarii and identification levels.

Dataset	Identification efficiency			
	Fixed Threshold		Category Based	
	Loose	Tight	Loose	Tight
Z→ee (Ideal)	99.0%	-	-	-
Z→ee (0 pb^{-1})	98.7%	-	-	-
Z→ee (100 pb^{-1})	99.0%	-	-	-
Z→ee (10 pb^{-1})	98.8%	67.1%	97.5%	82.8%
W→eν (10 pb^{-1})	98.7%	63.7%	97.1%	80.1%

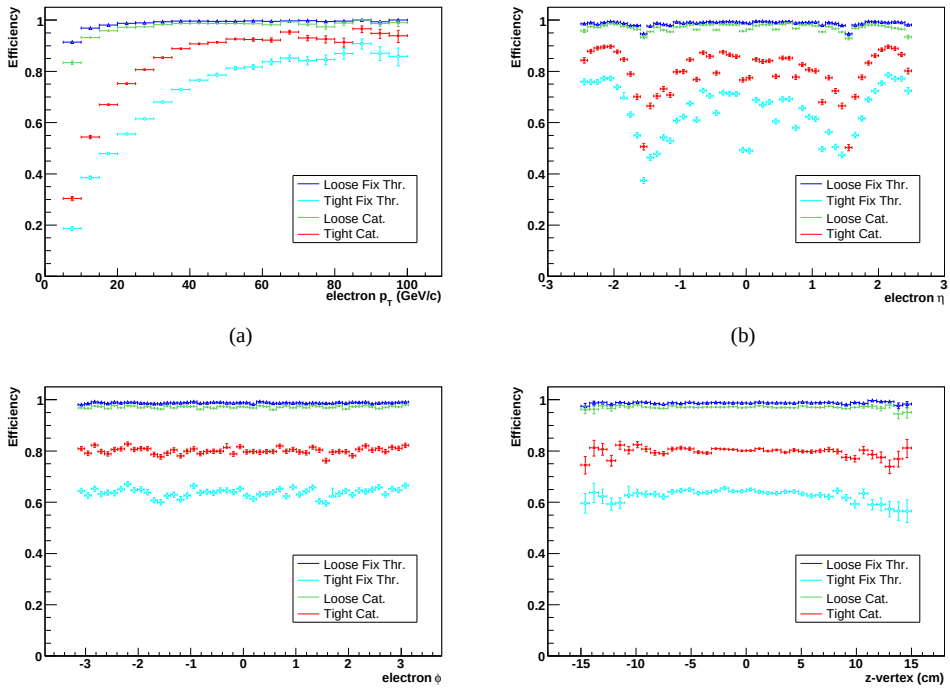


Figure 9: Selection efficiency for different identification selection levels from electrons in W→eν events as a function a) p_T, b) η, c) ϕ d) z vertex. Samples were simulated with according to 10 pb^{-1} scenario.

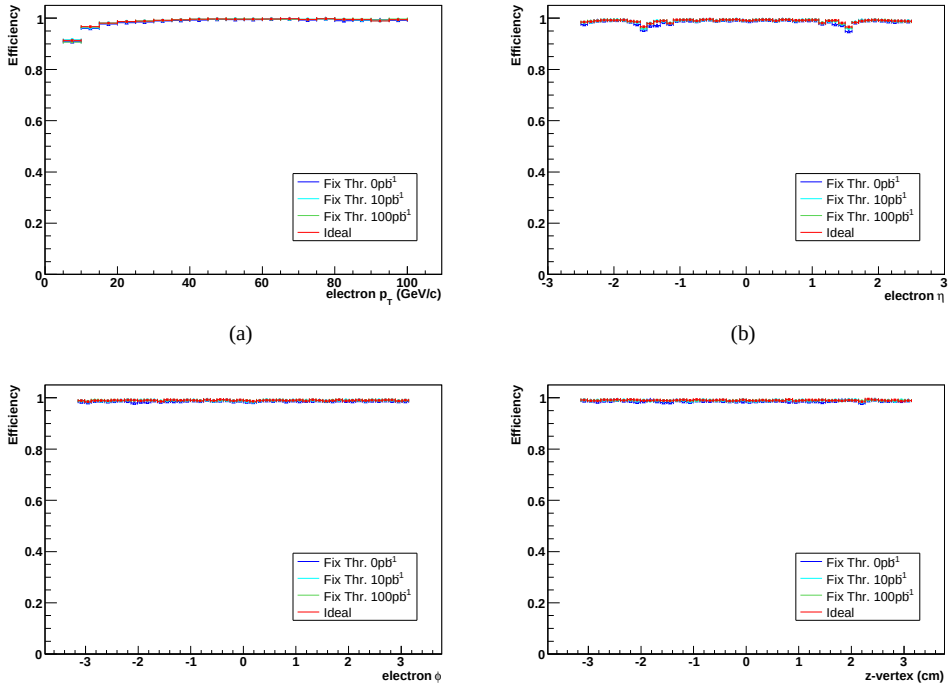
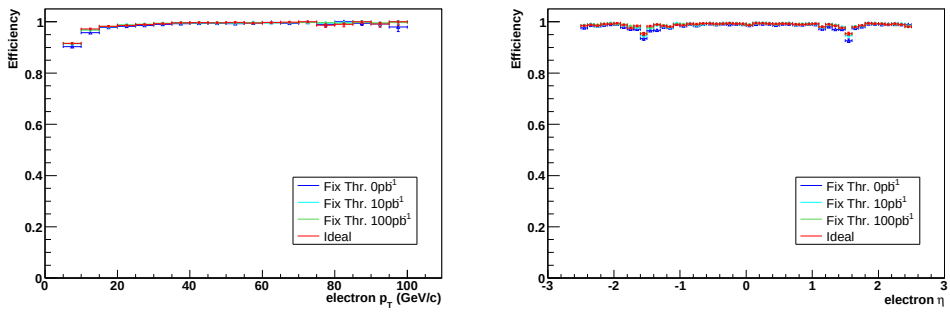
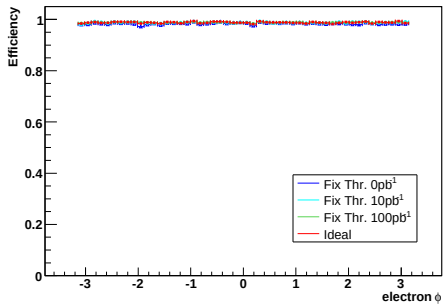


Figure 10: Selection efficiency for Loose Fixed Threshold selection for different scenarii in Z→ee events as a function a) p_T, b) η, c) ϕ d) z vertex.

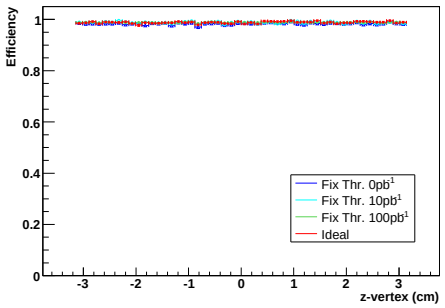


(a)

(b)



(c)



(d)

Figure 11: Selection efficiency for Loose Fixed Threshold for different scenarii in W→eν events as a function a) p_T, b) η, c) ϕ d) z vertex.

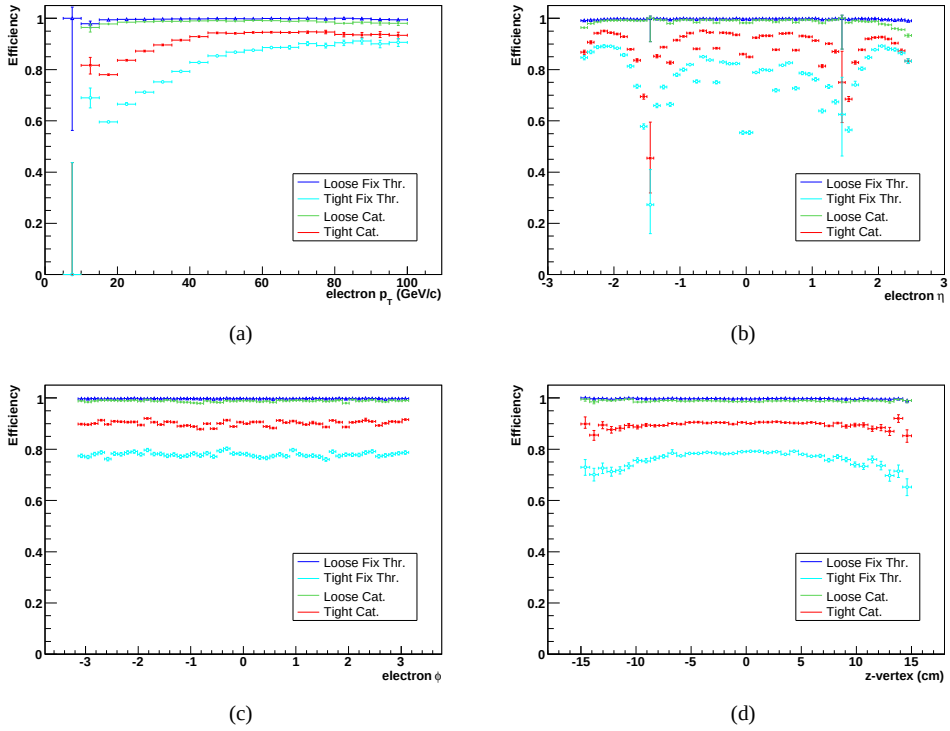


Figure 12: Selection efficiency for different identification selection levels from HLT electron candidates in Z→ee events as a function a) p_T, b) η, c) ϕ d) z vertex.

5.2 Identification efficiency for HLT candidates

The performance of electron identification have been tested also with HLT candidates. In addition to the requirements described previously the electron is asked to be an HLT candidate. The involved trigger paths are: Single Isolated and Single Relaxed electron, Double Isolated and Double Relaxed electrons. They imply an E_T cut on the SuperCluster transverse energy and, for the isolated paths a tracker isolation cut: a threshold is applied on the p_T of tracks within a cone around the electron direction, but outside a smaller veto cone in order to exclude the electron track itself. For a detailed description of the these trigger paths see[2].

Figures 12 and 13 show the identification efficiency distributions in Z→ee W→eν events as a function of p_T, η, ϕ and z of the primary vertex for HLT candidates. The distributions look unaltered and the overall efficiency is enhanced. Table 8 reports the efficiencies for the two signals simulated with the 10 pb^{-1} scenario.

Table 8: Identification efficiency for HLT electron candidates.

	Identification efficiency for HLT candidates			
Dataset	Fixed Thresholds		Category Based	
	Loose	Tight	Loose	Tight
Z→ ee (10 pb^{-1})	99.7%	77.9%	98.8%	90.1%
W→eν (10 pb^{-1})	99.6%	76.3%	98.6%	89.1%

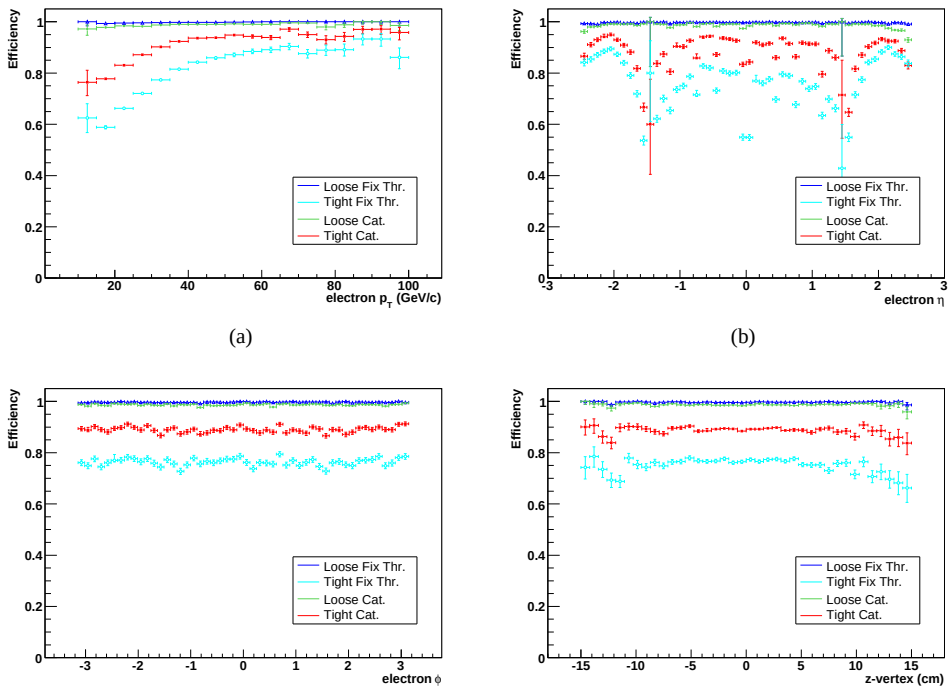


Figure 13: Selection efficiency for different identification selection levels from HLT electron candidates in W→eν events as a function a) p_T, b) η, c) ϕ d) z vertex.

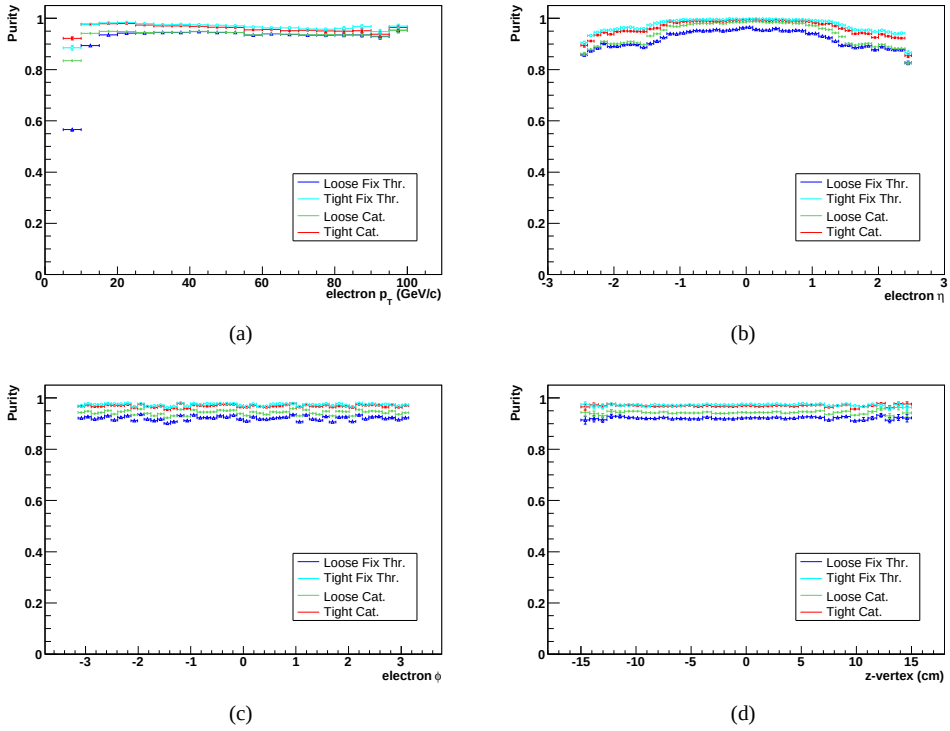


Figure 14: Selection purity for different identification selection levels from electrons in Z→ee events as a function a) p_T, b) η, c) ϕ d) z vertex.

5.3 Identification purity of prompt electrons

The identification purity was estimated from a sample of Z→ee and W→eν events simulated and reconstructed with misalignment and miscalibration conditions expected after 10 pb^{-1} of data taking. The purity was defined as the ratio between the number of electrons selected by the electron identification which were matched to prompt electrons and the total number of selected electrons. The results are shown in Figures 14 and 15. The purity for tight electron identification is above $95\%$for reconstructed electrons with $p_T > 10\mathrm{GeV}$. For Loose Fixed Threshold selection the purity is lower, about 90% for reconstructed electrons with $10 < p_T < 15\mathrm{GeV}$. The average purity is higher in the barrel than in the endcap. A significant fraction of electron contamination might be due to ambiguities in electron reconstruction resulting in electron double counting, as explained in the beginning of this section. Table 9 reports the purity for the two studied samples and the different identification levels.

Table 9: Purity for off-line electrons.

	Purity			
Dataset	Fixed Thresholds		Category Based	
	Loose	Tight	Loose	Tight
Z→ee (10 pb^{-1})	92.1%	97.4%	94.3%	96.0%
W→eν (10 pb^{-1})	88.5%	96.5%	93.0%	95.1%

5.4 Electron fake-rate

The electron fake-rate was estimated with the Gumbo Soup sample. This is essentially a weighted mixture of di-jet and photon plus jets events simulated and reconstructed with two misalignment and miscalibration conditions: 10 pb^{-1} and 100 pb^{-1} of data taking.

The fake-rate was defined in two different ways: per reconstructed jet and per SuperCluster. The fake-rate per jet is defined as the ratio between the number of reconstructed jets that match a reconstructed and identified electron divided by the total number of reconstructed jets. Considered jets are reconstructed with the `IterativeCone` algorithm within a cone of 0.5 and have $E_T > 5$ GeV. The jet-electron matching is performed as a geometrical

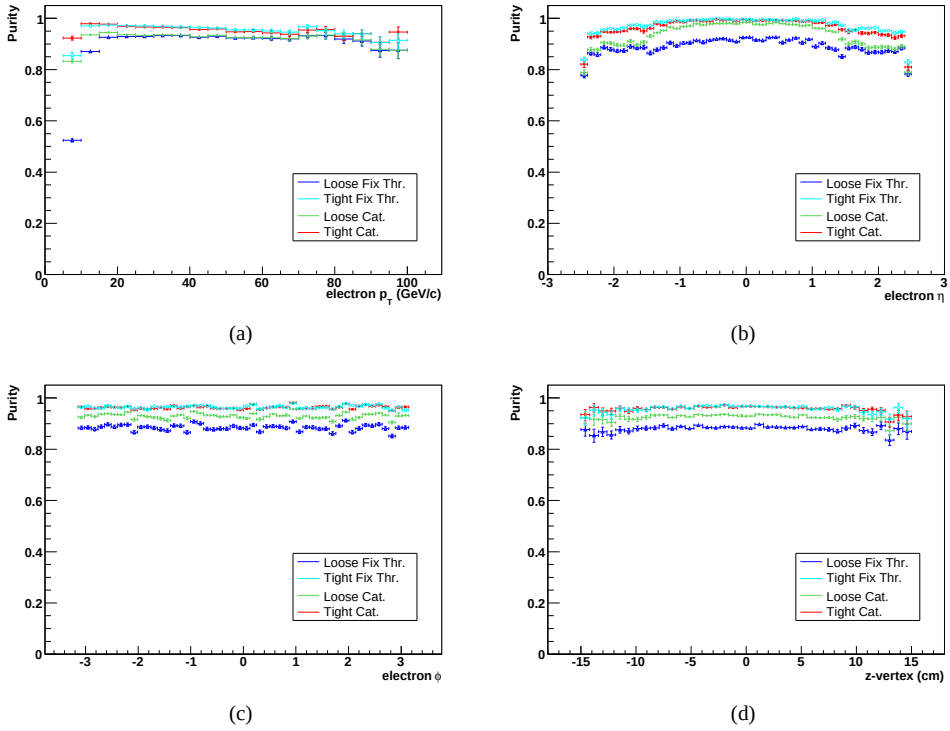


Figure 15: Selection purity for different identification selection levels from electrons in W→eν events as a function a) p_T, b) η, c) ϕ d) z vertex.

matching within a cone of 0.2.

The fake-rate per SuperCluster instead is defined as the ratio between the number of reconstructed SuperClusters that match a reconstructed and identified electron divided by the total number of reconstructed SuperClusters. An $E_T > 5$ GeV cut is applied to the SuperClusters. As in the previous definition the matching between SuperClusters and electrons is geometrical in a cone of 0.2.

The fake-rate results are shown in Figure 16 (per SuperCluster) and 18 (per reconstructed jet) as a function of E_T, η, ϕ and z position of the primary vertex. Both pictures refers to 10 pb^{-1} scenario. The fake-rate shows the same strong η dependence while is completely flat in ϕ as expected. Tight selections provide a fake-rate reduction by more than a factor 10 with respect to the Loose selections.

We have performed the comparison of different calibration scenario to check the variation on the fake-rate. Unfortunately for the background only two scenarii were available. Figures 17 and 19 show the comparison between 10 pb^{-1} and 100 pb^{-1} scenarii for the Loose Fixed Threshold identification and no appreciable difference can be noticed.

6 Conclusions

A new Cut Based Electron Identification for the LHC start-up period has been proposed. Two different methods of electron identification have been studied each one with two level of increasing purity.

A Fixed Threshold selection uses variables that are stable against a misaligned tracker and mis-calibrated ECAL, therefore are expected to be well measured during early data taking. An efficiency of 98% is reached with a reasonable low level of fake-rate (around 1% referring to SuperClusters with $E_T > 5$ GeV) for the Loose selection.

To reduce further the fake-rate, keeping almost the same efficiency, the electrons are categorized according to the estimate of energy fraction emitted by bremsstrahlung in the tracker material and the ratio of the energy collected in the electromagnetic calorimeter to momentum of the electron track. Simple sequential cuts are set according to the estimated ratio of signal to background events in one of each categories. With the Loose thresholds the fake-rate can be reduced by a factor 5 with the respect to the corresponding Loose Fixed Threshold selection.

Besides the Loose level of identification a set of Tight thresholds have been studied. This selections should be

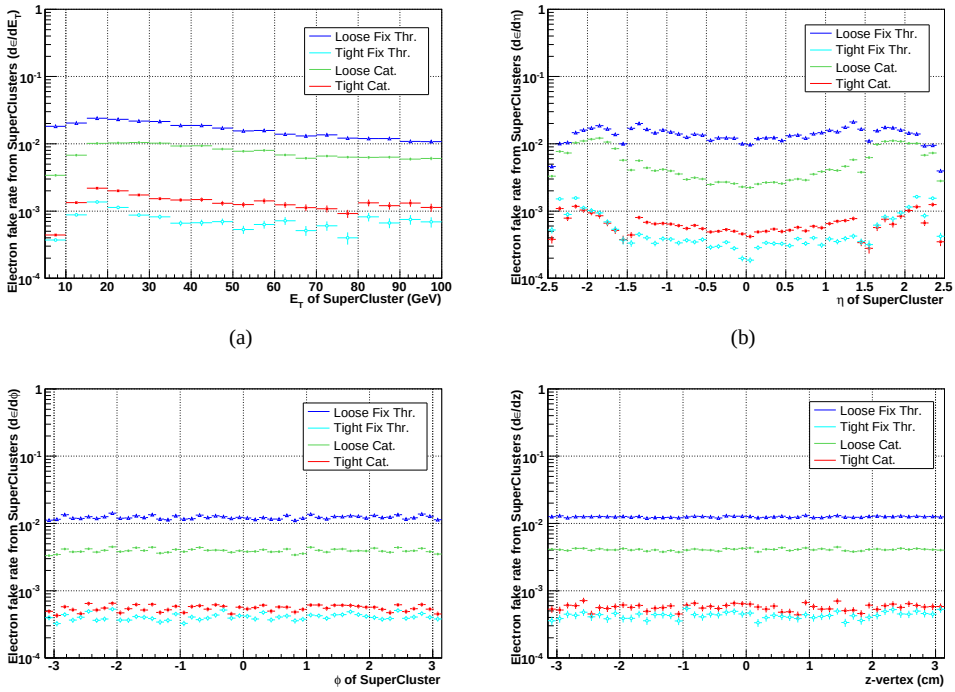


Figure 16: Fake-rate per SuperCluster in Gumbo soup events for the different identification levels as a function a) p_T, b) η, c) ϕ d) z vertex.

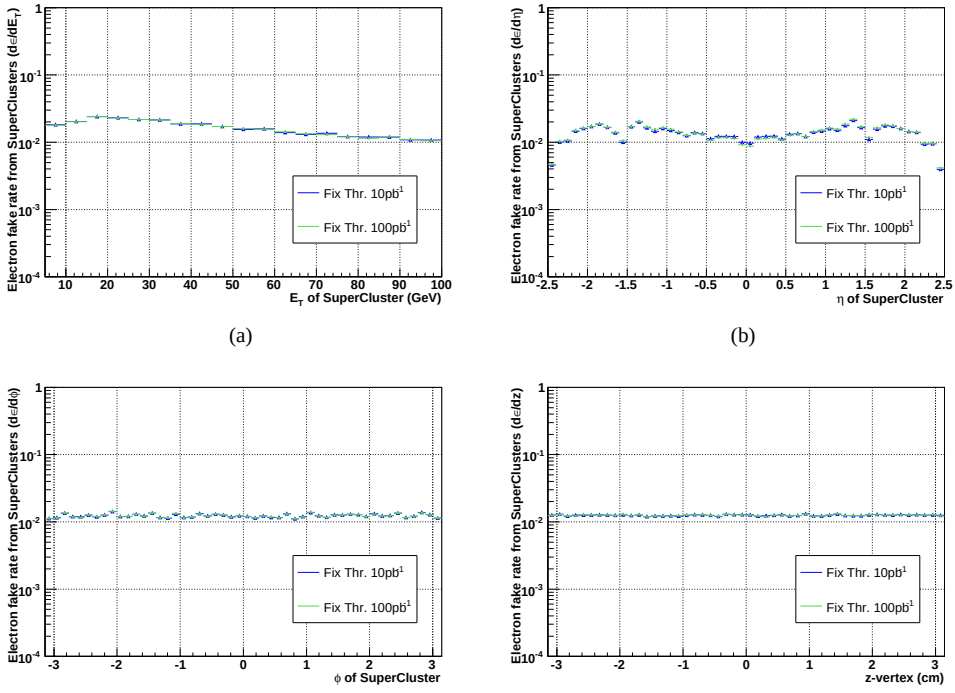


Figure 17: Comparison of the fake-rate per SuperCluster in Gumbo soup events for two different miscalibration/misalignment scenario as a function a) p_T, b) η, c) ϕ d) z vertex.

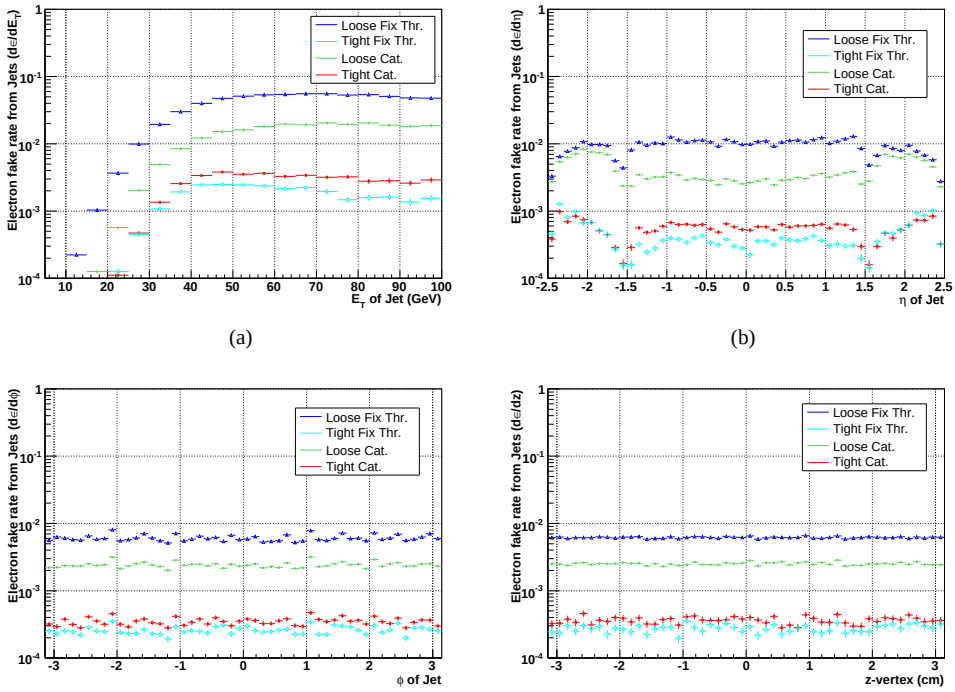


Figure 18: Fake-rate per Jet in Gumbo soup events for the different identification levels as a function a) p_T, b) η, c) ϕ d) z vertex.

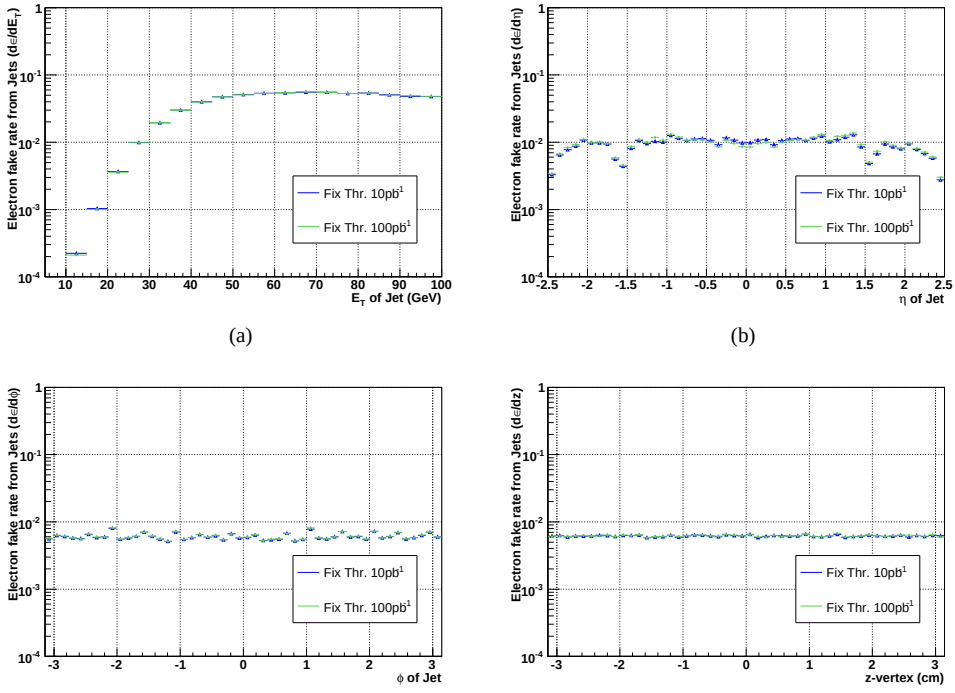


Figure 19: Comparison of the fake-rate per SuperCluster in Gumbo soup events for two different miscalibration/misalignment scenario as a function a) p_T, b) η, c) ϕ d) z vertex.

suitable for the W analysis where a higher purity of the electron sample is required. With tight selection the fake-rate can be further reduced by nearly a factor 10 with a corresponding 15% loose of efficiency.

This study has been particularly oriented to the start-up condition so we have checked how the performance of the electron identification should change with the calibration and alignment knowledge of the detector. In particular we have studied samples simulated with four different scenarii: start-up, 10 pb^{-1}, 100 pb^{-1} and ideal conditions. We have verified that either efficiency and fake-rate distributions do not change appreciably with the scenario and that the selection methods are stable enough for the initial data-taking conditions.

Computing, Software and Analysis Challenge (CAS07) CMS production

References

[1] Baffioni S. *et al.*, "Electron Reconstruction in CMS", CERN-CMS-NOTE-2006/040.

[2] Acosta D. *et al.*, "CMS High Level Trigger", CERN-CMS-AN-2007/009.

[3] Adam N. *et al.*, "Towards a Measurement of the Inclusive W→eν and Z→ee Cross Section in pp Collision at $\sqrt{s} = 14$ TeV", CERN-CMS-AN-2007/026.