Available on CMS information server **CMS AN -2009/178**

The Compact Muon Solenoid Experiment

Analysis Note

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

02 December 2009

Electron Identification in CMS

S. Baffioni[a], F. Beaudette[b], D. Benedetti[b], J. Branson[c], G. Daskalakis[d], E. Di Marco[e], C. Campagnari[f], C. Charlot[a], S. Harper[g], P.D. Kalavase[f], J.D. Keller[h], D. Lelas[i], P. Meridiani[b], M. Pieri[c], I. Puljak[i], R. Ranieri[c], N. Rompotis[l], C. Rovelli[e], R. Salerno[a,m], M. Sani[c], C. Seez[l], Y. Sirois[a], Y. Tu[c], A. Yagil[c]

a) LLR, Laboratoire Leprince-Ringuet, Ecole Polytechnique, IN2P3-CNRS, Palaiseau

b) CERN, European Organisation for Nuclear Research, Geneva

c) University of California, San Diego

d) Institute of Nuclear Physics "Demokritos", Attiki

e) Università di Roma I, INFN, Roma

f) University of California, Santa Barbara

g) RAL, Rutherford Appleton Laboratory, Didcot

h) University of Nebraska, Lincoln

i) FESB, Technical University of Split, Split

l) Imperial College, University of London, London

m) Università di Milano-Bicocca, INFN, Milano

Abstract

The electron identification algorithms and expected performances from a detailed Monte Carlo simulation are presented. The description of the selections and different optimization techniques are discussed. The method to check the identification variable shapes from data is described. Dedicated electron identification selections for not isolated, low p_T and high energy electrons are shown.

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 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.

After that the starting point for a more efficient electron selection is to introduce a classification of the electrons. Electrons categories have been originally proposed for electron selection in [1] and the same categorization is still used for the best momentum determination in the electron reconstruction [2]. Subsequently a classification based only on the on the brem fraction measured from the ratio of the difference between the track momentum at the vertex (p_{in}) and at the ECAL (p_{out}) and E/p has been proposed in [3].

Finally possible improvements using multivariate techniques and likelihood selection could be used later when the detector will be better understood [4].

This note complements previous work and documents the performances of the electron identification based on full Monte Carlo simulation prior to first collision data. The presented performances have been obtained using CMSSW version 3. Samples of $Z \to e^+e^-$, $W \to e\nu$, QCD di-jet and γ-jet events are used to illustrate the performances on an isolated electron signal and background electrons. In Table 1 the details for the samples are collected.

The note is organized as follow: in section 2, the identification algorithms and variables are described. The performances of the selections on isolated electrons are studied in section 3. Then, the description of the selections and different optimization techniques, based on real data are discussed in section 4.

Possible improvements using multivariate techniques, variable and likelihood selection are presented in section 5. The method to check the identification variable shapes from data is described in section 6. Finally dedicated electron identification selections for not isolated, low p_T and high energy electrons are presented respectively in sections 7, 8 and 9.

Channel	Dataset
$Z \to e^+e^-$	/Zee/Summer09-MC_31X_V3-v1/GEN-SIM-RECO
$W \to e\nu$	/Wenu/Summer09-MC_31X_V3-v1/GEN-SIM-RECO
di-jet	/QCD_EMEnriched_Pt20to30/Summer09-MC_31X_V3-v1/GEN-SIM-RECO
di-jet	/QCD_EMEnriched_Pt30to80/Summer09-MC_31X_V3-v1/GEN-SIM-RECO
di-jet	/QCD_EMEnriched_Pt80to170/Summer09-MC_31X_V3-v1/GEN-SIM-RECO
di-jet	/QCD_BCtoE_Pt20to30/Summer09-MC_31X_V3-v1/GEN-SIM-RECO
di-jet	/QCD_BCtoE_Pt30to80/Summer09-MC_31X_V3-v1/GEN-SIM-RECO
di-jet	/QCD_BCtoE_Pt80to170/Summer09-MC_31X_V3-v1/GEN-SIM-RECO
photon-jet	/PhotonJet_Pt15to20/Summer09-MC_31X_V3-v1/GEN-SIM-RECO
photon-jet	/PhotonJet_Pt20to30/Summer09-MC_31X_V3-v1/GEN-SIM-RECO
photon-jet	/PhotonJet_Pt30to50/Summer09-MC_31X_V3-v1/GEN-SIM-RECO
photon-jet	/PhotonJet_Pt50to80/Summer09-MC_31X_V3-v1/GEN-SIM-RECO
photon-jet	/PhotonJet_Pt80to120/Summer09-MC_31X_V3-v1/GEN-SIM-RECO
photon-jet	/PhotonJet_Pt120to170/Summer09-MC_31X_V3-v1/GEN-SIM-RECO
photon-jet	/PhotonJet_Pt170to300/Summer09-MC_31X_V3-v1/GEN-SIM-RECO

Table 1: List of the samples used to evaluate electron identification performance.

2 Algorithms and variables

Electrons, for example from W and Z decay, can be distinguished from other particles due to their unique characteristics that are primarily measured in the ECAL and tracker. Ideally, the electron track would match well with the cluster of energy found the ECAL, both in position and momentum. The track would also emanate directly from the event vertex and be isolated from other tracks and calorimeter energy deposits.

Electron identification makes use of a complete set of variables to distinguish between real electrons and background electrons. Here we give only a brief description of the most used variables:

- Hadronic to electromagnetic energy ratio: H/E.
- Energy-momentum matching variables between the energy of the super cluster or of the super cluster seed and the electron track measured momentum at the vertex or at the calorimeter: $\mathrm{E}/\mathrm{p}_{\mathrm{in}}$, $\mathrm{E}_{\mathrm{seed}}/\mathrm{p}_{\mathrm{in}}$ and $\mathrm{E}_{\mathrm{seed}}/\mathrm{p}_{\mathrm{out}}$.
- Geometrical matching between the electron track parameters at the vertex extrapolated to the super cluster and the measured super cluster position: $\Delta\phi_{\mathrm{in}}$ and $\Delta\eta_{\mathrm{in}}$.
- Calorimeter shower shape variables: the width of the ECAL cluster along the η direction computed for all the crystals in the 5×5 block of crystals centered on the highest energy crystal of the seed cluster, $\sigma_{\mathrm{i}\eta\mathrm{i}\eta}$ [1)] and the ratio of the energy sums over the 3×3 and 5×5 matrices centred on the highest energy crystal of the seed cluster, $\sum_9 / \sum_{25}$

Considering the high photon conversion rates due to the material budget in front of ECAL some cuts are designed to reject electrons from conversion:

- Impact Parameter: d0 of the electron track computed with respect to the reconstructed vertex.
- Missing Hits: number of crossed layers without compatible hits in the back-propagation of the track to the beam-line.

The performance of the electron identification depends of course on the degree of isolation imposed on the electron candidates. The most common isolation variables and their suggested cuts *find a reference* are:

- Tracker Isolation: sum of $\mathrm{p_T}$ of tracks with $\mathrm{p_T} > 0.7$ GeV/c and maximum distance to the vertex of 0.2 cm in a cone of 0.3 with an inner veto cone of 0.04,
- ECAL Isolation: sum of energy of ECAL RecHits with a Jurassic footprint removal (Jurassic width 1.5 crystals) in a cone of 0.4 with veto cone of 3 crystals. A Rechit noise cut is also applied: 0.08 GeV (E) for barrel and 0.1 GeV ($\mathrm{E_T}$) in the endcap,
- HCAL Isolation: sum of HCAL CaloTowers in a 0.4 cone with a 0.15 veto cone.

The distribution of both the isolation variables and the variables to reject electrons from conversions are shown in Figure 1.

2.1 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 quantities: H/E, $\Delta\phi_{\mathrm{in}}$, $\Delta\eta_{\mathrm{in}}$ and $\sigma_{\mathrm{i}\eta\mathrm{i}\eta}$.

Fig. 2 and Fig. 3 show the distributions for the variable used in the selection separately for barrel and endcap. Signal and background distributions are normalized to unity to enhance the shape differences.

2.2 Category Based Identification

A more efficient electron selection is to introduce a classification of the electrons to group electrons with the similar characteristics together.

The first electron categorization was introduced for optimal determination of the electron momentum and then it was also proposed to be used as basis of the electron identification [5]. This electron classification, initially proposed in [1], has been recently revisited [2]. The description of four, mutually exclusive, electron classes are reported in Appendix B

[1)] Recently $\sigma^{\mathrm{SC}}_{\mathrm{i}\eta\mathrm{i}\eta}$ was introduced, it has the same definition of the default $\sigma_{\mathrm{i}\eta\mathrm{i}\eta}$ variable but the sum is computed for all the crystals belonging to the electron super cluster. Despite the higher rejection power this variable is probably more sensitive to ECAL noise. On going studies are checking the actual benefit.

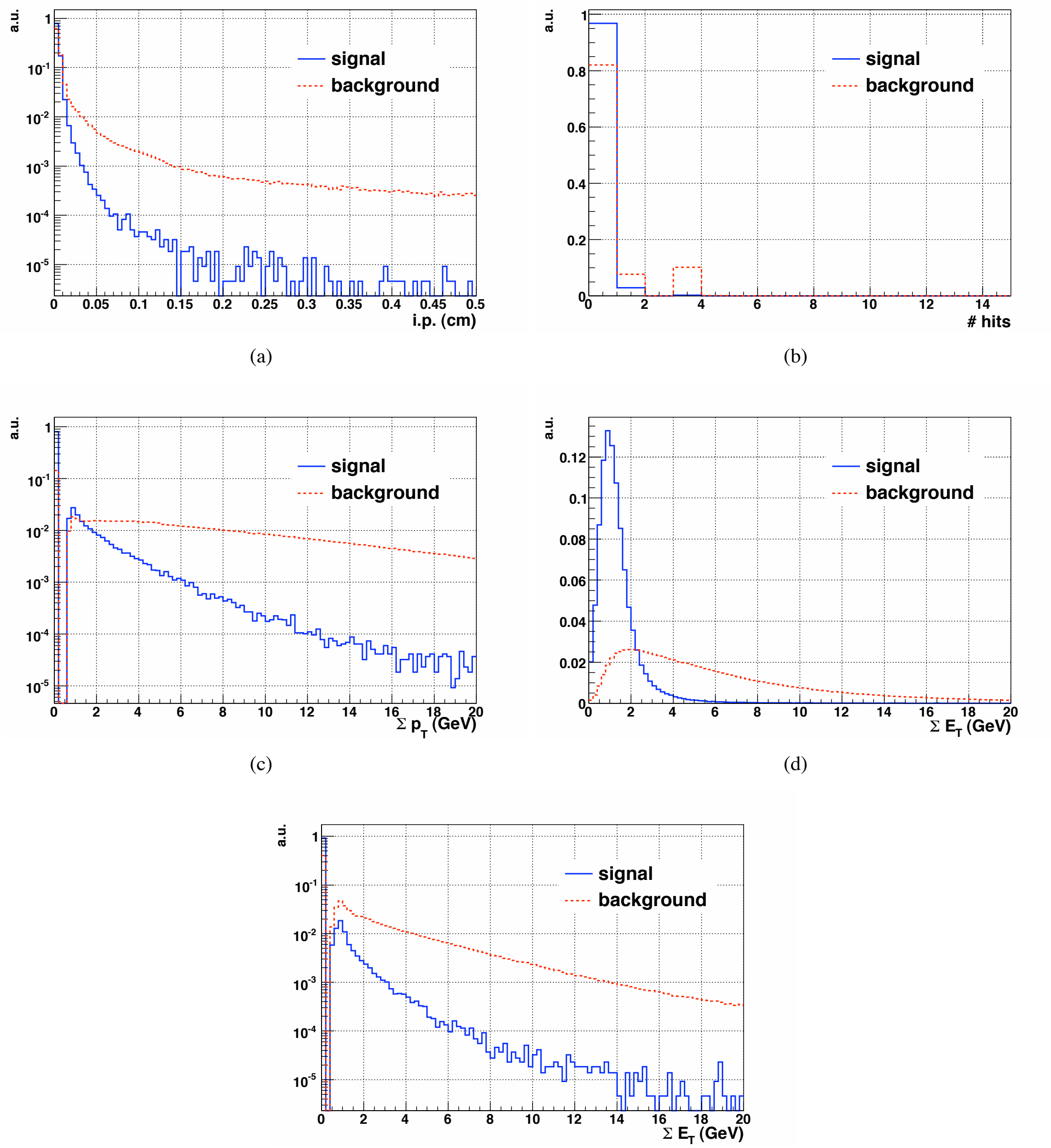
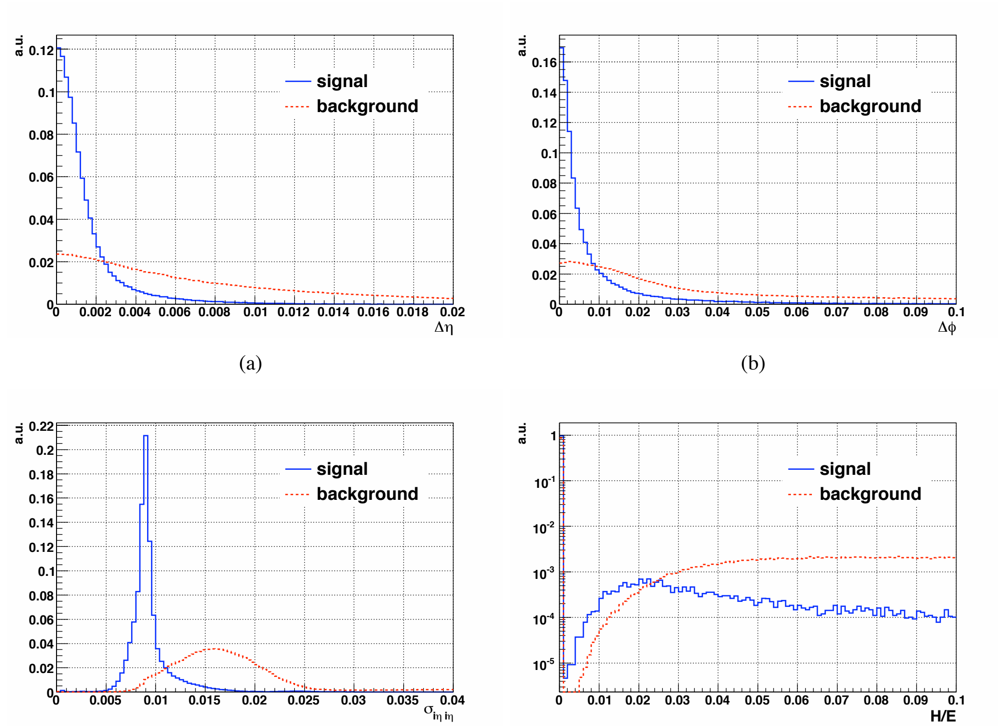
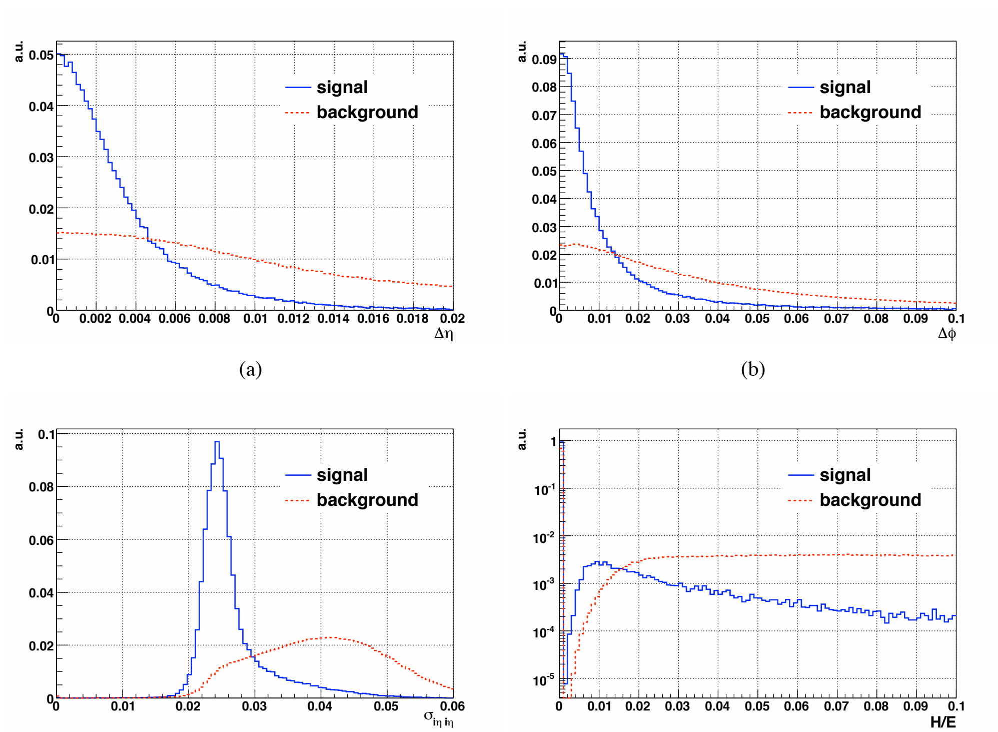


Figure 1: Distribution of the isolation variables and variables to reject electron from conversion in the barrel: (a) impact parameter, (b) number of missing hits, (c) Tracker Isolation, (d) ECAL Isolation, (e) HCAL Isolation. Signal and background distributions are normalized to unity. Similar distributions with larger spread are obtained for electrons in the endcaps.



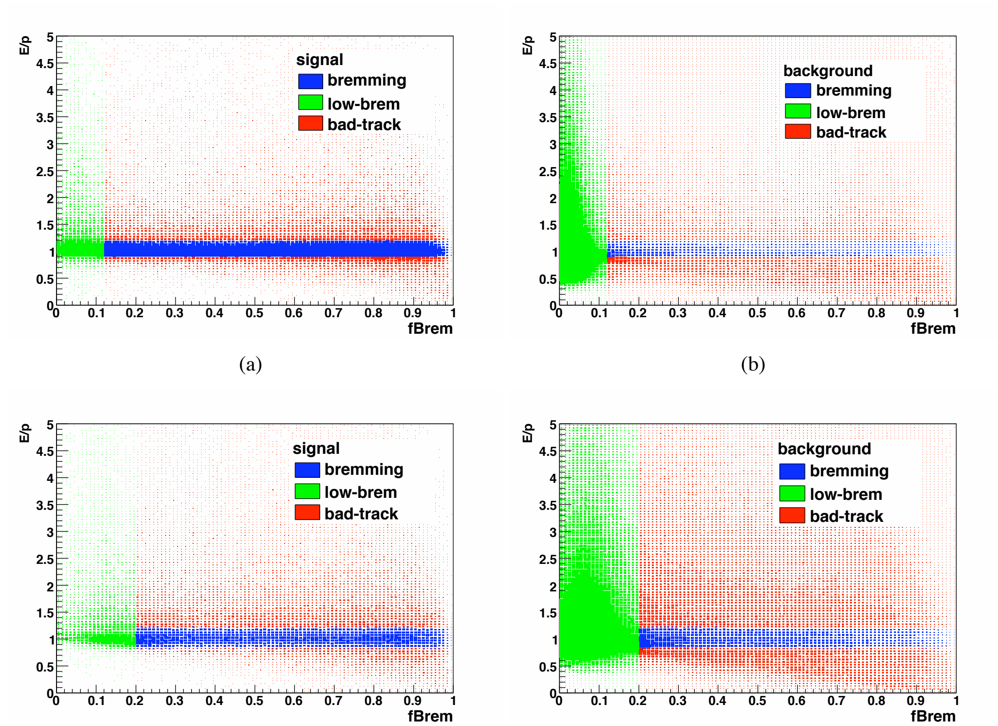
(a) (b) (c) (d)

Figure 2: Distribution of the variables used in Fixed Threshold selection in the barrel: (a) $\Delta\eta_{\rm in}$, (b) $\Delta\phi_{\rm in}$, (c) $\sigma_{\rm i\eta i\eta}$, (d) H/E. Signal and backgorund distributions are normalized to unity.



(a) (b) (c) (d)

Figure 3: Distribution of the variables used in Fixed Threshold selection in the endcap: (a) $\Delta\eta_{\rm in}$, (b) $\Delta\phi_{\rm in}$, (c) $\sigma_{\rm i\eta i\eta}$, (d) H/E. Signal and backgorund distributions are normalized to unity.



(a) (b) (c) (d)

Figure 4: Class definition used in the Category Based Electron Identification. The class population is shown for signal and background respectively in barrel (a)(b) and endcap (c)(d).

A classification driven by the considerations of separating out different regions of S/B and different background sources which have different signatures have been introduced in [3].
Some electron tracks are measured to loose significant energy in the tracker material and thus are very unlikely to be from other particles that do not radiate like electrons. These electrons are particularly well identified if the track momentum and the ECAL energy match indicating that both are well measured. We can maintain high efficiency on these electrons by not applying cuts that are too tight. On the other hand some electrons do not radiate much energy in the inner parts of the tracker and are thus not well separated from normal charged particles that are plentiful. Since the thickness of the tracker material varies, these non-radiating electrons an important fraction of the best measured electrons. To reduce the background due to charged particles and overlaps, some tight cuts need to be placed on these electron candidates. One particularly important cut removes low $\mathrm{E/p_{in}}$ candidates that likely come from charge pions or kaons. A large fraction of electrons also have a track that is mis-measured, primarily due to large energy loss in the tracker before three points on the track are measured. This category of electron might be faked by an overlap between a lower momentum charged particle and a high $\mathrm{E_T}$ π^0. To help reduce fakes yet keep reasonable efficiency, we need to place tighter track-ECAL matching cuts on this type of candidate.

We distinguish these three categories of electrons:

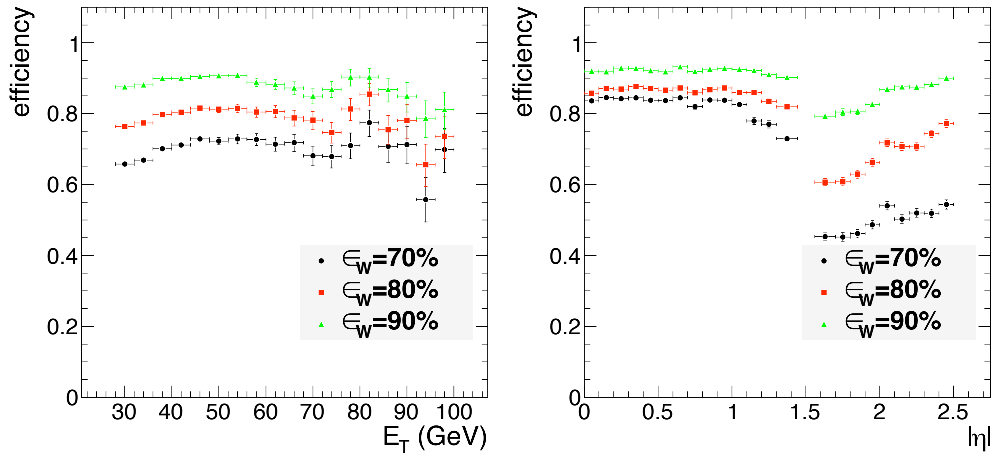
1. *low-brem electrons*: ($0.9 < \mathrm{E/p_{in}} < 1.2$ - fbrem < 0.12 (barrel), $0.82 < \mathrm{E/p_{in}} < 1.22$ - fbrem < 0.2 (endcap)), fake-like region with high population from both real and fake electrons,

2. *bremming electrons*: ($0.9 < \mathrm{E/p_{in}} < 1.2$ - fbrem > 0.12 (barrel), $0.82 < \mathrm{E/p_{in}} < 1.22$ - fbrem > 0.2 (endcap)), electrons-like region with little contamination from fakes,

3. *bad track*: (left regions), region with not many real electrons.

Fig. 4 represents graphically the classes described above.

We have tried to keep the number of categories small and to have all categories with a large number of good electrons so that the choice of cuts can be robust and can be made with low statistics in the early data, however, the large differences between the detectors in the barrel and endcap require us to set separate cuts for these two regions so there are six categories of electron candidates, three in the barrel and three in the endcap.



(a) (b)

Figure 5: Selection efficiencies as a function of $\mathrm{E_T}$ and η for Fixed Threshold selection $\mathrm{E_T} > 30$ GeV.

Cuts are applied on the following variables: H/E, $\Delta\eta_{\mathrm{in}}$, $\Delta\phi_{\mathrm{in}}$, $\sigma^{\mathrm{SC}}_{\mathrm{i}\eta\mathrm{i}\eta}$, $\mathrm{E_{seed}/p_{in}}$, d0 and on the missing hits number

3 Performance of the methods

In this section the performance of the different selection described above are shown. The performance of electron identification is estimated using simulated data by measuring the efficiency for the selection of prompt electrons and estimating the fake rate for background events.

The efficiency on signal samples 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 to prompt electrons.

The fake rate is reported in two ways: as efficiency of selecting a reconstructed electron in the background sample, or as the fraction of reconstructed SuperCluster belonging to a selected electron.

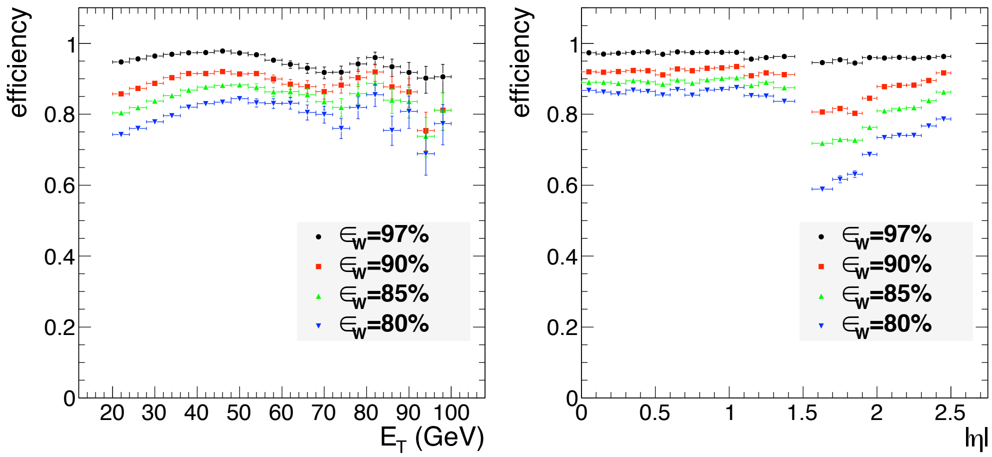
3.1 Fixed Threshold Identification

3.1.1 Signal Efficiency

In the following the efficiency plots obtained with the Fixed Threshold Electron Id in $\mathrm{W} \to \mathrm{e}\nu$ events are shown. To compute the efficiency signal events are selected requiring electrons with $\mathrm{E_T} > 30$ GeV and matched to an HLT candidate (trigger "1ele15 egamma"). The Iterative Method, described in the next Section, which has been used to optimized this selection can provide several set of cuts corresponding to a MonteCarlo efficiency for signal events ranging from 90% to 65%, cut values are reported in Appendix A. As an example Fig. 5 shows the overall efficiency of the selection for different working points as a function of electron $\mathrm{E_T}$ and η. Since these selections are primarly meant for $\mathrm{W} \to \mathrm{e}\nu$ analysis similar sets of cuts have been optimized for $\mathrm{E_T} > 20$ GeV which is more suitable for $\mathrm{Z} \to \mathrm{e^+e^-}$ analysis. Fig. 6 shows the overall efficiency for this alternative selection for different working points as a function of electron $\mathrm{E_T}$ and η.

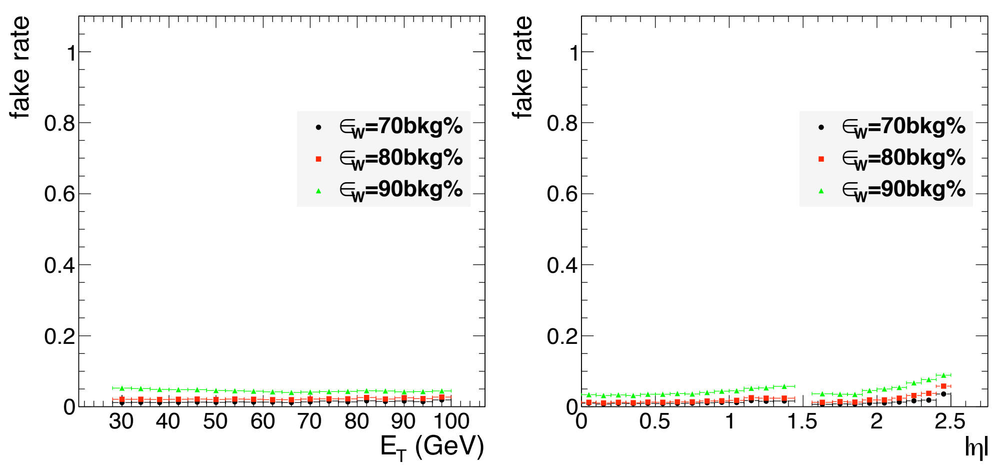
3.1.2 Background Rejection

The background sample used in the analysis is made of the samples described in Sec. 1. Fig. 7 shows the background efficiency as a function of $\mathrm{E_T}$and η under the same conditions as the one described for the signal in the previous Section for $\mathrm{E_T} > 30$ GeV for several working points. Fig. 8 shows instead the background efficiency for the selection optimized with $\mathrm{E_T} > 20$ GeV for several working points.



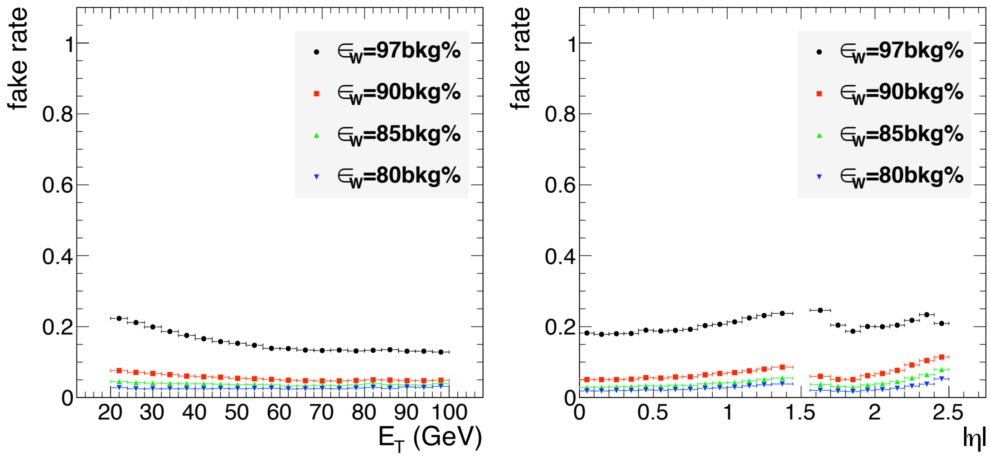
(a) (b)

Figure 6: Selection efficiencies as a function of $\mathrm{E_T}$ and η for Fixed Threshold selection $\mathrm{E_T} > 20$ GeV.



(a) (b)

Figure 7: Selection efficiency on the background as a function of $\mathrm{E_T}$ and η for Fixed Threshold selection $\mathrm{E_T} >$ 30 GeV.



(a) (b)

Figure 8: Selection efficiency on the background as a function of $\mathrm{E_T}$ and η for Fixed Threshold selection $\mathrm{E_T} >$ 20 GeV.

3.2 Category Based Identification

3.2.1 Signal Efficiency

The selection efficiency for Category Based Electron ID was evaluated using sample of either $\mathrm{Z} \rightarrow \mathrm{e^+e^-}$ and $\mathrm{W} \rightarrow \mathrm{e}\nu$ events. In Fig. 9 it is shown the efficiency of the electron identification for different tightness. The different set of cuts presented here are obtained with the "Differential S/B method" that will be described in the next Section in detail and the cut values for "Loose" and "Tight" selections are reported in Appendix A.

For electrons with $\mathrm{E_T}$ of 30 GeV the efficiency in the barrel is greater than 97% for the "VeryLoose" while it is greater than 75% for the "Hyper Tight1". In the two mentioned selections the efficiency reaches a plateau at 99% and 91% for electrons with $\mathrm{E_T} > 40$ GeV.

The adopted optimization procedure tends to maximize the overall efficiency for the signal sample. In the endcap, where the background is higher, the procedure sets the cuts tighter than in the barrel and the resulting signal efficiency is lower.

No relavant difference in the performance have been observed for electrons from W decays.

3.2.2 Background Rejection

In Fig. 10 the background rejection (fake rate) of the Category Based Electron Id is shown as the fraction of reconstructed SuperCluster belonging to a selected electron.

It is important to notice that the QCD di-jet sample used in these studies has been pre-selected at generator level to enhance its electromangetic component. Such a filter contained a requirement for at least one electromagnetic cluster with energy above 20 GeV in the event hence underestimating the background fraction for $\mathrm{E_T} < 20$ GeV. Besides the listed background a Minimum Bias sample has been used to optimize correctly the cuts in this low energy range.

In the barrel the estimated fake rate goes from 10^{-2} at 12 GeV to 10^{-1} for high energy electrons in the "VeryLoose" selection. It is largely reduced for "Hyper Tight1" selection: $5{\times}10^{-3}$ at 12 GeV to a $5{\times}10^{-2}$ for higher energy electrons. Roughly each intermediate level of selection provides almost a factor 2 rejection on the background.

Regarding the endcap it can be done a similar remark as the one done for the efficiency. Setting the cuts the overall efficiency has been maximized tightening the selection in the endcap where the background is higher. This results in a smaller fake rate in the endcap for each selection level.

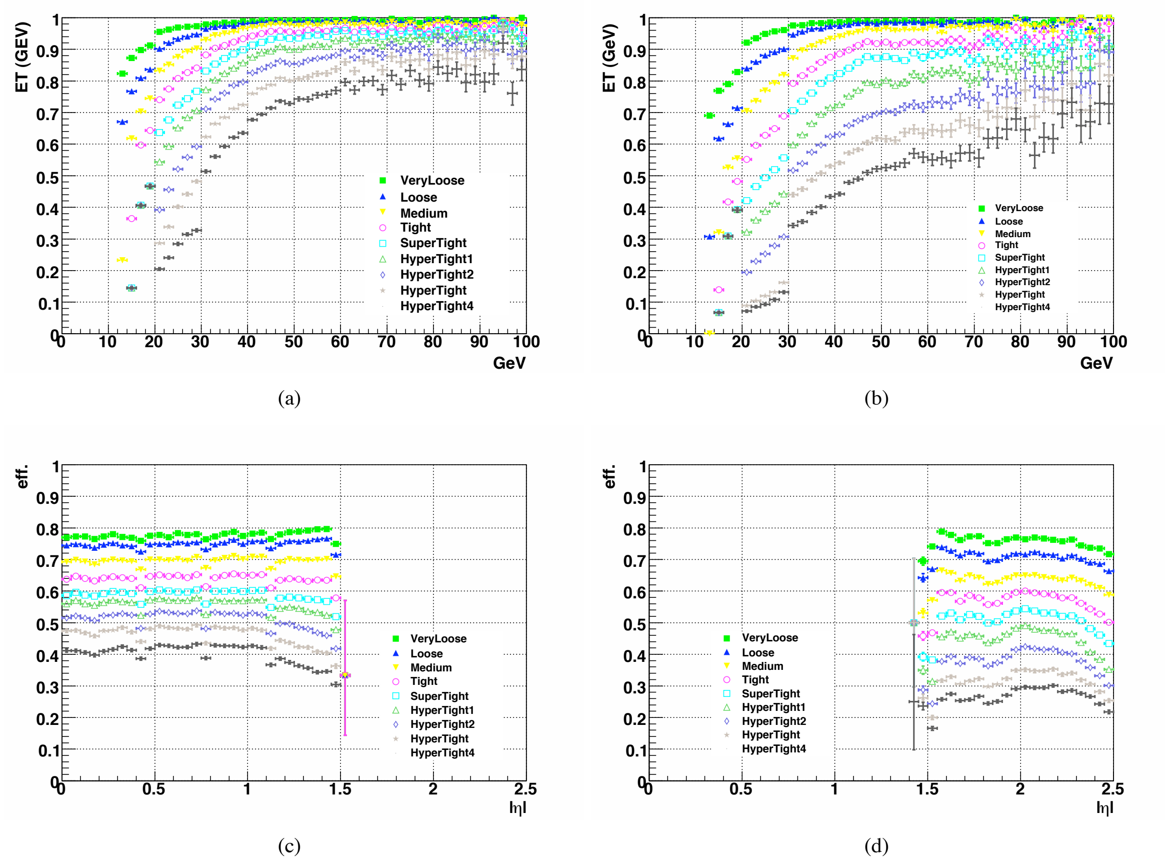
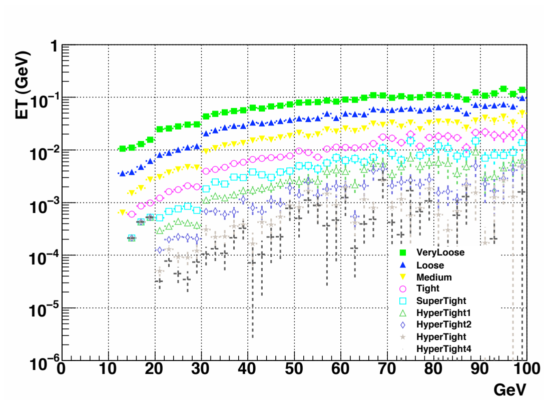
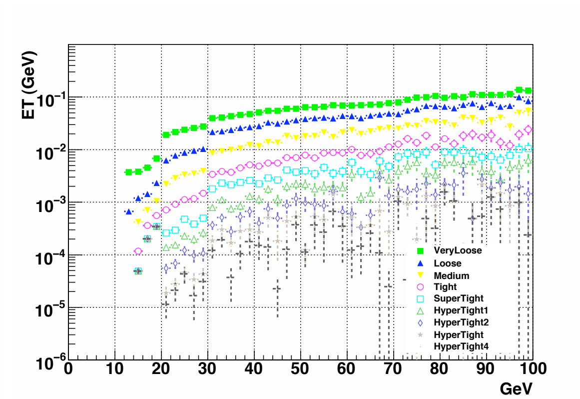


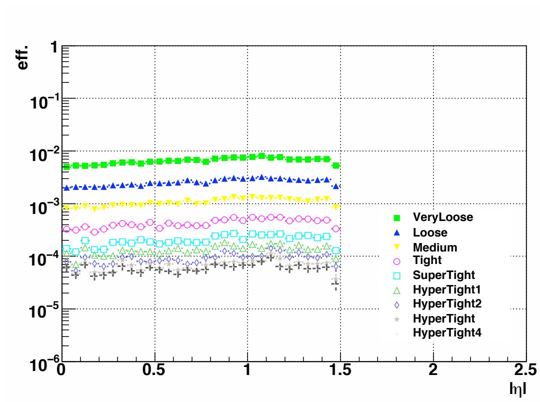
Figure 9: Selection efficiencies as a function of $\mathrm{E_T}$ and η for different selection tightness for Category Based Electron ID in $\mathrm{Z} \rightarrow \mathrm{e^+e^-}$ events. The efficiencies are reported separately for barrel (a)(c) and endcap (b)(d).



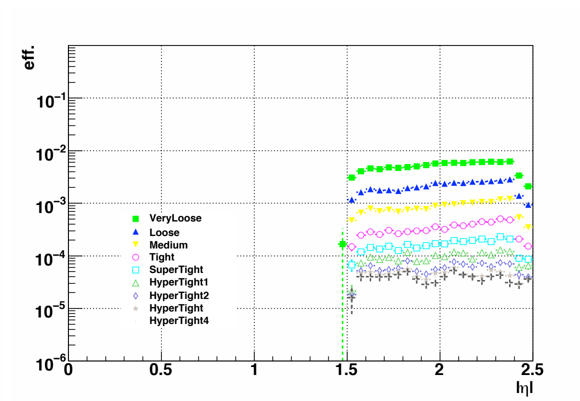
(a)



(b)



(c)



(d)

Figure 10: Background rejection quoted as the fraction of SuperCluster belonging to tag electro as a function of $\mathrm{E_T}$ and η for different selection tightness for Category Based Electron ID. The fake rate is reported separately for barrel (a)(c) and endcap (b)(d).

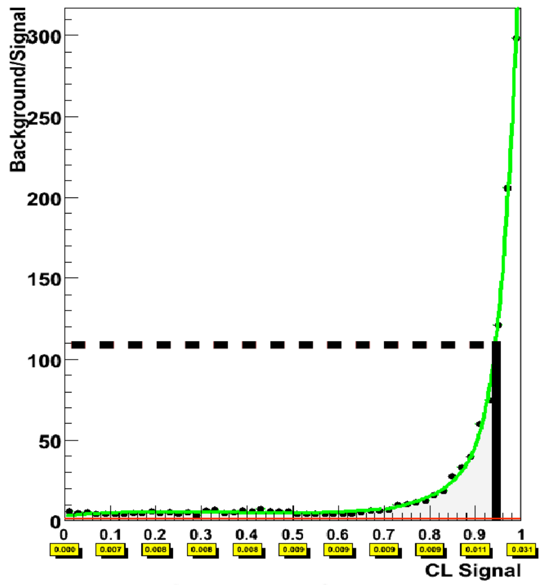


Figure 11: Example of n-1 b/s distribution for $\sigma_{\mathrm{i}\eta\mathrm{i}\eta}$ variable. The orizontal line correspond to the chosen b/s value, the vertical one defines the cut value.

4 Optimization techniques

4.1 Differential S/B method

Rather than use a cut optimizer that can easily over-train on low statistics data, a simple procedure to set the cuts is used:

- produce the "n-1" distributions for each cut variable, that is, distributions where all the cuts have been applied besides.
- use a very safe procedure to fit a smooth curve to the background to signal ratio as a function of this one cut variable in question.
- set the cut for this variable at a pre-specified value of b/s.
- set cuts using the same b/s specification for all the cut variables and iterate a few times.

In this way each cut is removing events with the same purity, that is the same signal to background ratio, thus the overall purity of accepted electrons is maximized for a given efficiency. This clearly means that cuts will be "tighter" in the bad-track category, than in the bremming-electron category, and tighter in the endcap than in the barrel.

The Fig. 11 shows an example b/s distribution with the cut chosen for a particular b/s specification. It has been found that this procedure gives a stable result even for very low statistics signal and background samples.

The result of this procedure to set the cuts clearly must depend on the signal and background samples chosen. For our baseline electron id cuts, we have used di-electrons with masses above $40\ \mathrm{GeV/c^2}$ which is dominated by the Z resonance as signal. For background we have used di-jet events. The cut selection is done at the single electron level. With this choice of signal and background, the background to signal ratio decreases rapidly with $\mathrm{E_T}$ for $\mathrm{E_T}$ less than about half the Z mass. Low $\mathrm{E_T}$ electrons can dominate the fake rate if measures are not taken to deal with the $\mathrm{E_T}$ dependence. At the same time, we want good efficiency and simply cutting out low $\mathrm{E_T}$ electrons would be a bad choice. Therefore, the cuts are set in three $\mathrm{E_T}$ regions, $\mathrm{E_T} > 30$ GeV, $20 < \mathrm{E_T} < 30$ GeV and $12 < \mathrm{E_T} < 20$ GeV. Electrons below 12 GeV $\mathrm{E_T}$ are cut out for this identification primarily used to detect electrons from Z and W decay. We expect to have other dedicated sets of cuts for low $\mathrm{E_T}$ electrons based on signals other than the Z. In addition to the $\mathrm{E_T}$ bins, we scale the isolation cuts with $\mathrm{E_T}$ to enable us to smoothly loosen the cuts as $\mathrm{E_T}$ increases even within one of these $\mathrm{E_T}$ bins.

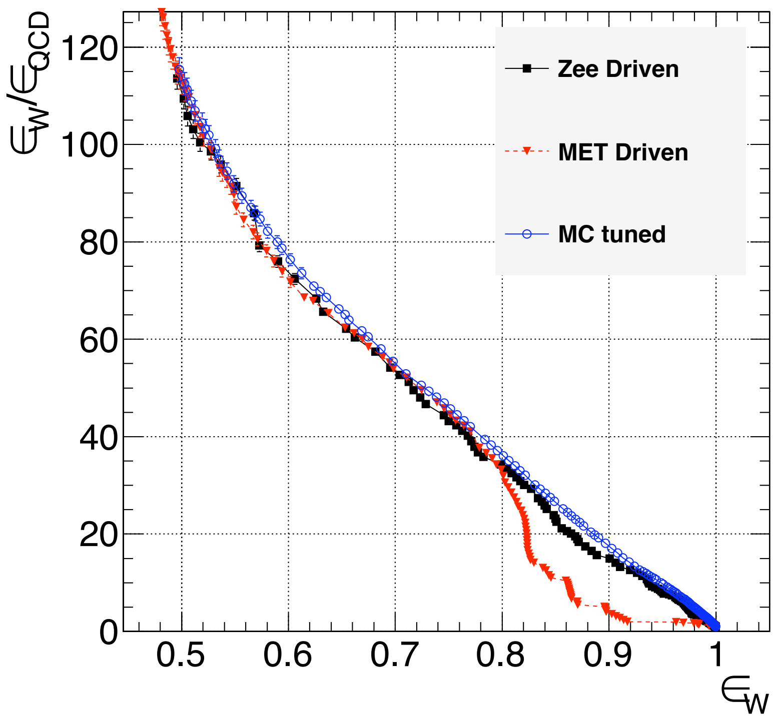


Figure 12: Demonstration of the iterative technique with realistic signal and background samples for 10 pb^{-1}

At this time, we have determined cuts for 9 different b/s values. Each level of cuts reduces the background rate by about a factor of 2, thus the cuts range from very high efficiency that might be useful in multilepton events, to very large background rejection that might be useful to detect single electrons.

4.2 Iterative Method

An alternative method, or algorithm, to tune cuts for electron selection has been investigated. It aims to provide, for any given tightness of selection, the optimal sharing of the cutting power (and efficiency loss) between the cutting variables, demanding that the cuts values are chosen, for any given overall background rejection, so that the overall efficiency is maximum. The details of the algorithm and its testing with Autumn08 (CMSSW 2.1) MC samples can be found in [6] (the software is available in /UserCode/EGamma/EGammaElectronAnalysis/Part3/). The algorithm has so far been used for tuning cuts on relatively small number of selection variables for the inclusive W and Z cross-section measurement analysis in the electron channel [7]. For this usage the barrel and endcap regions are taken as separate categories and isolation variables are treated as selection variables on the same footing as the other electron ID variables.

The method requires 2 inputs: true electron candidates ("signal"), and fake electron candidates ("background"). These inputs can be pure, as can be obtained using MC truth, or contaminated, as will be the case for samples obtained from data. The algorithm was tested during the October 2009 exercise. For this test the following selection variables were used: $\sigma_{i\eta i\eta}$, HoE, $\Delta\phi_{\rm in}$, $\Delta\eta_{\rm in}$, and the standard electron isolation variables (cone of 0.4) for ECAL, HCAL and track isolation. The results can conveniently be shown as a trajectory in the background rejection versus signal efficiency plane, where each point on the trajectory represents a set of selection cuts obtained by the algorithm.

Figure 12 shows the results of a study performed on Summer09 MC samples taking electron candidates with $p_T > 30$ GeV/c, and where the electron supercluster falls within the ECAL fiducial region ($|\eta| < 2.5$ and the transition region $1.4442 < |\eta| < 1.56$ excluded). In this particular example the $\Delta\phi_{\rm in}$ cut has been constrained to be no smaller than 0.02. This was found appropriate in the W cross-section analysis, since tighter cuts on $\Delta\phi_{\rm in}$ result in an electron efficiency which varies too strongly with η. The performance of the selection without this constraint is slightly better in terms of signal efficiency versus background rejection.

The axes of Fig. 12 refer to the useful background rejection ($\epsilon_{signal}/\epsilon_{background}$) and ϵ_{signal} that are measured with pure Monte Carlo samples of $\mathrm{W} \rightarrow \mathrm{e}\nu$ as "signal" and QCD jet samples as "background"[2)]. The blue line joining the points marked with open circles shows the series of optimized selection points that are obtained when the algorithm is run with pure $\mathrm{W} \rightarrow \mathrm{e}\nu$ and pure QCD jet samples as signal and background inputs.

The other lines in Fig. 12 show what happens when realistic impure samples are used as input to the algorithm – although what is plotted at each point on the line is the background rejection and signal efficiency measured with pure Monte Carlo samples. For both lines the input background samples are single electron triggers with missing $\mathrm{E_T}$ cut ($\not{E}_\mathrm{T} < 20\ \mathrm{GeV/c^2}$) and events containing a second electron candidate with $\mathrm{E_T} > 20$ GeV removed. This selects an adequately pure sample of background. Calorimetric $\not{E}_\mathrm{T}$ is used here, although it is worth mentioning that track corrected and particle flow $\not{E}_\mathrm{T}$ provide a cleaner separation of $\mathrm{W} \rightarrow \mathrm{e}\nu$ events from jet background events in simulated events reconstructed with CMSSW 3.

The red line joining the points marked by solid triangles shows what happen when the input signal sample consists of single electron triggers with a missing $\mathrm{E_T}$ cut ($\not{E}_\mathrm{T} > 30\ \mathrm{GeV/c^2}$). It can be seen that only when the selections have reached some level of tightness, thus cleaning the input signal sample, do we manage to obtain the same level of optimization as that obtained with the Monte Carlo pure samples. Once this is achieved, at a signal efficiency of about 80%, the selections obtained by the realistic impure input samples track those obtained with Monte Carlo pure samples extremely well. The black line joining the points marked with solid squares shows what happens when the input signal is a rather pure electron sample obtained using a tag and probe like technique to select electrons from Z decays. This line tracks the MC truth line rather closely.

5 Multivariate techniques

An alternative approach to the cut-based selection described above is the usage of a multivariate technique based on a likelihood algorithm combining several discriminating variables.
The same electron classification discussed in Sec. 2.2 is used to improve the discriminating power of the algorithm. Since most of the jets mis-reconstructed as electrons are classified as showering, the big brem and narrow classes are poorly populated and this makes the modelling of the probability density functions (*PDFs*) for the two intermediate classes difficult. Gloden, big brem and narrow electrons are therefore merged in in a unique class, characterized by having a single cluster in the ECAL, which we define as *non-showering*.
A set of 6 variables is used in the likelihood construction: H/E, $\sigma_{\mathrm{i}\eta\mathrm{i}\eta}$, $\Delta\phi_{\mathrm{in}}$, $\Delta\eta_{\mathrm{in}}$, $\mathrm{E_{seed}/p_{out}}$ and $\sum_9 / \sum_{25}$ Probability Density Functions (PDFs) are constructed for each variable from control samples on data, as described in Sec. 6.
The cross-correlation between these variables was checked to be small enough (at the percent level) both for signal and background. The variables are therefore combined to compute the likelihood $L_{k,c}(\xi)$ for:

- two particle hypotesis $\xi = \{e, jet\}$,
- 4 kinematic bins
  $k = \{(\mathrm{p_T} < 15\mathrm{GeV/c}; \mathrm{b}arrel), (\mathrm{p_T} > 15\mathrm{GeV/c}; \mathrm{b}arrel), (\mathrm{p_T} < 15\mathrm{GeV/c}; \mathrm{e}ndcap), (\mathrm{p_T} > 15\mathrm{GeV/c}; \mathrm{e}ndcap)\}$,
- 2 electron classes
  $c = \{\mathrm{n}on - showering, showering\}$:

The likelihood function is defined as the product of the single variable PDF ($\mathcal{P}_{k,c}(x;\xi)$):

$$\begin{aligned} L_{k,c}(\xi) \quad &= \quad \mathcal{P}_{k,c}(\mathrm{E_{seed}/p_{out}};\xi) \cdot \mathcal{P}_{k,c}(H/E;\xi) \cdot \mathcal{P}_{k,c}(\Delta\eta_{\mathrm{in}};\xi) \cdot \\ &\quad\quad \mathcal{P}_{k,c}(\Delta\phi_{\mathrm{in}};\xi) \cdot \mathcal{P}_{k,c}(\sum_9 / \sum_{25};\xi) \cdot \mathcal{P}_{k,c}(\sigma_{\mathrm{i}\eta\mathrm{i}\eta};\xi) \cdot \end{aligned} \tag{1}$$

Weighting the individual likelihoods with their *a priori* probabilities p_ξ, we define the likelihood ratio as:

$$r = \frac{p_e L(e)}{p_e L(e) + p_{jet} L(jet)} \tag{2}$$

Since the a priori probabilities depend on the trigger settings, and these are not yet defined, we set them all equal to 1, i.e. assuming no a priori knowledge.

[2)] QCD heavy and light flavor dijets and γ+jets.

6 Shapes from data

Obtaining identification variables shapes using data, descriptions of shape extraction from data: selection used for Cut-based and Likelihood PDF, similar strategies for the signal (Tag&Probe like) but different for the background.

6.1 Cut-Based Identification

The identification and isolation variable shapes are checked using data. A set of pure signal electrons from $\mathrm{Z} \to \mathrm{e}^+\mathrm{e}^-$ and background electrons from QCD di-jet events are extracted from data to measure all the distributions. Moreover, for every set of cuts (for example the "Loose" one) the distributions of electron identification (isolation) variables are checked by applying either only the electron isolation (identification) cuts or all the cuts except the one on the variable under study (N-1 plot). Data and MC are compared at all three of these levels to see what discrepancies are.

6.1.1 Signal selection

The goal is to select events with electrons from $\mathrm{Z} \to \mathrm{e}^+\mathrm{e}^-$ decay with the littlest contamination of fake electrons from QCD or real electrons from either $\mathrm{W} \to \mathrm{e}\nu$ or (semi)leptonic $\mathrm{t\bar{t}}$ decays. The selection stategy is Tag&Probe-like, with a tight selection on one electron ("Tag") and no requirements on the second electron ("Probe") in order to have an unbiased sample of electrons from Z decay. Events with two reconstructed electrons with SuperCluster transverse energy $\mathrm{E_T^{SC}}$ greater than 20 GeV are selected. A cut on the missing transverse energy $\not{E}_\mathrm{T} < 30\ \mathrm{GeV/c^2}$ is applied to reject the fraction of $\mathrm{W} \to \mathrm{e}\nu$ events with a jet faking the second electron.
Each electron is scanned for eligibility as a "Tag", defined as a reconstructed electron with this properties:

- is in ECAL fiducial region: $|\eta|$ <1.4442 (barrel) or 1.560< $|\eta|$ <2.5 (endcaps),
- is associated to an HLT electron candidate which triggered the event,
- passes the "SuperTight" identification and isolation cuts.

For each "Tag" electron found in the event, the other electrons are labeled as "Probe" if the invariant mass of the Tag&Probe pair is compatible with the reconstructed Z mass value within $5\ \mathrm{GeV/c^2}$. There can be more than one "Probe" electrons per event, for example if both are "Tag" electrons satisfying the mass compatibility.
The number of signal $\mathrm{Z} \to \mathrm{e}^+\mathrm{e}^-$ events selected per $10\ \mathrm{pb}^{-1}$ is 3k while 70 are the background events, resulting in a S/B ratio of about 40. The statistics of "Probe" electrons with $\mathrm{E_T^{SC}} > 20$ GeV to study the electron id. and isolation distributions is 3.6k in the barrel and 1.5k in the endcaps.
At start-up a looser identification selection will be applied to select the "Tag" electrons and a wider window around the Z mass central value will be opened when pairing "Tag" and "Probe" candidates.

6.1.2 Background selection

Background electrons are selected from QCD di-jet events requiring at least one jet reconstructed with the iterative cone algorithm with uncorrected transverse energy $\mathrm{E_T}^{\mathrm{uncorr}}$ exceeding 20 GeV. The highest-$\mathrm{E_T}^{\mathrm{uncorr}}$ jet is elected as "Tag" if its electromagnetic energy less than 90% of the total energy. A total of 13.2 millions events are selected per $10\ \mathrm{pb}^{-1}$ with 4.6k $\mathrm{W} \to \mathrm{e}\nu$ and 1.5k $\mathrm{Z} \to \mathrm{e}^+\mathrm{e}^-$ decays.
All the electrons found outside the jet cone ($\Delta\mathrm{R} \geq 0.4$) associated to an HLT electron candidate firing the trigger and passing a minimal identification set of cuts (the "VeryLoose" identification or isolation cuts) are labeled as "Probes" and form the set of background electrons. The background electrons with $\mathrm{E_T^{SC}} > 20$ GeV collected are 211k in the barrel and 313k in the endcaps.

As an example, the distributions of the electron variables used for electron identification are shown in Figures 13 and 14 for the barrel only. The distributions that can be obtained from data are represented with points for both the signal and background selections. The histograms superimposed represent the distribution of electrons from $\mathrm{Z} \to \mathrm{e}^+\mathrm{e}^-$ decay from the signal selection and fake electrons in QCD di-jet and γ+jet events from the background selection.

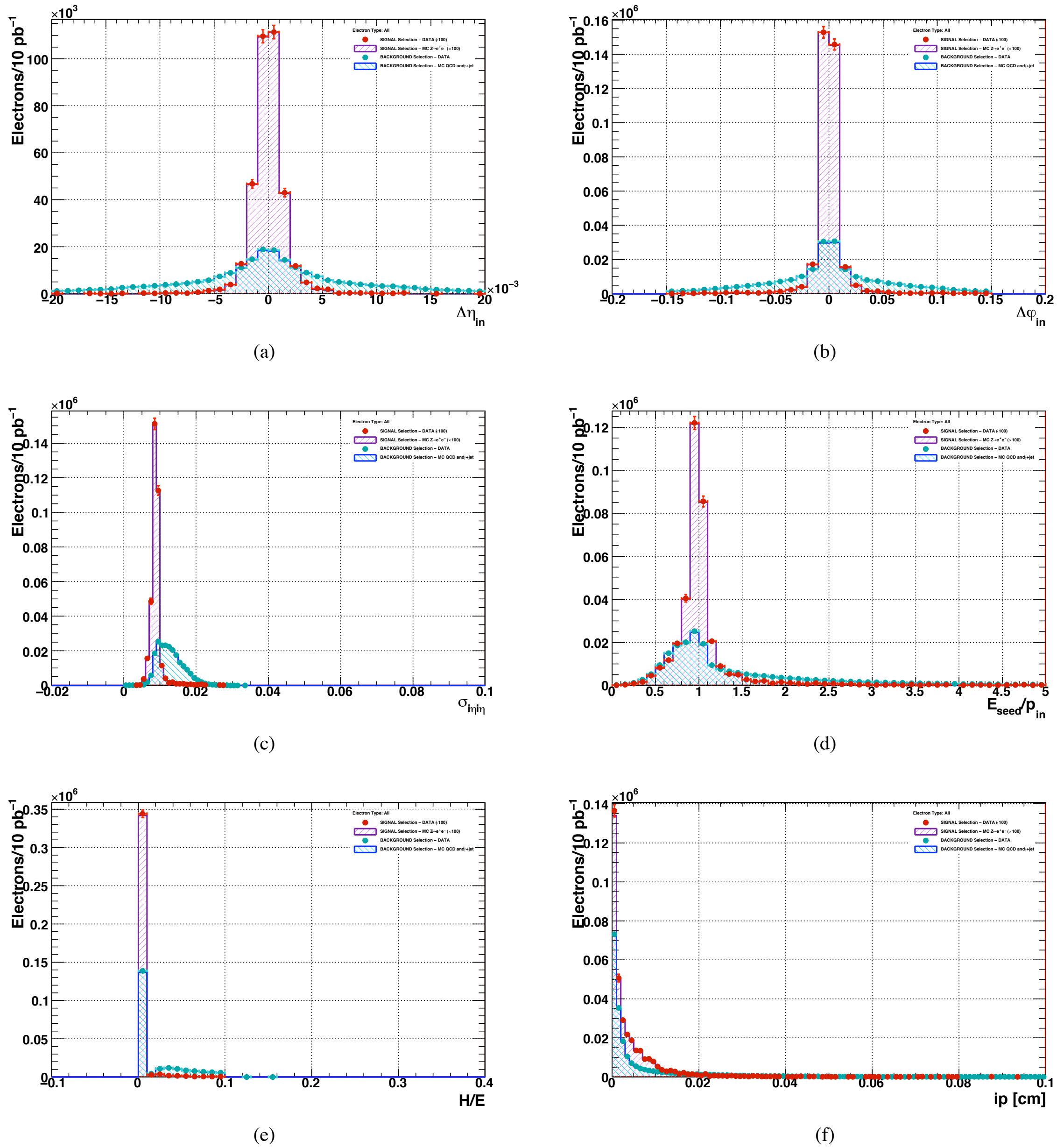


Figure 13: Electron identification variable distributions for signal and background electrons in ECAL barrel: $\Delta\eta_{\rm in}$ (a), $\Delta\phi_{\rm in}$ (b), $\sigma_{\rm i\eta i\eta}$ (c), fbrem (d), H/E (e) and track transverse impact parameter. Distributions that can be obtained from data are represented with points. The histograms are the distribution of $\rm Z \rightarrow e^{+}e^{-}$ electrons from signal selection and QCD di-jet and γ+jet fake electrons from background selection. Signal distributions are multiplied by 100.

6.2 Likelihood PDFs

In this section we discuss the extraction of the PDFs for the variables used in the likelihood algorithm from data, similarly to what was done in the previous section for the Cut-Based approach. The distributions as extracted from a pure Monte Carlo sample are compared with those extracted from a data-like sample after a background subtraction procedure.

6.2.1 Signal selection

Real electrons are selected with a Tag&Probe procedure similar to the one discussed in the previous section. Both the ‘Tag’ and the ‘Probe’ electrons are required to have transverse energy $\rm E_T^{SC}$ greater than 5 GeV and to fall within the ECAL fiducial region. No further requirements are applied on the ‘Probe’ electron, while the ‘Tag’ electron

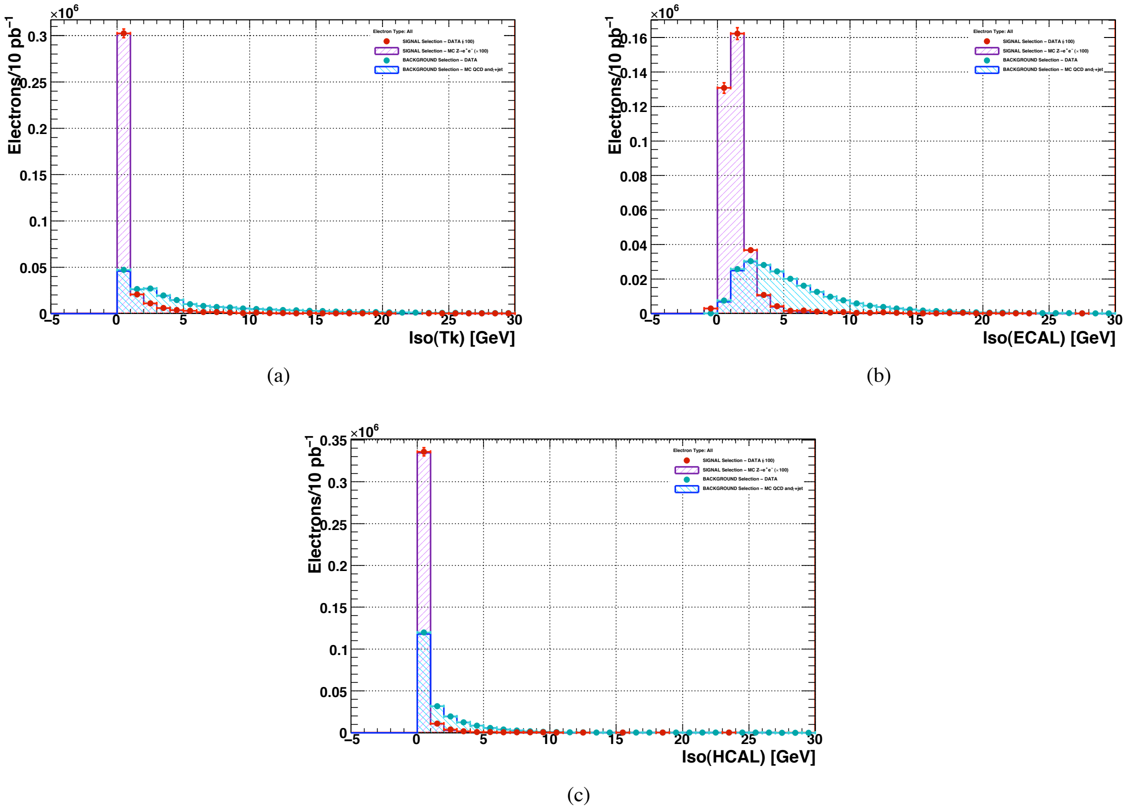


Figure 14: Electron isolation variable distributions for signal and background electrons in the ECAL barrel: Tracker (a), ECAL (b) and HCAL (c). Distributions that can be obtained from data are represented with points. The histograms are the distribution of $\mathrm{Z} \rightarrow \mathrm{e}^{+}\mathrm{e}^{-}$ electrons from signal selection and QCD di-jet and γ+jet fake electrons from background selection. Signal distributions are multiplied by 100.

is asked to be loosely identified and isolated. The loose isolation requirement is $\sum \mathrm{p_T}/\mathrm{p_T}^{\mathrm{electron}} < 0.2$, the sum running over the tracks in a cone of $\Delta\mathrm{R} < 0.4$ around the electron track. Offline reconstructed electrons are not required to match a HLT electron candidate. To have the largest electron sample, if the electron ‘Probe’ also fulfill the quality criteria for a ‘Tag’, the roles are inverted and the first electron is used as a ‘Probe’.

For each ‘Tag’ electron found in the event, the ‘Probe’ electron bringing to the invariant mass closest to the Z mass is retained. A different approach is used here with respect to the Cut-Based analysis case. A loose selection is applied on the invariant mass of the Tag&Probe pair

- $60\ \mathrm{GeV/c^2}, 110\ \mathrm{GeV/c^2}$

We then apply a statistical background subtraction which makes use of the full Z lineshape extracted from data and we assign to any event the probability to be signal or background through a maximum likelihood fit to the di-electron invariant mass. The invariant mass for signal is modeled with a function defined as:

$$f(x; m, \sigma_L, \sigma_R, \alpha_L, \alpha_R) = N \times \exp\left[-\frac{(x-m)^2}{2\sigma_{L/R}^2 + \alpha_{L/R}(x-m)^2}\right] \tag{3}$$

where the σ_L and α_L (σ_R and α_R) corresponds to resolution and tail parameters of the distribution for $x - m < 0$ ($x - m > 0$). The QCD background is parameterized with a second order polynomial. The parameterization of the signal and the background components are shown in Fig. 15

To apply our strategy on a sample similar to the one that is selected on data, we merge signal and background events to get $10\ \mathrm{pb}^{-1}$ equivalent data. We then perform an unbinned maximum likelihood fit to this dataset fixing the background shape to Monte Carlo, while leaving the signal lineshape floating as well as the signal and background yields. As output of the fit we get the signal and background yields, as well as the signal function parameters, consistent with the expected values. The fit to the data-like sample is shown in Fig. 16.

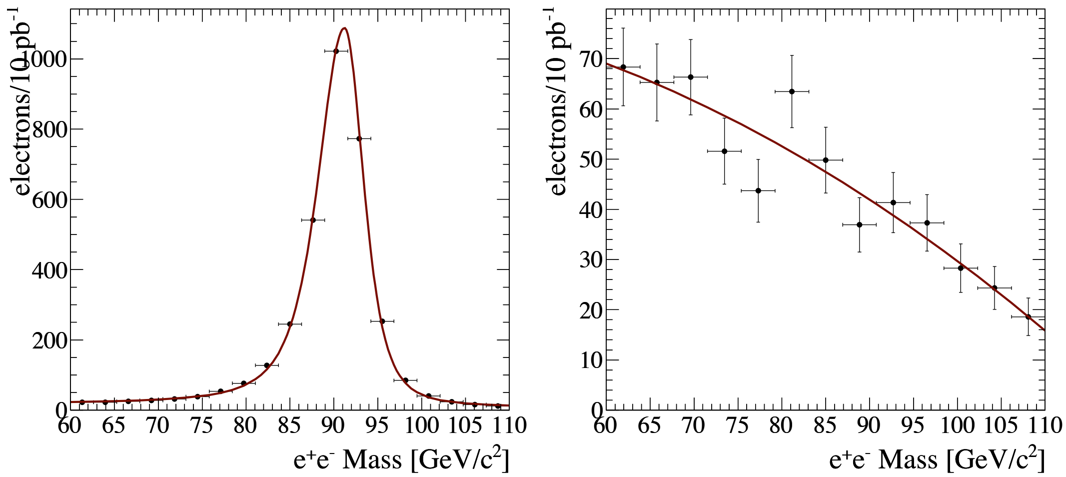


Figure 15: Distribution of the tag and probe electrons invariant mass for $\mathrm{Z} \rightarrow \mathrm{e}^{+}\mathrm{e}^{-}$ events (left) and QCD events (right). The result of the fit with the parameterization discussed in the text is superimposed.

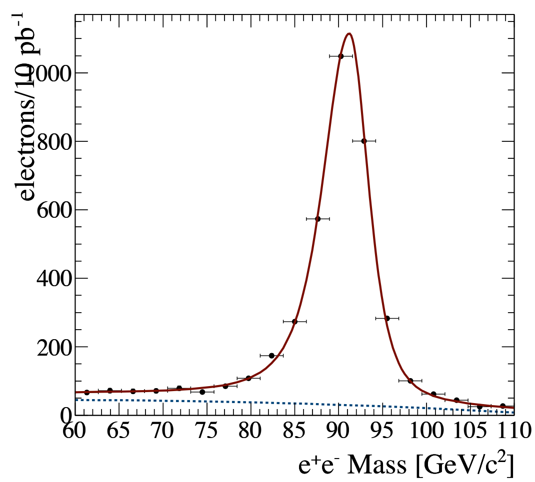


Figure 16: Distribution of the Tag&Probe electrons invariant mass for a data-like merged sample formed by $\mathrm{Z} \rightarrow \mathrm{e}^{+}\mathrm{e}^{-}$ and QCD events corresponding to an integrated luminosity of $10\ \mathrm{pb}^{-1}$. The result of the fit with the parameterization discussed in the text is superimposed.

The value of the likelihood is then used to compute the signal $_sWeight$ [9], which is proportional to the probability for that event of being signal. We form the distributions of electron identification variables weighting each event with its signal $_sWeight$. We call these distributions $_sPlots$. Properties of $_sWeight$ are that the resulting distributions are background subtracted with the correct normalization and correct uncertainties.

The distributions of the variables used in the likelihood algorithm for electrons from $\mathrm{Z} \rightarrow \mathrm{e}^{+}\mathrm{e}^{-}$ are shown in Figures 17 (barrel only). The distributions obtained from data after the background subtraction procedure described above (dots) are compared with references obtained from a pure $\mathrm{Z} \rightarrow \mathrm{e}^{+}\mathrm{e}^{-}$ sample after the signal selection A very good agreement is obtained.

6.2.2 Background selection

Background electrons are selected from QCD di-jet events passing a jet HLT path. The highest transverse energy jet reconstructed with the SisCone [8] algorithm and falling in the pseudorapidity region $|\eta|$<2.5 is used as a ‘Tag’. The ‘Tag’ jet corrected transverse energy $\mathrm{E_T}^{\mathrm{corr}}$ is required to exceed $30\ \mathrm{GeV}/c$. No selection is applied on the jet electromagnetic energy. Fake electrons reconstructed in the jet samples are used as ‘Probes’. Probe candidates are electrons reconstructed within the ECAL acceptance ($|\eta| < 2.5$) and have transverse energy $\mathrm{E_T^{SC}}$ greater than

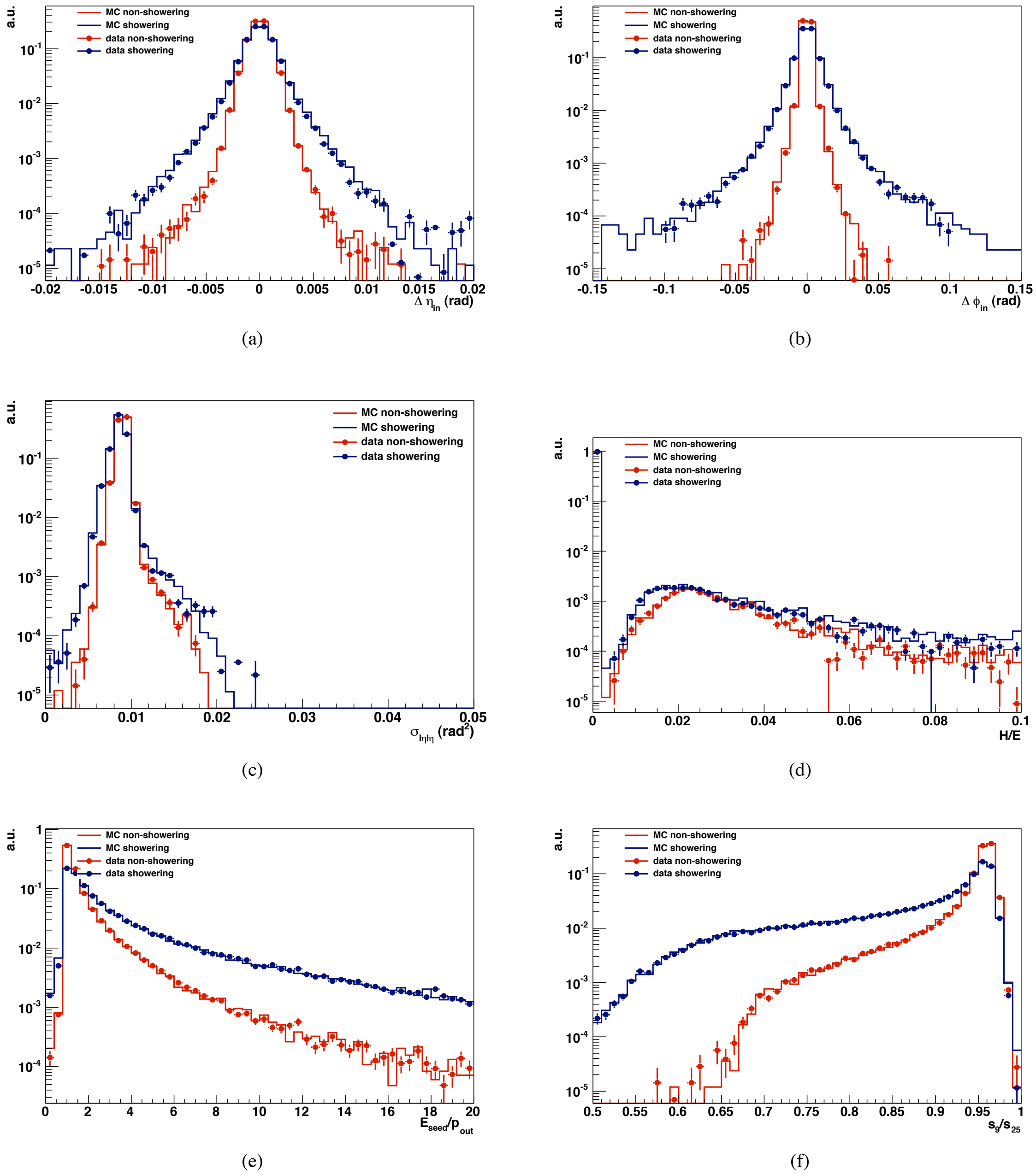


Figure 17: Electron identification variables distributions for signal electrons in the ECAL barrel: $\Delta\eta_{\mathrm{in}}$, $\Delta\varphi_{\mathrm{in}}$, $\sigma_{\mathrm{i}\eta\mathrm{i}\eta}$, H/E, $\mathrm{E_{seed}/p_{out}}$, $\sum_9 / \sum_{25}$ Distributions that can be obtained from data are represented with points. The histograms are the distribution of $\mathrm{Z} \rightarrow \mathrm{e^+e^-}$ electrons from signal selection. Showering and not showering electrons are shown separately

5 GeV. If several Probe candidates are reconstructed the one with the largest distance in the transverse plane with respect to the tag is chosen.

In order to reject events coming from $\mathrm{Z} \rightarrow \mathrm{e^+e^-}$ decays, if more than one electron is reconstructed in the event (the probe plus an additional one), the invariant mass of the two ones with the highetst $\mathrm{p_T}$ is computed. If that is consistent with Z nominal mass ($\mathrm{m_{e^+e^-}} > 70\ \mathrm{GeV/c^2}$), the event is discarded. This request reduces the $\mathrm{Z} \rightarrow \mathrm{e^+e^-}$ contamination to a negligile level, and it is more efficient in vetoing Z's than a requirement on probe-'Tag' invariant mass because of the poorer energy estimation for the tag jet with respect the electrons.

In order not to select photons produced in association with jest as Probes, the Probe candidate is required to be not isolated both in ECAL and in the tracker:

- $\sum \mathrm{p_T}^{\mathrm{tracks}} / \mathrm{p_T}^{\mathrm{electron}} > 5$,
- $\sum \mathrm{E_T}^{\mathrm{ECALhits}} / \mathrm{E_T}^{\mathrm{electron}} > 2$,

considering a cone of $\Delta \mathrm{R} < 0.4$ around the electron. With this requirement the number of photons plus jets events where the probe is the photon (which could give a electron-like ECAL cluster shape) is reduced to a negligible level. The surviving events, though not negligible, are the events where the fake electron come from the jet, so consistent with the QCD di-jet events. We use two kinematic variables to select the di-jet events, namely:

- $\Delta\phi$, the opening angle between the Tag and the Probe in the transverse plane
- the calorimetric missing transverse energy.

As in the signal case, a statistical background subtraction is applied to assign to any event the $_sWeight$ performing a 2-dimensional maximum likelihood fit to the selected sample wih these variables. In the fit the selected γ+jets and QCD di-jets events are considered as signal. Given the small number of expected events the $\mathrm{Z} \rightarrow \mathrm{e^+e^-}$ contribution is neglected. Events from $\mathrm{W} \rightarrow \mathrm{e}\nu$ or (semi)leptonic $\mathrm{t\bar{t}}$ decays are considered separately as backgrounds.

The functional form in Eq.3 is used to parameterize the transverse missing energy for the signal and the two background species. $\Delta\phi$ is parameterized with Eq.3 function for the signal, a second order polynomial for the $\mathrm{W} \rightarrow \mathrm{e}\nu$ background and a third order polynomial for $\mathrm{t\bar{t}}$. The parameterization of the signal and the background components for the two variables used in the fit are shown in Fig. 18

We produce a Monte Carlo sample representative of $10\ \mathrm{pb}^{-1}$ of data after that the selection has been applied merging toghether the simulated events of the signal and background sources described before. On this data-like sample we perform an unbinned maximum likelihood fit, we compute the $_sWeight$ and we use this as an event weight to produce background subtracted distributions of the electron identification variables for the fake electrons.

Figures 19show the distribution of the $_sPlots$ for fake electrons in the barrel only. As for the signal, the Figures above show a good agreement between the variables computed on a pure QCD di-jet sample and the ones obtained on data-like sample after the background selection is applied.

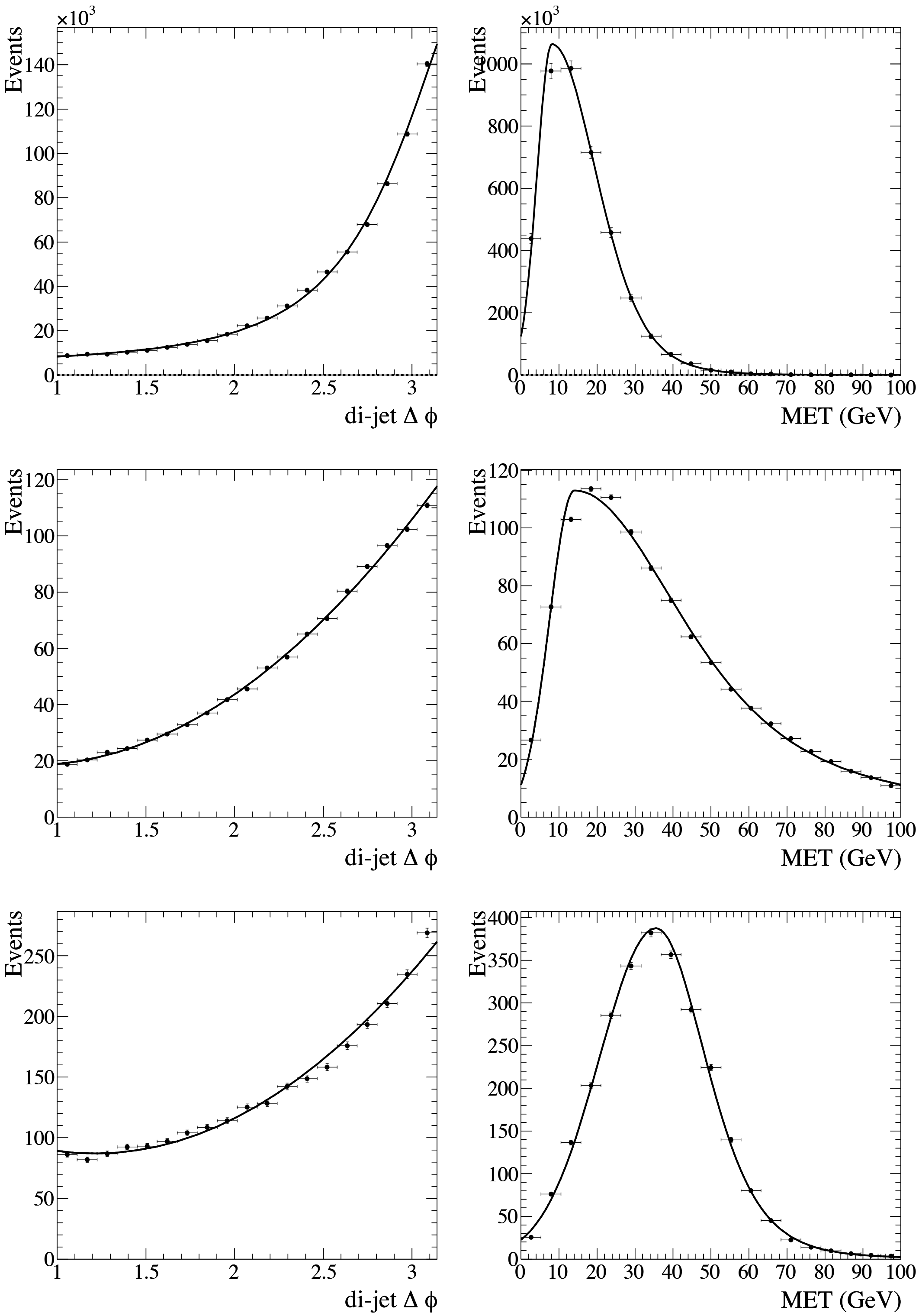


Figure 18: Distribution of the opening angle between the Tag and the Probe in the transverse plane (left) and the missing transverse energy (right) for QCD and photon plust jets (top), $t\bar{t}$ (middle) and $W \to e\nu$ (bottom) events. The result of the fit with the parameterizations discussed in the text is superimposed.

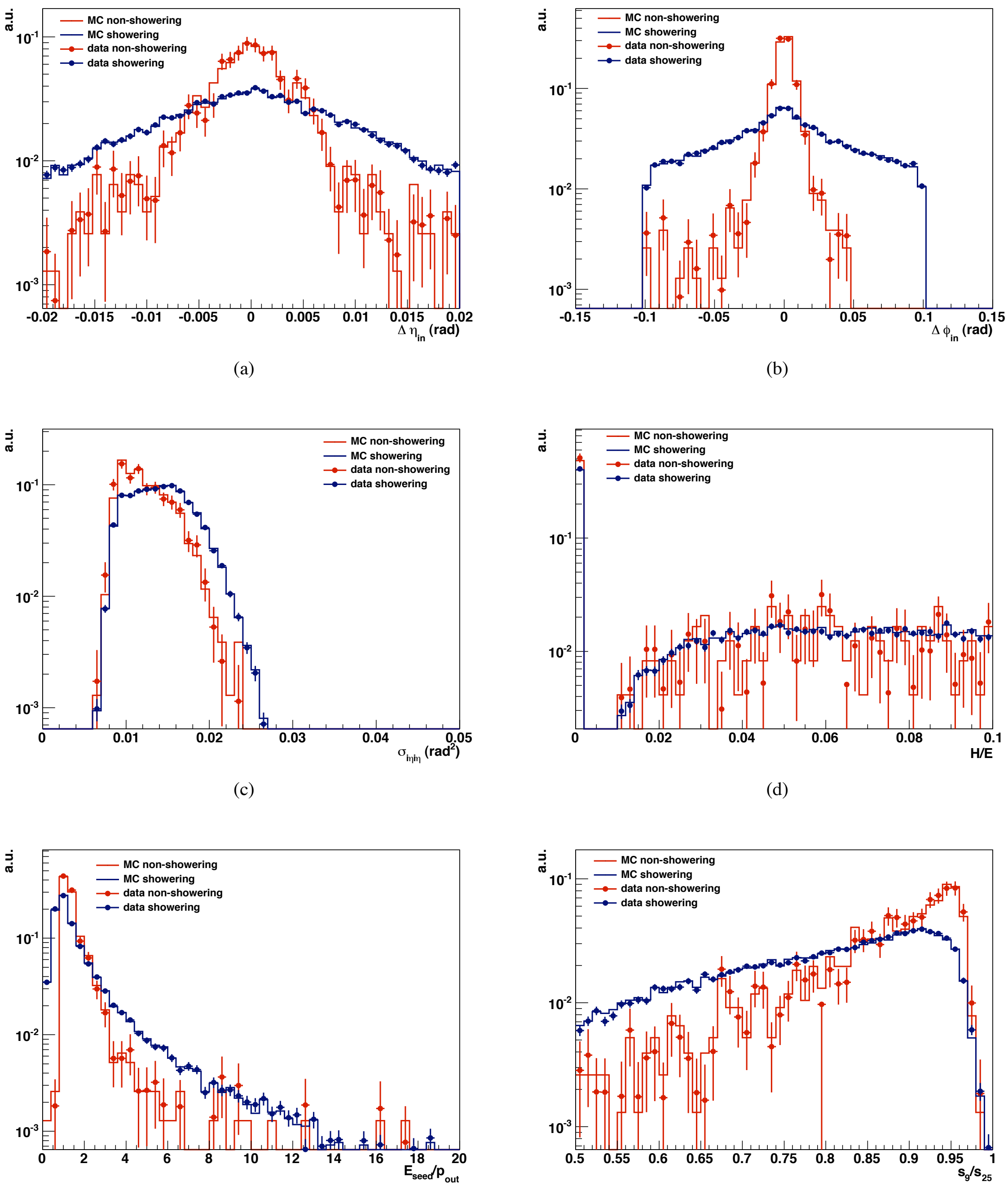


Figure 19: Electron identification variables distributions for fake electrons from backgrounds in the ECAL barrel: $\Delta\eta_{\rm in}$, $\Delta\phi_{\rm in}$, $\sigma_{\rm i\eta i\eta}$, H/E, $\rm E_{seed}/p_{out}$, $\sum_9/\sum_{25}$ Distributions that can be obtained from data are represented with points. The histograms are the distribution for fake electrons reconstructed in QCD di-jets samples with the background selection. Showering and not showering electrons are shown separately

7 No isolated electrons

Electrons in jets are mainly reconstructed and identified using a technique developed in the context of the particle-flow event reconstruction [10] [11].

A dedicated particle-flow clustering algorithm has been developed which, thanks to the good ECAL granularity, is able to separate overlapping showers. Within a jet, it allows the energy deposits of the hadrons and from the electrons to be disentangled. However, it also means that the electron cluster and each of the Bremsstrahlung photon clusters will be often reconstructed separated, hence the need for a strategy to gather them together, as to compute the total energy deposit of the electron.

In the particle flow, a GSF-track-driven Bremsstrahlung recovery strategy has been developed in order to collect the energy deposited in the ECAL and in the preshower by the emitted Bremsstrahlung photons. Starting from a GSF track, a tangent is extrapolated at each track measurements towards the ECAL in an attempt to mimic a Bremsstrahlung emission. A geometrical matching is then performed between the ECAL clusters and the extrapolated positions of the GSF track and each of the track tangents. The ECAL clusters associated to other KF tracks (charged hadrons) are discarded during this procedure.

For a pseudo-rapidity $|\eta| > 1.6$ the recovered ECAL clusters are then geometrically matched also with preshower clusters. With this technique, it is possible to build a particle-flow supercluster (cluster of particle-flow clusters) that contains all the ECAL and preshower energy of the initial electron with a limited absorption of the jet energy.

The most important observable for the electron-charged hadron discrimination is the energy matching between the track momentum and the ECAL. Using the particle-flow supercluster, three discriminating observables based on the energy matching can be built. These are $\mathrm{E}_e/\mathrm{p}_{\mathrm{out}}$ (momentum-energy match for electron at the ECAL); $\mathrm{E}_{\gamma(s)}/(\mathrm{p}_{\mathrm{in}} - \mathrm{p}_{\mathrm{out}})$ (Bremsstrahlung photon energy matches the change in track momentum); and $\mathrm{E}_e{+}\mathrm{E}_{\gamma(s)}/\mathrm{p}_{\mathrm{in}}$ (supercluster energy matches the initial momentum). Moreover two boolean observables $EarlyBrem$ and $LateBrem$, correlated with a possible bias of the track-ECAL energy matching, are introduced. Indeed when a Bremsstrahlung photon emission occurs in the pixel detector, the p_{in} is often underestimated with respect to the initial electron momentum leading to the bias $\mathrm{E}_{\mathrm{TOT}} > \mathrm{p}_{\mathrm{in}}$. However, in such a case, an ECAL cluster is linked with one of the first three Bremsstrahlung track tangents and this information is stored in the ($EarlyBrem$) observable. Instead, when a Bremsstrahlung photon emissions occurs later in the tracker, the particle-flow clustering might not be able to disentangle the electron energy deposit from a late Bremsstrahlung emission leading to the bias $\mathrm{E}_e > \mathrm{p}_{\mathrm{out}}$. When it happens, the azimuthal position of the merged particle-flow cluster (electron and the late Bremsstrahlung contribution) will not be centered on the GSF track extrapolation in the ECAL surface but more towards one of the latest track tangents. If this happen this information is saved in the $LateBrem$ observable.

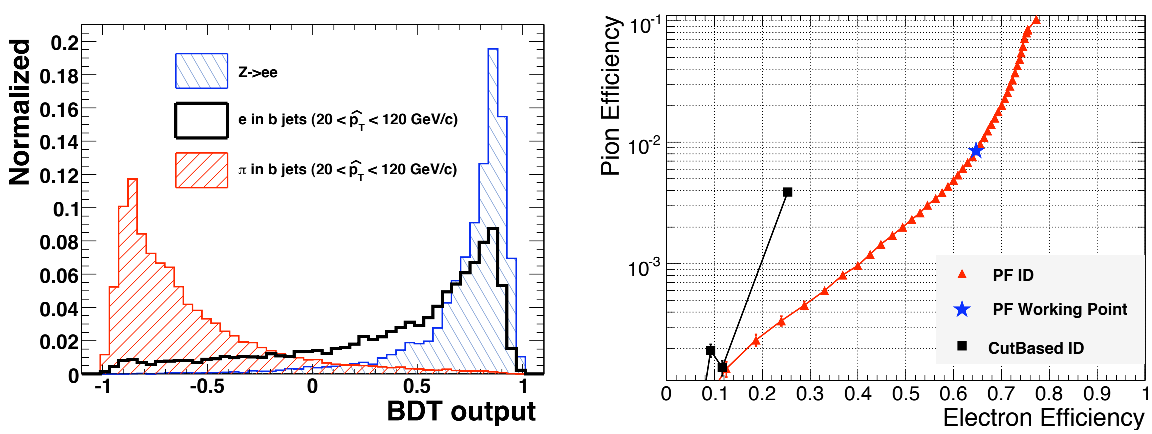
The Tracker and ECAL information is also used to match in pseudo-rapidity the ECAL cluster and the GSF track ($\eta_{\mathrm{GSF}} - \eta_{\mathrm{Cluster}}$). Moreover, the lateral shower shape of the ECAL cluster associated to the electron is expected to be narrow as is typical for genuine electromagnetic showers. Because the Bremsstrahlung photon contribution enlarges the shower in the azimuthal direction, the $\sigma_{\eta\eta}$ [1] variable has been chosen to measure the shower width only in the pseudo-rapidity direction. The HCAL information was used to compute the hadronic fraction, built from ratio of the HCAL energy (H) and the ECAL energy E of the clusters linked to the GSF track (H/(H+E)).

Finally, pure tracking observables, both for the GSF and KF track, were considered. Pure tracking observables are important especially when the particle-flow supercluster absorbs energy from the neutral particles of the jet, leading to a mismatch of the Tracker and ECAL observables or when an electron goes in one of the ECAL cracks. For radiating electrons, the GSF track momentum at the outermost state is smaller than the momentum at the innermost state. The so-called fbrem. variable, measured with the GSF track, has then a flatter behavior for electrons with respect to charged hadrons. Moreover the KF track tends to have a larger χ^2_{KF} and a smaller number of reconstructed tracker hits than the GSF one, because the Kalman Filter is not able to follow the change of curvature. The resolution $\sigma_{\mathrm{p_T}}/\mathrm{p_T}$ and the χ^2 of the GSF track are also included in the list of pure tracking observables. They do not show a good discrimination power, but they are used to quantify the reliability of the parameters measured by the GSF track. The final list of the observables for the electron identification are summarized below:

- $E_{TOT}/\mathrm{p}_{\mathrm{in}}$
- $E_e/\mathrm{p}_{\mathrm{out}}$
- $E_{\gamma(s)}/(\mathrm{p}_{\mathrm{in}} - \mathrm{p}_{\mathrm{out}})$
- EarlyBrem

- LateBrem
- $\mathrm{Log}(\sigma_{\eta\eta})$ (only for the ECAL cluster linked to the GSF track)
- $\eta_{\mathrm{GSF}} - \eta_{\mathrm{Cluster}}$ (only for the ECAL cluster linked to the GSF track)
- H/(H+E)
- $\mathrm{fbrem}_{\mathrm{GSF}} = (\mathrm{p_{in}} - \mathrm{p_{out}})/\mathrm{p_{in}}$
- χ^2_{KF} and χ^2_{GSF}
- # $\mathrm{hits_{KF}}$
- $\sigma_{\mathrm{p_T}}/\mathrm{p_T}$ GSF track

These observables are combined into a single discriminator using a multivariate Boosted Decision Tree (BDT) method. The training was done using a combined sample of isolated and non-isolated electrons and pions applying a selection which make the background and signal sample flat in $\ln(\mathrm{p_T})$ and η. The BDT output for isolated electrons ($\mathrm{Z} \to \mathrm{e^+e^-}$) and non-isolated electrons and pions (b jets $\hat{\mathrm{p_T}}[20 - 120]\ \mathrm{GeV}/c$) is shown in Fig. 20 for all the GSF tracks with $\mathrm{p_T} > 2\ \mathrm{GeV}/c$. A very good separation between electrons and pions is achieved when the electrons are isolated. The pion-electron separations remains good for electrons in jets. The efficiency for non-isolated electrons and pions for a sample of b jets with $\hat{\mathrm{p_T}}[20 - 120]\ \mathrm{GeV}/c$ is shown in Fig. 20 as a function of the BDT output. The results obtained with the developed multivariate method are also compared with the cut-based identification algorithm developed for isolated electrons.



(a) (b)

Figure 20: Output of the Boosted Decision Tree (BDT) (a). Electron vs pion efficiency as a function of the BDT output (b)

Because of the ability of particle-flow to reconstruct electrons within jets, these electrons are ideal for the use of b-jet tagging with soft leptons. Leptons in jets are very useful for b-jet tagging due to the large branching fraction of b-quarks to leptons. A framework for performing soft lepton tagging using muons has existed for quite some time within CMSSW. Thanks to particle-flow, soft lepton tagging with electrons is now also possible.

There are currently two simple soft lepton taggers utilizing electrons, which use two distinct variables as discriminators:

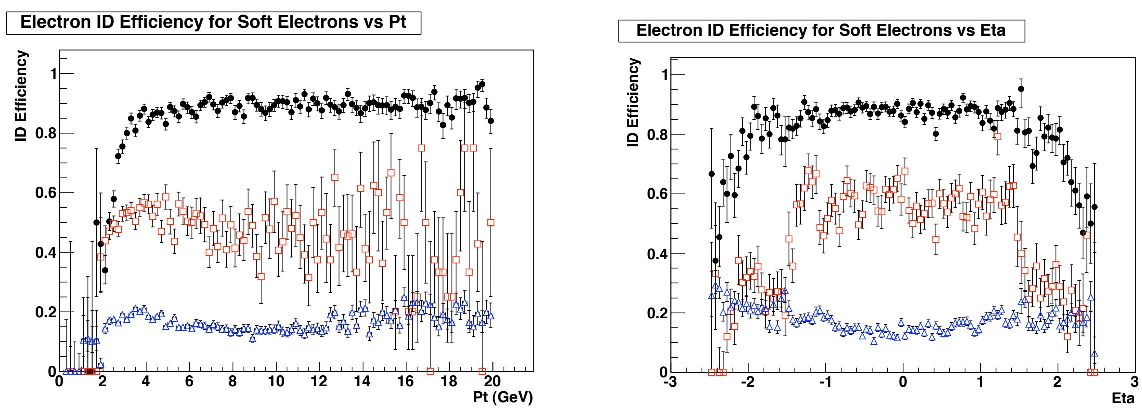
- the transverse component of the electron momentum with respect to the jet axis (known as $\mathrm{p_{T,Rel}}$),
- the signed impact parameter significance of the electron track with respect to the primary vertex (the sign of the variable comes from the dot product of the distance vector between the primary vertex and the electron point of closest approach and the jet axis vector).

This section will focus primarily on the $\mathrm{p_{T,Rel}}$ tagger, which is the one most likely to be used with early data.

Before the output electrons from particle-flow can be used in soft lepton tagging, some cut-based preselection is performed to remove fake electrons and electrons from conversions. Including these in the tagging will enrich the tagged sample with jets from light partons. The variables used to preselect electrons for tagging are

- $\Delta\mathrm{R}$ between GSF Track and ECAL Cluster,
- Supercluster Energy / Momentum at Vertex,
- Particle-flow MVA variable
- $\Delta\mathrm{R}$ between the first and last tracker hit,
- The radial distance of the first tracker hit from the beam pipe,
- The distance along the z-direction between the first tracker hit and the detector center.

Of the variables listed above, the particle-flow mva variable is the most discriminating against fake electrons, while the radial and z distance of the first tracker hit are more discriminating against conversions. The mva cuts for preselection are mva > -0.1 for electrons in the barrel region, and mva > -0.24 in the forward. The electron-id efficiency (defined as the number of reconstructed electrons surviving the cuts divided by the total number of reconstructed electrons) is shown in Fig. 21 as a function of $\mathrm{p_T}$ and η for real soft electrons (the black dots), conversion electrons (the red squares), and fake electrons (the blue triangles) in b-jets.



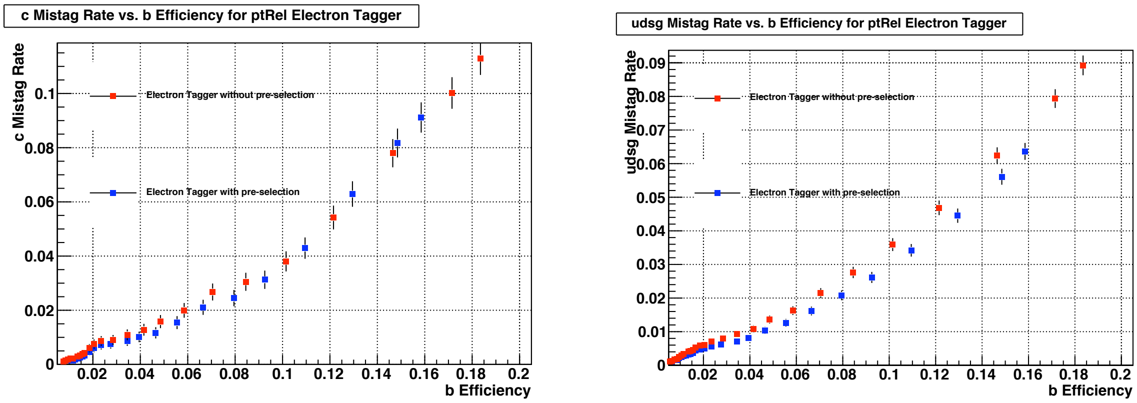
(a) (b)

Figure 21: The electron-id efficiency for the tagging electron preselection cuts as a function of $\mathrm{p_T}$ and η for real soft electrons (black dots), conversion electrons (red squares), and fake electrons (blue triangles), in b-jets.

The effect of the preselection of soft electrons on the $\mathrm{p_{T,Rel}}$ tagger is shown in Fig. 22, with both the c-jet mistag rate (shown in (a)) and the udsg-jet mistag rate (shown in (b)). The tagging efficiency is defined as the number of tagged b-jets divided by the total number of b-jets. Thus, the maximum expected tagging efficiency is $\sim 19\%$, when one considers the b $\rightarrow$ e branching ratio plus any cascade decays (i.e. b$\rightarrow$ c $\rightarrow$ e). The mistag rate is defined the same as the tagging efficiency, but is applied only to c and light jets. As can be seen in Fig. 22, the overall tagging efficiency/mistag rate for all jet flavours decreases with the inclusion of soft electron preselection, and the udsg-jet mistag rate slightly improves with the inclusion of the preselection.

8 Low $\mathrm{p_T}$ electrons

Electrons with low $\mathrm{p_T}^e \lesssim 10\ \mathrm{GeV}/c$ are better reconstructed and identified using the technique developed for non-isolated electrons (Sec. 7) in the context of the particle-flow event reconstruction. Indeed for low $\mathrm{p_T}^e$, the ϕ window used for the standard superclusters starts to be too small and some electrons which radiate lead to electron and photon clusters more separated than 0.3 rad in the magnetic field [2]. The particle-flow supercluster (Sec. 7) does not suffer by this limitation because the ϕ window for the ECAL cluster recovery varies dynamically with the curvature of the GSF track. This region is always delimited by the extrapolated positions towards the ECAL surface of the GSF track and the first track tangent used for the Bremsstrahlung recovery. The performances of the ECAL clusters recovery with this method has been illustrated in [2] using a sample of electrons with a flat

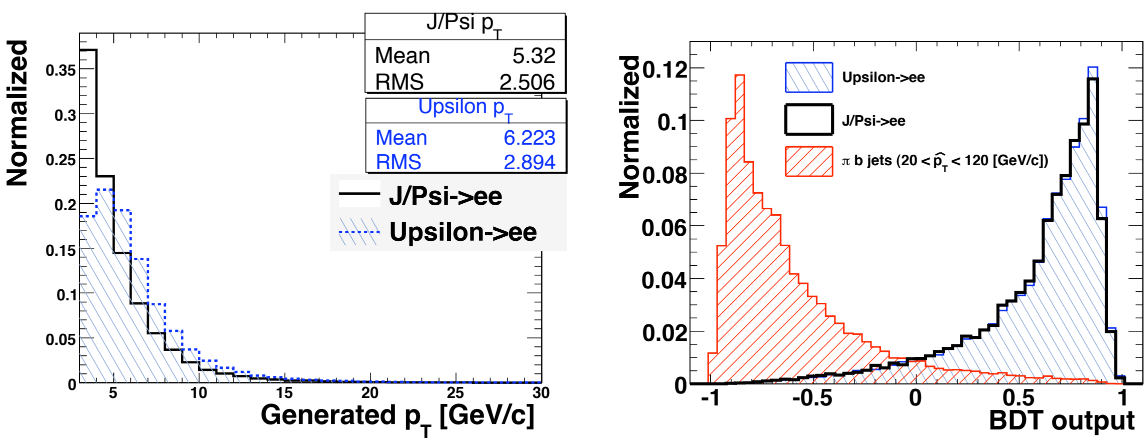


(a) (b)

Figure 22: The effect of soft electron preselection on the $\mathrm{p_{T,Rel}}$ tagger. The c-jet mistag rate vs. b-jet efficiency is shown in (a), while the udsg-jet mistag rate vs. b-efficiency is shown in (b).

$\mathrm{p_T}$ distribution between $2-10$ GeV$/c$. Around 96% of the maximum recoverable energy is contained in the particle-flow supercluster.

Thanks to this technique the energy matching between the track momentum and the ECAL still holds also for low $\mathrm{p_T}^e$ electrons. Moreover the multivariate Boosted Decision Tree (BDT) method, described in Sec. 7, has been trained applying a selection which makes the background and signal sample flat in $\ln(\mathrm{p_T})$ and η in order to take into account the correlations of the input variables with those quantities. The generated $\mathrm{p_T}$ spectrum and the BDT output for a sample of $\mathrm{J}/\Psi \rightarrow$ ee and $\Upsilon \rightarrow$ ee is shown in Fig. 23. The final identification efficiency for the chosen



(a) (b)

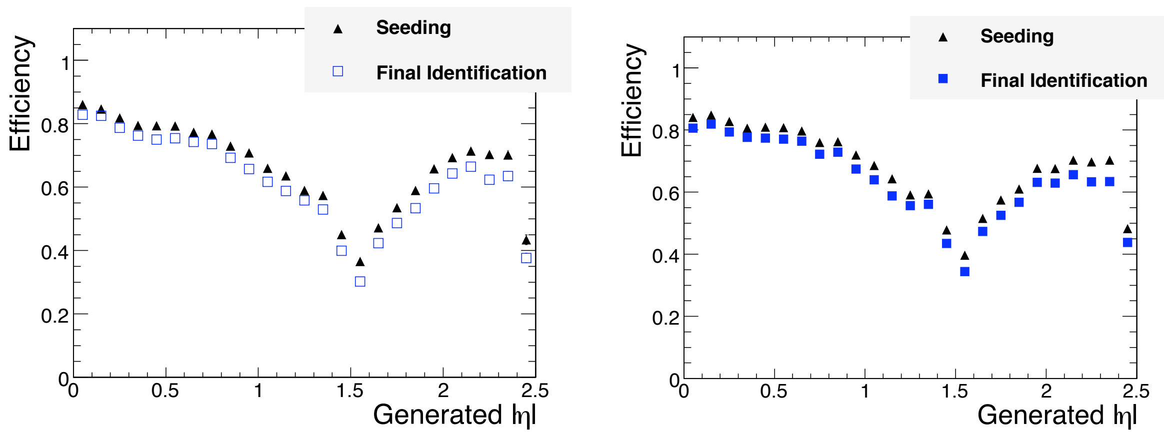
Figure 23: $\mathrm{p_T}$ spectrum (a) and output of the Boosted Decision Tree (BDT) (b) for a $\Upsilon \rightarrow$ ee and a $\mathrm{J}/\Psi \rightarrow$ ee sample. Electrons with $\mathrm{p_T} > 3$ GeV$/c$ are MC preselected. For comparison the BDT output for pions in a b jets sample with $20 < \hat{\mathrm{p_T}} < 120$ GeV$/c$ is also presented.

particle-flow working point BDT > -0.1 is compared with the electron seeding efficiency in Fig. 24. The final identification electron reconstruction efficiency is 65% (94.0%) and 64% (93.6%) for the $\Upsilon \rightarrow$ ee and $\mathrm{J}/\Psi \rightarrow$ ee sample. The values in brackets indicate the efficiencies with respect to the electron seeding. A detailed study for the J/Ψ, Υ background rejection goes beyond the scope of this note, however as it is shown in Fig. 20 the pion efficiency for the chosen BDT > -0.1 selection is less than 1% for a b jets sample with $20 < \hat{\mathrm{p_T}} < 120$ GeV$/c$.

9 Electron Selection at High Energy

The requirements for a good high energy electron selection are:

- use simple variables that are well modeled in the Monte Carlo simulation



(a) (b)

Figure 24: Electron seeding efficiency and Final Identification efficiency for a sample of $J/\Psi \rightarrow ee$ (a) and $\Upsilon \rightarrow ee$ (b). Electrons with $p_T > 3$ GeV/c are MC preselected.

- have little energy dependence and any evolution with energy to be well understood
- be highly efficient

These requirements differ from the requirements of the electron identification at normal energies, where normal energy is defined as the Z pole, where simplicity and energy dependence are not as important. High energy electron identification be simple, robust and have a well understood evolution with energy as there is no known source of high energy electrons with sufficient purity to measure the electron identification efficiency. This means, unlike for the case of normal energy electrons which have the Z pole, the selection efficiencies must be estimated with Monte Carlo simulation. Therefore the electron ID approach at high energies must be more conservative than that of the normal energy electron ID and the variables chosen must be well modeled in the Monte Carlo simulation with a well understood energy dependence. Additionally at high energy electrons are rare and have a smaller jet backgrounds so a good high energy electron selection is also highly efficient at the cost of increased fake rate to maximise the sensitivity to new physics.

Currently the benchmark process used for tuning the high energy electron identification is Z'→ee although it is used by all exotica analyses using high energy electrons. The high energy electron electron selection is an evolution of the HEEP selection first documented in [12]. The latest selection can be found on the following twiki [13] where there are also instructions on how to run it.

9.1 Samples and Method

To evaluate the electron and jet efficiencies, the samples in table 9.1 were used. For electron samples the efficiency is defined as the fraction of reconstructed electrons matched to a true prompt electron passing the selection. For jet samples the efficiency is defined as the fraction of reconstructed electrons passing the selection. Jet samples are weighted to the same luminosity.

9.2 Selection

The selection cuts are detailed in Table 3. All high energy electrons are required to be seeded by the ecal not the tracker. This is because the tracker momentum resolution is poor at high energies and while tracker driven seeding is more efficient than ecal driven seeding for low E_T or non-isolated electrons, it offers no gain for high E_T isolated electrons and would introduce an additional complication. The differences between normal energy and high energy selection variables can be found below. It should be stressed that in cases, these variables work at normal energies as well but either will have some special high energy behaviour irrelevant to normal energies or are just a simpler more conservative variable. The first difference is that all isolation variables use a cone of 0.3 rather than 0.4. A study in [12] found little difference in performance between the cone sizes so the smaller cone size was chosen to minimise the area of the detector used.

Channel	Dataset
di-jet	/QCD_Pt30/Summer09-MC_31X_V3-v1/GEN-SIM-RECO
di-jet	/QCD_Pt80/Summer09-MC_31X_V3-v1/GEN-SIM-RECO
di-jet	/QCD_Pt170/Summer09-MC_31X_V3-v1/GEN-SIM-RECO
di-jet	/QCD_Pt300/Summer09-MC_31X_V3-v1/GEN-SIM-RECO
di-jet	/QCD_Pt470/Summer09-MC_31X_V3-v1/GEN-SIM-RECO
di-jet	/QCD_Pt800/Summer09-MC_31X_V3-v1/GEN-SIM-RECO
$\mathrm{Z} \rightarrow \mathrm{e}^{+}\mathrm{e}^{-}$	/Zee/Summer09-MC_31X_V3-v1/GEN-SIM-RECO
$\mathrm{Z} \rightarrow \mathrm{e}^{+}\mathrm{e}^{-} M_Z > 120$ GeV/c^2	private 3_1_2 production
$\mathrm{Z} \rightarrow \mathrm{e}^{+}\mathrm{e}^{-} M_Z > 200$ GeV/c^2	private 3_1_2 production
$\mathrm{Z} \rightarrow \mathrm{e}^{+}\mathrm{e}^{-} M_Z > 500$ GeV/c^2	private 3_1_2 production
$\mathrm{Z} \rightarrow \mathrm{e}^{+}\mathrm{e}^{-} M_Z > 800$ GeV/c^2	private 3_1_2 production
RS Graviton $M_G = 500$ GeV/c^2	private 3_1_2 production
RS Graviton $M_G = 750$ GeV/c^2	private 3_1_2 production
RS Graviton $M_G = 1000$ GeV/c^2	private 3_1_2 production
RS Graviton $M_G = 1250$ GeV/c^2	private 3_1_2 production

Table 2: List of the samples used to evaluate high energy electron selection performance. The official samples have approximately 2 million events each while the private $\mathrm{Z} \rightarrow \mathrm{e}^{+}\mathrm{e}^{-}$ and RS Graviton samples have approximately 40K and 20K each.

Variable	Barrel	Endcap
$\mathrm{E_T}$	> 25 GeV	> 25 GeV
η^{SC}	$\lvert\eta^{\mathrm{SC}}\rvert \leq 1.442$	$1.560 \leq \lvert\eta^{\mathrm{SC}}\rvert \leq 2.5$
isEcalDriven	true	true
H/E	< 0.05	< 0.05
$\mathrm{E}^{2\times5}/\mathrm{E}^{5\times5}$	$< 0.94 \| \|\mathrm{E}^{1\times5}/\mathrm{E}^{5\times5} > 0.83$	n/a
$\sigma_{\mathrm{i}\eta\mathrm{i}\eta}$	n/a	< 0.03
$\lvert\Delta\eta_{\mathrm{in}}\rvert$	< 0.005	< 0.007
$\lvert\Delta\phi_{\mathrm{in}}\rvert$	< 0.09	< 0.09
isol EM + Had Depth 1	$< 2 + 0.03 \times \mathrm{E_T}$	< 2.5 GeV for $\mathrm{E_T} < 50$ GeV else $< 2.5 + 0.03 \times (\mathrm{E_T} - 50)$ GeV
isol Had Depth 2	-	< 0.5 GeV
isol Trk p_T	< 7.5 GeV$/c$	< 15 GeV$/c$

Table 3: The high energy electron selection cuts

The third difference is that in the high energy electron selection exploits the depth segmentation of the HCAL endcap. As a real electron will deposit almost zero energy in the second depth of the HCAL, this gives excellent jet discrimination at high energies. It is not as significant for low energies as low energy jets will also deposit little energy in the second depth of the HCAL.

Finally the selection replaces the shower-shape variable $\sigma_{\mathrm{i}\eta\mathrm{i}\eta}$ with cuts $\mathrm{E}^{2\times5}/\mathrm{E}^{5\times5}$ and $\mathrm{E}^{1\times5}/\mathrm{E}^{5\times5}$ in the barrel. These are defined as the energy in a energy in a X in η and 5 in ϕ block of crystals containing the seed crystal. In the case of the E^{2x5}, of the two possible blocks, the block with the highest energy is chosen. In the $\eta - \phi$ geometry of the barrel these variables are logically identical to $\sigma_{i\eta i\eta}$ but is simpler and better performing.

It is also worth noting that the the tracker momentum resolution is poor at high p_T as the track has little curvature. This means that the high energy electron selection must avoid tracker based momentum measurements and for this reason variables such as f_{brem} and E/p_{in} are not included in the selection.

9.3 Evolution of Isolation With Energy

The calorimeter isolation is ideally defined as the transverse energy in cone centered on the electron's supercluster minus the energy deposited by the electron, known as the electrons footprint. In practice, the electron footprint removal does not remove all the energy deposited by the electron and therefore there is a small dependence on electrons $\mathrm{E_T}$, which is only significant at high energies. Fig. 25 shows the efficiency of the calorimeter depth 1 isolation cut vs $\mathrm{E_T}$. With no scaling, there is a large drop in efficiency at high $\mathrm{E_T}$. A scaling term of 2-3% is required to stablise the efficiency vs $\mathrm{E_T}$. The tracker isolation does not suffer from this problem and is independent

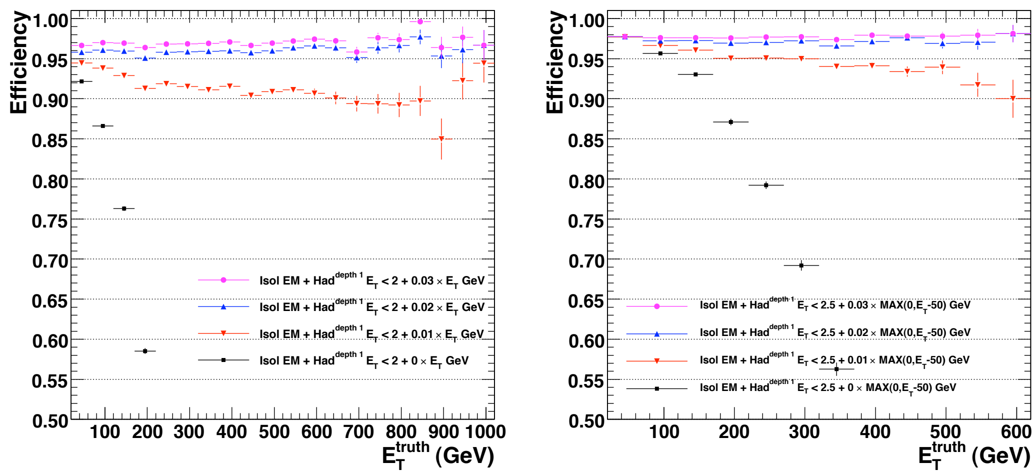


Figure 25: The efficiency of the isol EM+Had Depth 1 cut for various scalings with $\mathrm{E_T}$ for barrel (left) and endcap (right) electrons.

of electron $\mathrm{E_T}$.

9.4 Performance

The efficiency of the identification cuts, the isolation cuts and the total selection cuts vs $\mathrm{E_T}$ are shown in figure 26 for both electrons and jets, divided into barrel and endcap electrons. As can be seen from the figure the efficiency is flat for $\mathrm{E_T} > 75$ GeV. The 3% efficiency loss at low energy in the endcap is mainly due to the $\Delta\eta_{\mathrm{in}}$ cut but also some contribution from the $\sigma_{\mathrm{i}\eta\mathrm{i}\eta}$ and $\mathrm{H/E}$ cuts. The efficiency vs η is shown in figure 27 for low ($\mathrm{E_T} <$ 150 GeV) and high ($\mathrm{E_T} > 150$ GeV) energy electrons. There is a strong η dependence in the ID cuts due to $\Delta\eta_{\mathrm{in}}$. Intermodule gaps in the barrel also cause the $\mathrm{H/E}$ cut to have an η dependence. The isolation efficiency gradually increases by about 2% with increasing η in the endcap. The effects are reduced at high energy, except for the $\mathrm{H/E}$ cut which becomes more inefficient by the intermodule gaps in the barrel. The efficiency vs ϕ is shown in figure 28 for low ($\mathrm{E_T} < 150$ GeV) and high ($\mathrm{E_T} > 150$ GeV) energy electrons. There is a 2% variation in efficiency in ϕ in the endcap due to the $\mathrm{H/E}$ cuts around the region where the crystal gaps are no longer off-pointing.

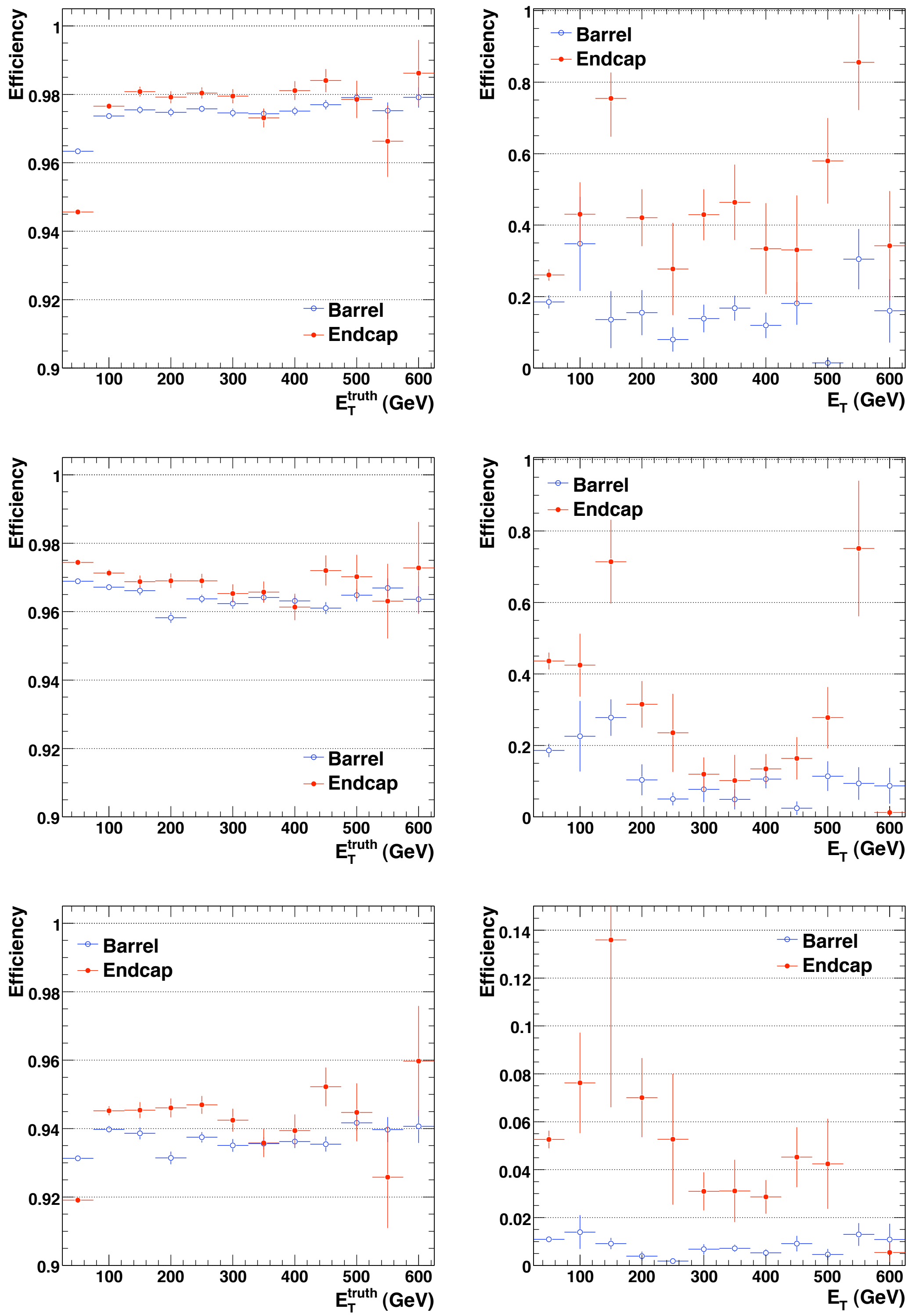


Figure 26: The efficiency of the high energy electron selection vs $\mathrm{E_T}$ for electrons (left) and jets (right) w.r.t to a pre-selected electron passing $\mathrm{E_T}$ and η cuts. The top plot is for the efficiency of the ID cuts after isolation, the middle plot is the efficiency of the isolation after ID and the bottom plot is the total efficiency. The variables $\mathrm{H/E}$ and and $\Delta\eta_{\mathrm{in}}$ drive the decreased efficiency in the first bin.

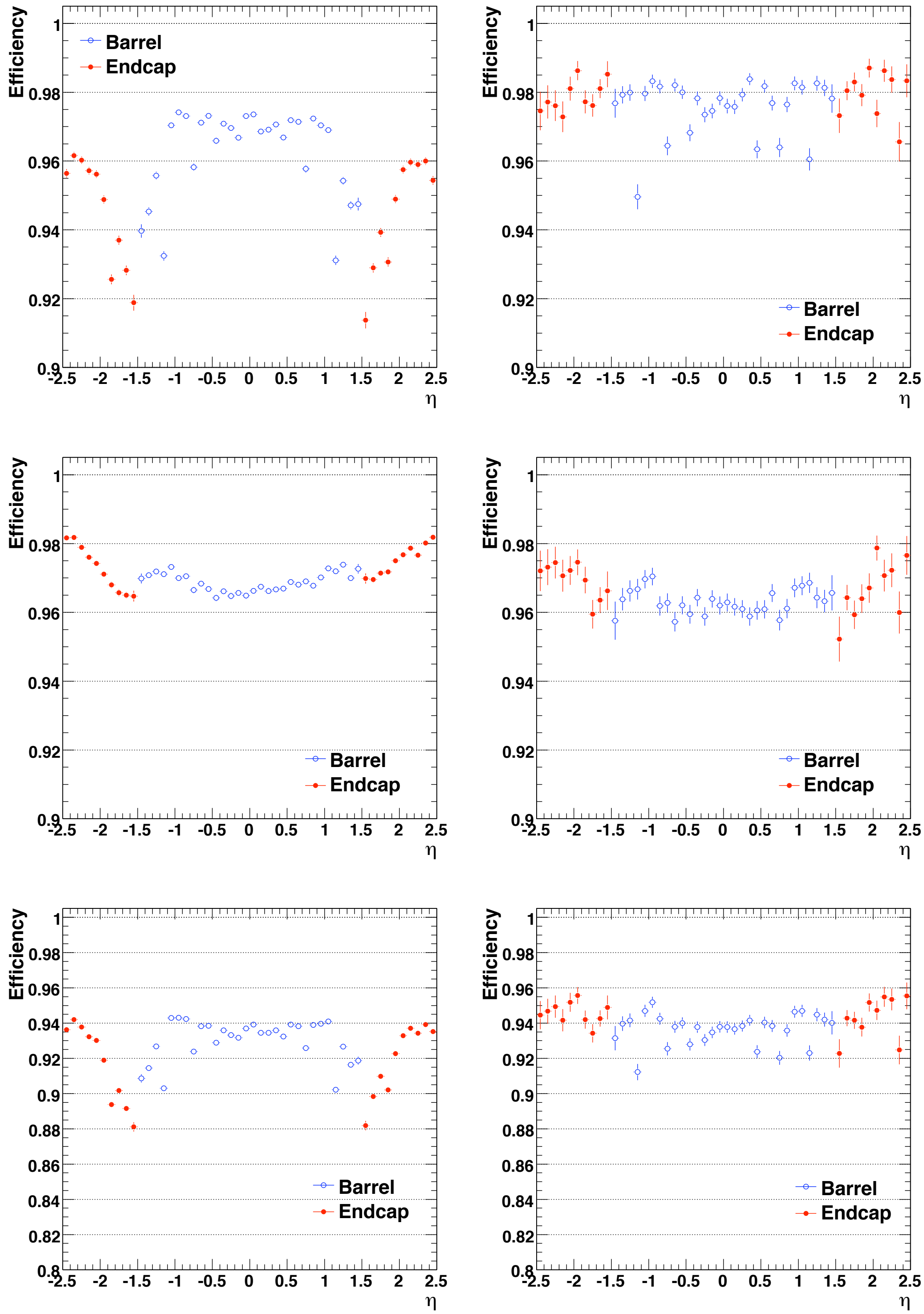


Figure 27: The efficiency of the high energy electron selection vs η for $\mathrm{E_T} < 150$ GeV (left) and $\mathrm{E_T} > 150$ GeV (right) w.r.t to a pre-selected electron passing $\mathrm{E_T}$ and η cuts. The top plot is for the efficiency of the ID cuts after isolation, the middle plot is the efficiency of the isolation after ID and the bottom plot is the total efficiency. The $\Delta\eta_{\mathrm{in}}$ variable drives η dependence at low energies.

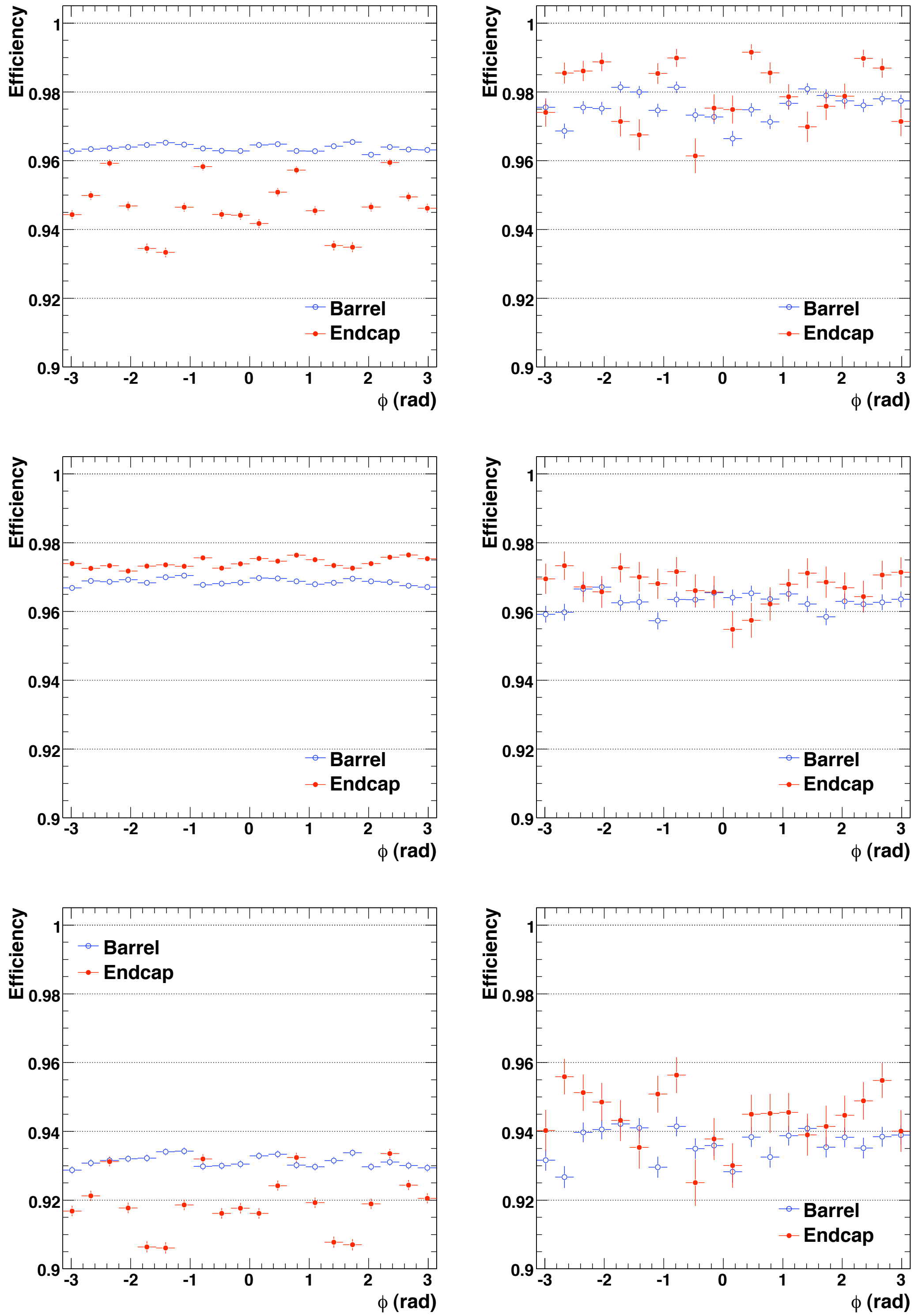


Figure 28: The efficiency of the high energy electron selection vs ϕ for $\mathrm{E_T} < 150$ GeV (left) and $\mathrm{E_T} > 150$ GeV (right) w.r.t to a pre-selected electron passing $\mathrm{E_T}$ and η cuts. The top plot is for the efficiency of the ID cuts after isolation, the middle plot is the efficiency of the isolation after ID and the bottom plot is the total efficiency. The $\mathrm{H/E}$ variable drives ϕ dependence in the endcap.

10 Rejection of Electrons from Conversions.

Due to the non-negligible material budget of the CMS tracking system, large multiple scattering, bremsstrahlung and high photon conversion rates are all prevalent. Electrons from photon conversions constitute approximately 15÷35% [3)] of electrons in QCD events, depending on the cuts applied. Electrons from photon conversions are therefore a non-negligible background to prompt electrons from hadron collisions and must be rejected efficiently. To that end, we have developed three cuts designed to reject such electrons:

- Impact Parameter: electrons from photon conversions will have, on average, a greater transverse distance from the beamspot (impact parameter or d0) than electrons from prompt sources. This is due to the fact that a photon conversion occurs in material, either in the beam pipe or in the tracker layers. Extrapolating the electron candidates track to the nominal beam position in the xy plane results in a large impact parameter for electron candidates from photon conversions while a prompt electron coming from the interaction region will have a small impact parameter. Requiring a small impact parameter ($< 200\ \mu$m) rejects a large fraction of electrons from photon conversions while having a negligable effect on prompt electrons.
- Hit Pattern: photon conversions occur later inside the tracker volume and not at the primary vertex. Therefore, the first valid hit of a resulting electron track may not necessarily be located in the innermost tracker layer. We call a hit valid if it is used in the final out of the track. Extrapolating the track of an electron from a photon conversion back to the beam-line, one could cross active detector layers which do not have hits compatible with the track (in other words, a missing hit). For prompt electrons, whose trajectories start from the beam-line, we do not expect any missing hits in the crossed inner tracker layers. We can therefore use this expectation of no missing hits at inner radii to reject electrons from photon conversions. Determining whether a track has missing hits is possible via its associatied Hit Pattern object. We find that requiring the number of expected layers with a missing hit be $\leq$1 efficiently rejects a large fraction of electrons from photon conversions.
- Search for the Conversion Partner Track: the tracks of the resulting electrons from a conversion decay are parallel to each other at the decay point, and remain so in the rz plane. This is a unique feature that is the basis of the algorithm we use. To exploit this geometry, all Combinatorial Track Fitter (CTF) tracks within a cone of $\Delta\mathrm{R} < 0.3$ around the electron GSF track and with charge opposite that of the GSF track, are pre-selected. For each of these tracks, the following two quantities are defined:
  - $\Delta\cot(\Theta) = \cot(\Theta_{\mathrm{CTF\ Track}}) - \cot(\Theta_{\mathrm{GSF\ Track}})$
  - The Dist is defined as the two-dimensional distance (xy plane) between the two tracks when the CTF track in question and the electron GSF track would be parallel when extrapolated. This distance is calculated analytically by a simple intersection of he- lices method using the track parameters of the two tracks as input. Figure 29 shows the definition of dist, as well as the sign convention used.

  Requiring that the $|\Delta\cot(\Theta)| < 0.02$ and $|\mathrm{Dist}| < 0.02$ cm efficiently rejects a significant portion of the remaining electrons from photon conversions.

  The cuts above are described in more detail in analysis note [14].

11 Conclusions

The electron identification for CMS has been presented. Simple cut based selections and a complete set of variables to distinguish between real electrons and background electrons are described while when the detector will be better understood possible improvements using multivariate techniques can be introduced. The performance of electron identification was estimated using simulated data by measuring the efficiency for the selection of prompt electrons and estimating the fake rate for background events, different levels of tightness are shown. The way of obtaining identification variables shapes using data both for the signal and for the background has been established. Methods to tune the cuts for electron selection with real data have been investigated. The method of identified electrons in jets using a technique developed in the context of the particle-flow event reconstruction has been studied. the same technique as been applied to electrons with very low $\mathrm{p_T}$. The requirements for a good high energy electron selection are described and the selection method is established.

[3)] The rate of electron candidates from photon conversions in QCD was studied in CMSSW_1_6_7, where large statistics QCD samples were available. As many things have changed since CMSSW_1_6_7, the numbers quoted above are meant only as a rough guide to give the reader a scale of the problem.

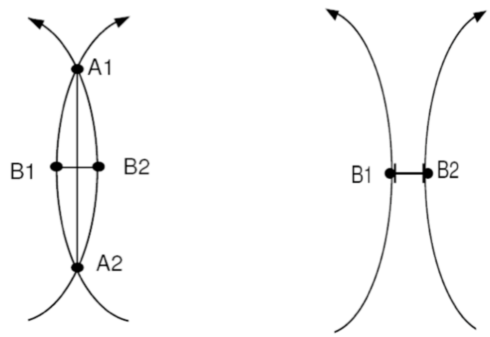


Figure 29: The Dist quantity is the two dimensional distance between points B1 and B2 in the xy plane as seen above. At these points, the two tracks from the photon conversion are parallel. The Dist is defined to be negative when the two tracks overlap, and is positive notherwise.

References

[1] S. Baffioni *et al.*, “Electron Reconstruction in CMS”, The European Physical Journal C, **49** (2007) 1099-1116

[2] F. Beaudette *et al.*, “Electron Reconstruction in CMS” CMS AN-2009/164.

[3] J. Branson *et al.*, “A cut based method for electron identification in CMS” CMS AN-2008/082.

[4] E. Di Marco *et al.*, “Electron Identification in the CMS experiment based on a likelihood algorithm” CMS AN-2008/110.

[5] S. Baffioni *et al.*, “Electron Selection and Identification in CMS”, CMS AN-2005/065.

[6] G. Daskalakis *et al.*, “Data driven selection cut tuning for electrons” CMS AN-2009/108.

[7] The CMS Collaboration, “W and Z production in the electron channel at sqrts = 10 TeV” CMS PAS EWK-2009-004.

[8] G. P. Salam and G. Soyez, “A practical Seedless Infrared-Safe Cone jet algorithm”, JHEP 05 (2007) 086, [arXiv:0704.0292].

[9] M. Pivk and F. R. Le Diberder, Nucl. Instrum. Meth. A **555** (2005) 356 [arXiv:physics/0402083].

[10] The CMS Collaboration, “Particle-Flow Event reconstruction in CMS and Performance for Jets, Taus, and E_T^{miss} ”, CMS PAS PFT-09/001.

[11] The CMS Collaboration, “Electron reconstruction within the Particle Flow Algorithm”, CMS PAS PFT-09/006 (in preparation).

[12] O. Charaf *et al.*, “Electron ID at High Energies”, CMS AN-2008/045.

[13] Electron ID and Isolation for high energy electron: *https://twiki.cern.ch/twiki/bin/view/CMS/HEEPELectronID*

[14] W. Andrews *et al.*, “Study of Photon Conversion Rejection”, CMS AN-2009/159.

A Cut values

In this appendix some are reported an example of cut values for each selection described in the previous chapters.

Efficiency	95.00%	93.00%	90.00%	87.00%	85.00%	83.00%	80.00%	75.00%
Barrel								
Tk Iso.	7.63E+00	7.63E+00	6.07E+00	6.07E+00	4.93E+00	3.42E+00	3.42E+00	2.81E+00
ECAL Iso.	5.48E+00	4.76E+00	4.76E+00	4.76E+00	4.76E+00	4.76E+00	3.87E+00	3.87E+00
HCAL Iso.	5.70E+00	5.70E+00	5.70E+00	5.70E+00	5.70E+00	5.70E+00	5.70E+00	5.70E+00
$\sigma_{\mathrm{i}\eta\mathrm{i}\eta}$	1.03E-02	1.03E-02	1.03E-02	1.03E-02	1.03E-02	1.03E-02	1.03E-02	9.95E-03
$\Delta\phi_{\mathrm{in}}$	8.96E-02	5.97E-02	3.36E-02	2.63E-02	2.63E-02	2.63E-02	2.01E-02	2.01E-02
$\Delta\eta_{\mathrm{in}}$	7.13E-03	7.13E-03	7.13E-03	7.13E-03	7.13E-03	7.13E-03	7.13E-03	7.13E-03
H/E	5.42E-02	5.42E-02	5.42E-02	3.12E-02	3.12E-02	3.12E-02	3.12E-02	3.12E-02
Endcap								
Tk Iso.	7.28E+00	6.18E+00	6.18E+00	5.04E+00	4.22E+00	2.87E+00	1.29E+00	1.04E+00
ECAL Iso.	3.23E+00	2.79E+00	2.79E+00	2.40E+00	2.40E+00	2.40E+00	2.40E+00	2.40E+00
HCAL Iso.	2.13E+00	1.68E+00	1.68E+00	1.68E+00	1.19E+00	1.19E+00	1.19E+00	7.85E-01
$\sigma_{\mathrm{i}\eta\mathrm{i}\eta}$	3.04E-02	2.93E-02	2.93E-02	2.93E-02	2.93E-02	2.93E-02	2.93E-02	2.93E-02
$\Delta\phi_{\mathrm{in}}$	7.00E-01	7.00E-01	2.71E-02	2.27E-02	2.27E-02	2.27E-02	2.27E-02	2.27E-02
$\Delta\eta_{\mathrm{in}}$	8.55E-03	8.55E-03	8.55E-03	8.55E-03	6.10E-03	6.10E-03	6.10E-03	4.45E-03
H/E	3.38E-02	3.38E-02	3.38E-02	2.24E-02	2.24E-02	2.24E-02	2.24E-02	2.09E-03

Table 4: Example of cut values optimized by the iterative method. Each column reports a different set correspondind to different level of tightness. This cuts have been set for $\mathrm{E_T} > 30$ GeV. In tuning this cuts the $\Delta\phi_{\mathrm{in}}$ cut has been restricted to be greater than 0.02. This was done because despite this cut is very efficient, too tight values can lead to huge efficiency dependence.

Efficiency	97.00%	95.00%	93.00%	90.00%	87.00%	85.00%	80.00%
Barrel							
Tk Iso.	8.13E+00	6.34E+00	5.66E+00	4.94E+00	3.83E+00	3.39E+00	2.63E+00
ECAL Iso.	(n.a)	(n.a)	5.53E+00	4.77E+00	4.77E+00	4.77E+00	4.09E+00
HCAL Iso.	4.65E+00	4.65E+00	4.65E+00	4.65E+00	4.65E+00	4.65E+00	4.65E+00
$\sigma_{\mathrm{i}\eta\mathrm{i}\eta}$	1.08E-02	1.05E-02	1.02E-02	1.02E-02	1.02E-02	1.02E-02	1.02E-02
$\Delta\phi_{\mathrm{in}}$	(n.a)	(n.a.)	9.48E-02	6.06E-02	4.32E-02	4.32E-02	3.20E-02
$\Delta\eta_{\mathrm{in}}$	6.43E-03	6.43E-03	6.43E-03	6.43E-03	6.43E-03	6.43E-03	6.43E-03
H/E	5.00E-01	6.26E-02	6.26E-02	6.26E-02	6.26E-02	3.41E-02	3.41E-02
Endcap							
Tk Iso.	9.82E+00	8.33E+00	6.91E+00	5.91E+00	4.93E+00	4.21E+00	3.46E+00
ECAL Iso.	(n.a)	3.43E+00	2.91E+00	2.91E+00	2.91E+00	2.91E+00	2.34E+00
HCAL Iso.	2.83E+00	2.11E+00	2.11E+00	2.11E+00	1.56E+00	1.56E+00	1.56E+00
$\sigma_{\mathrm{i}\eta\mathrm{i}\eta}$	3.16E-02	3.16E-02	3.05E-02	2.96E-02	2.96E-02	2.96E-02	2.96E-02
$\Delta\phi_{\mathrm{in}}$	(n.a)	(n.a.)	(n.a.)	3.91E-02	2.64E-02	2.24E-02	1.65E-02
$\Delta\eta_{\mathrm{in}}$	1.05E-02	8.57E-03	8.57E-03	8.57E-03	8.57E-03	8.57E-03	6.58E-03
H/E	4.34E-02	4.34E-02	3.24E-02	3.24E-02	3.24E-02	3.24E-02	2.12E-02

Table 5: Example of cut values optimized by the iterative method. Each column reports a different set correspondind to different level of tightness. This cuts have been set for $\mathrm{E_T} > 20$ GeV.

	Barrel			Endcap		
Variable	**bremming**	**low-brem**	**bad-track**	**bremming**	**low-brem**	**bad-track**
$\mathrm{E_T} > 30$ GeV						
$\Delta\eta_{in}$	9.58E-03	4.06E-03	1.22E-02	1.37E-02	8.37E-03	1.27E-02
$\Delta\phi_{in}$	3.72E-02	1.14E-01	1.18E-01	4.88E-02	1.17E-01	1.19E-01
E_{seed}/p_{in}	8.78E-01	8.02E-01	8.14E-01	9.42E-01	7.35E-01	7.74E-01
H/E	8.87E-02	9.34E-02	9.49E-02	9.86E-02	4.31E-02	8.78E-02
$\sigma_{\mathrm{i}\eta\mathrm{i}\eta}$	1.72E-02	1.15E-02	1.43E-02	3.44E-02	2.95E-02	3.04E-02
Tk Iso.	2.43E+01	8.45E+00	1.44E+01	2.78E+01	6.02E+00	1.05E+01
ECAL Iso.	3.34E+01	2.81E+01	7.32E+00	2.74E+01	7.33E+00	2.17E+01
HCAL Iso.	1.35E+01	9.93E+00	7.56E+00	1.48E+01	8.10E+00	1.08E+01
Impact Par.	2.46E-02	7.60E-02	9.66E-02	8.85E-02	4.41E-01	2.05E-01
Missing Hits	5.50E+00	1.50E+00	5.50E+00	2.50E+00	2.50E+00	2.50E+00
$20 < \mathrm{E_T} < 30$ GeV						
$\Delta\eta_{in}$	1.10E-02	3.36E-03	9.77E-03	1.50E-02	6.75E-03	1.09E-02
$\Delta\phi_{in}$	6.06E-02	5.48E-02	1.17E-01	7.00E-02	3.55E-02	1.17E-01
E_{seed}/p_{in}	8.29E-01	9.09E-01	8.29E-01	8.13E-01	8.60E-01	8.97E-01
H/E	9.70E-02	5.09E-02	9.80E-02	9.91E-02	3.21E-02	9.28E-02
$\sigma_{\mathrm{i}\eta\mathrm{i}\eta}$	1.45E-02	1.08E-02	1.28E-02	3.47E-02	3.07E-02	3.16E-02
Tk Iso.	1.41E+01	1.02E+01	1.45E+01	1.91E+01	6.10E+00	1.41E+01
ECAL Iso.	9.38E+01	1.02E+02	1.21E+01	2.60E+01	8.91E+00	1.00E+01
HCAL Iso.	4.27E+01	2.01E+01	9.11E+00	1.04E+01	6.89E+00	5.59E+00
Impact Par.	2.92E-02	2.93E-02	6.19E-02	2.51E-02	1.59E-01	8.15E-02
Missing Hits	3.50E+00	5.50E+00	5.00E-01	1.50E+00	2.50E+00	5.00E-01
$\mathrm{E_T} < 20$ GeV						
$\Delta\eta_{in}$	1.31E-02	3.29E-03	9.61E-03	1.48E-02	6.18E-03	1.32E-02
$\Delta\phi_{in}$	8.82E-02	3.53E-02	1.15E-01	7.98E-02	2.35E-02	4.89E-02
E_{seed}/p_{in}	8.17E-01	8.03E-01	8.02E-01	8.48E-01	7.96E-01	8.04E-01
H/E	6.92E-02	8.88E-02	9.77E-02	7.55E-02	8.23E-03	8.30E-04
$\sigma_{\mathrm{i}\eta\mathrm{i}\eta}$	1.76E-02	1.11E-02	1.34E-02	3.50E-02	3.09E-02	3.48E-02
Tk Iso.	8.92E+00	9.48E+00	8.03E+00	8.49E+00	7.26E+00	1.23E+01
ECAL Iso.	1.72E+01	1.88E+01	1.62E+01	1.41E+01	1.13E+01	1.14E+01
HCAL Iso.	8.89E+00	5.42E+01	1.56E+01	4.86E+00	1.05E+01	3.36E+00
Impact Par.	7.29E+00	1.20E-02	5.76E+00	6.89E+00	1.78E+00	5.89E+00
Missing Hits	5.00E-01	1.50E+00	5.00E-01	5.00E-01	5.00E-01	5.00E-01

Table 6: Cut values for the Category based loose selection. For each variable six cuts are set: bremming, low-brem and bad-track electrons, barrel and endcap separately.

	Barrel			Endcap		
Variable	**bremming**	**low-brem**	**bad-track**	**bremming**	**low-brem**	**bad-track**
$\mathrm{E_T} > 30$ GeV						
$\Delta\eta_{in}$	9.15E-03	3.02E-03	6.10E-03	1.35E-02	5.65E-03	7.93E-03
$\Delta\phi_{in}$	3.69E-02	3.07E-02	1.17E-01	4.75E-02	2.16E-02	1.17E-01
E_{seed}/p_{in}	8.78E-01	8.59E-01	8.74E-01	9.44E-01	7.37E-01	7.73E-01
H/E	8.71E-02	2.89E-02	7.83E-02	9.46E-02	2.45E-02	3.63E-02
$\sigma_{\mathrm{i}\eta\mathrm{i}\eta}$	1.31E-02	1.06E-02	1.15E-02	3.06E-02	2.80E-02	2.93E-02
Tk Iso.	6.53E+00	4.60E+00	6.00E+00	8.63E+00	3.11E+00	7.77E+00
ECAL Iso.	2.00E+01	2.72E+01	4.48E+00	1.35E+01	4.56E+00	3.19E+00
HCAL Iso.	1.09E+01	7.01E+00	8.75E+00	3.51E+00	7.75E+00	1.62E+00
Impact Par.	2.39E-02	2.70E-02	7.68E-02	2.31E-02	1.78E-01	9.57E-02
Missing Hits	5.50E+00	1.50E+00	5.00E-01	1.50E+00	2.50E+00	5.00E-01
$20 < \mathrm{E_T} < 30$GeV						
$\Delta\eta_{in}$	1.02E-02	2.66E-03	1.06E-02	9.03E-03	7.66E-03	7.23E-03
$\Delta\phi_{in}$	3.72E-02	2.46E-02	4.26E-02	6.12E-02	1.42E-02	3.90E-02
E_{seed}/p_{in}	8.60E-01	9.67E-01	9.17E-01	8.12E-01	9.15E-01	1.01E+00
H/E	6.71E-02	4.80E-02	6.14E-02	9.24E-02	1.58E-02	4.90E-02
$\sigma_{\mathrm{i}\eta\mathrm{i}\eta}$	1.31E-02	1.06E-02	1.15E-02	3.17E-02	2.90E-02	2.89E-02
Tk Iso.	5.42E+00	4.81E+00	4.06E+00	6.47E+00	2.80E+00	3.45E+00
ECAL Iso.	1.22E+01	1.31E+01	7.42E+00	7.67E+00	4.12E+00	4.85E+00
HCAL Iso.	1.16E+01	9.90E+00	4.97E+00	5.33E+00	3.18E+00	2.32E+00
Impact Par.	1.02E-02	1.68E-02	4.30E-02	1.66E-02	5.94E-02	3.08E-02
Missing Hits	3.50E+00	5.50E+00	5.00E-01	5.00E-01	5.00E-01	5.00E-01
$\mathrm{E_T} > 20$ GeV						
$\Delta\eta_{in}$	1.04E-02	2.22E-03	1.30E-02	1.48E-02	4.65E-03	1.48E-02
$\Delta\phi_{in}$	7.57E-02	1.53E-02	3.38E-02	2.02E-02	1.51E-02	2.07E-02
E_{seed}/p_{in}	8.35E-01	9.50E-01	8.34E-01	8.43E-01	7.49E-01	1.20E+00
H/E	1.00E-02	5.26E-02	2.80E-02	1.73E-02	1.79E-05	1.80E-06
$\sigma_{\mathrm{i}\eta\mathrm{i}\eta}$	1.35E-02	1.06E-02	1.10E-02	3.48E-02	2.90E-02	3.47E-02
Tk Iso.	4.41E+00	6.00E+00	5.58E+00	4.35E+00	3.95E+00	1.47E-01
ECAL Iso.	1.09E+01	1.39E+01	9.47E+00	1.03E+01	1.17E+01	1.03E+01
HCAL Iso.	1.99E-01	5.73E+00	3.55E+00	1.06E-02	3.73E+00	7.31E-03
Impact Par.	2.27E+00	6.78E-03	4.99E+00	5.41E+00	3.25E-01	5.84E+00
Missing Hits	5.00E-01	5.00E-01	5.00E-01	5.00E-01	5.00E-01	5.00E-01

Table 7: Cut values for the Category based tight selection. For each variable six cuts are set: bremming, low-brem and bad-track electrons, barrel and endcap separately.

B Category Based Identification based on the energy momentum classes

At the reconstruction level, the optimal determination of the electron momentum is based on a classification of electrons according to both tracker and ECAL estimate of the amount of the bremsstrahlung emission. Such classification was also proposed to used as basis of the electron identification, it separates background like shower patterns from well behaved cluster-track patterns [5].

The electron classification, initially proposed in [1], has been recently revisited [2]. Here we give only a brief description of the properties of four, mutually exclusive, electron classes:

1. *golden electrons*: this class represents the most precisely measured electrons, which are least affected by bremsstrahlung and have a good track-supercluster match. The pattern in the ECAL is characterized by a single seed cluster.

2. *big brem electrons*: this class contains the non-golden electrons characterized by a single seed cluster in ECAL, but with a large fraction of the initial energy radiated very early or very late in the tracker, resulting in the well behaved energy measurement in the ECAL.

3. *showering electrons*: this class contains the electrons which are badly measured, due to bremsstrahlung loses resulting in a supercluster made of multi sub-clusters or badly match the track momentum.

4. *crack electrons*: this class contains the electrons with those impacting either in the transition region between the barrel and the endcaps or in the ECAL inter-module borders or the innermost ring of an ECAL endcap.

A set of cuts on the following five variables have been studied: H/E, $\sigma_{\mathrm{i}\eta\mathrm{i}\eta}$, $\Delta\eta_{\mathrm{in}}$, $q \times \Delta\phi_{\mathrm{in}}$ and $\mathrm{E_{seed}/p_{out}}$. Thresholds are different for the different classes. Fig 30 represents graphically the variables used in this algorithms divided in the classes described above. Distributions are normalized to unity to enhance the shape differences.

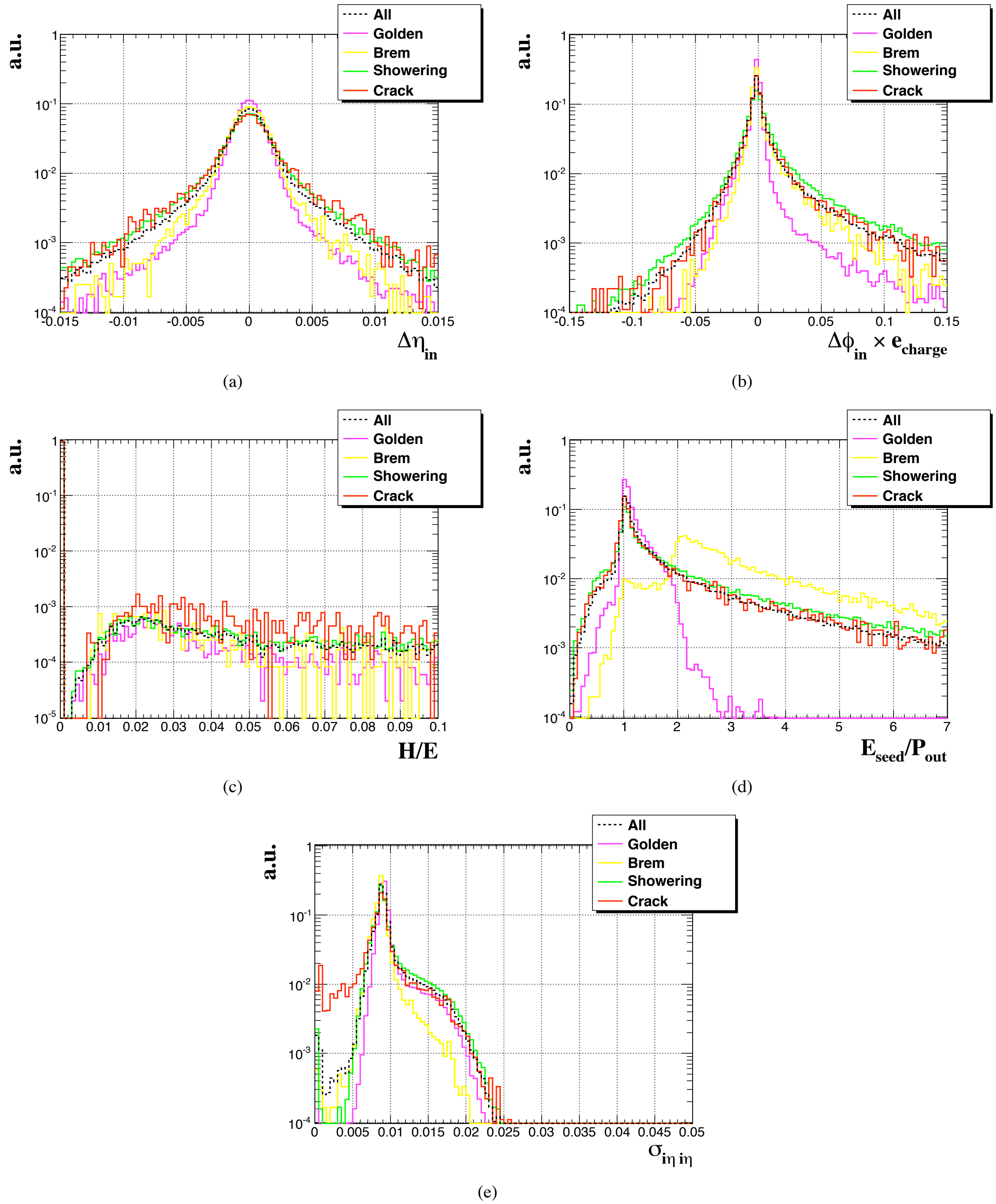


Figure 30: Distribution of electron identification observables in the η range of the ECAL barrel. (a) $\Delta\eta_{\rm in}$, (b) $q \times \Delta\phi_{\rm in}$, (c) H/E, (d) $\sigma_{\rm i\eta i\eta}$, (e) $\rm E_{seed}/p_{out}$. The distributions are shown for different classes of electrons and for all the classes summed up (dashed). Similar distributions with larger spread are obtained for electrons in the endcaps.